



An original Data, Information and Knowledge management approach for model-based engineering projects

Mouna El Alaoui, Souad Rabah, Vincent Chapurlat, Victor Richet, Robert Plana

► To cite this version:

Mouna El Alaoui, Souad Rabah, Vincent Chapurlat, Victor Richet, Robert Plana. An original Data, Information and Knowledge management approach for model-based engineering projects. 10th IFAC Conference on Manufacturing Modelling, Management and Control (MIM) 2022, IMT Atlantique, Nantes, Jun 2022, Nantes, France. pp.3262-3267, 10.1016/j.ifacol.2022.10.135 . hal-03834889

HAL Id: hal-03834889

<https://hal.science/hal-03834889>

Submitted on 31 Oct 2022

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

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

An original Data, Information and Knowledge management approach for model-based engineering projects

M. El Alaoui*, S. Rabah *, V. Chapurlat*, V. Richet**, R. Plana**

* *Laboratory for the Science of Risks, IMT Mines Ales, Ales, France*
(e-mail: mouna.el-alaoui@mines-ales.fr; souad.rabah-chaniour@mines-ales.fr;
vincent.chapurlat@mines-ales.fr).

** *Assystem Energy and Operation Services (AEOS), Courbevoie, France* (e-mail:
rplana@assystem.com; vrichet@assystem.com)

Abstract: Data, Information and Knowledge (DIK) problematics are of undeniable growing actuality; many efforts were made to explore and make progress in this domain. Therefore, these DIK have been defined by several characteristics conventionally studied such as volume, variety, variability... and a few solutions have been uncovered and revealed. In parallel, the systems Engineering (SE) and specifically Model-Based Systems Engineering (MBSE) approaches, that encourage, among other things, the use of models instead of documents in critical infrastructure engineering namely in the energy infrastructure domain, participate in the growing volume and complexity of DIK. In this particular context, it is important to enrich MBSE by adapting the existing DIK advances to the MBSE needs, which is initiated in this paper, first by providing DIK definition, second, by checking what are the main issues that shall be solved in order to help stakeholders use and manage them in large MBSE-driven projects, involving various business actors during more or less long periods. At last, this article proposes a methodological contribution bridging the supposed gap between MBSE and so-called data management fields of research.

Copyright © 2022 The Authors. This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Keywords: System Engineering, Model Based System Engineering (MBSE), Knowledge management, SE Ontology, Large engineering projects, data, information, knowledge, model

1. INTRODUCTION

The current energy transition is characterized by constantly growing and complex needs in terms of delivery performance. Particularly, cost and time overrun are becoming mitigated to meet the future energy demands. This is also exacerbated by the trends to have energy infrastructures featuring larger size and featuring more requirements in terms of availability, safety and security. In term of project delivery, this will translate into a higher complexity that will have to be handled. To accommodate this complexity, the different actors are mentioning the digital transformation of the engineering that will couple a data centric approach with a system modelling approach. As the number of energy infrastructures projects are increasing, there is an important feature that will arise related to industrialization of the projects to capitalize from one project to another and to the lessons learnt knowledge that will be essential to optimize the delivery of the future projects.

It has to be further mentioned that the digital transformation process will not be very simple as it will have to incorporate the legacy of company, stakeholders and projects. This will generate what is referred to as heterogeneous and complex data that will have to be handled. There is an observed lack of methodology to handle the delivery of complex projects that need to be covered and that will be the topic of the paper under discussion.

In the literature, the complexity is addressed through System Engineering (SE) principles and processes that are defined and formalized in the technical standard (ISO 29110, 2015), and

the Guide to the Systems Engineering Body of Knowledge known as (SEBoK, 2016) that was created in collaboration between the International Council on Systems Engineering (INCOSE), The IEEE Systems Council, and Stevens Institute of Technology where we find similar issues related to different aspects of Data Information and Knowledge (DIK) handled in such situations. In this article, we propose to introduce a new way to develop a Data Information and Knowledge (DIK) System that will apply to the future energy infrastructures. This article intends, at first, to define and characterize the DIK concepts, second, to focus on the encountered problems and locks with a concrete example and illustration, and finally, to propose a methodological contribution to provide some mitigation elements for these specific elements.

2. PROBLEM STATEMENT

2.1. DIK definition

Indeed, in SE-based projects, it is essential to acknowledge the difference between data, information, and knowledge. As a matter of fact, data can be defined as raw unprocessed events (i.e., signals, codes...), that have yet to be processed in order to make sense of what they represent. On the other hand, information is considered to be data that has been interpreted in a given context to make sense of it. In the literature, many authors tried to describe the term of ‘Information’. For instance, (Terra and Angeloni, 2003) gathered 3 main definitions: “Organized data”, “Data endowed with relevance and purpose”, and “Interpreted data”. The last term however is much more complex and challenging, philosophically,

“knowledge can only reside in one’s mind and is the result of human experience and reflection based on a set of beliefs that are at the same time individual and collective” (Terra & Angeloni, 2003). In other words, it is much more difficult to capture it and/or manage it since it mainly relies on human cognitive capacities such as experiences, values, contextual information and intuition all of which are not palpable nor scalable. All of the above justify that knowledge is a concept that it hardly captured, evaluated, shared and more generally managed. Particularly, all along SE processes, various models result from different activities (e.g., requirements or architectural engineering). Independently from their nature, format, aims and objectives, they must be then considered being effectively seen and categorized at a moment as:

- a **data**: any model is built by a project stakeholder according to their specific objectives and points of views and therefore must be contextualized in order to be understood by the rest of the stakeholders, or;

- an **information**: when a model is considered as an information, it is then a perfectly understandable and unambiguous source of information for all the stakeholders who are then able to refine, improve or question it or use it to build their own representations, or;

- a **knowledge**: a model may for instance formalize a modelling pattern, reusable and customizable for various purposes. It can be then considered as a knowledge considering any pattern validated and proven, or;

- a **container** that includes, displays, and put in context both various data, information and/or knowledge that could be relevant to handle, share and trace.

2.2. DIK characterization

Considering these definitions, in most engineering projects, these Data, Information, and Knowledge (denoted DIK in the following) are characterized by several features as synthesized in Figure 1 and studied in the literature, selected, and detailed below:

- **Volume**: in terms of size and quantity making it vital to be able to manage important amounts of DIK of variable sizes.

- **Variety**: the different types and formats they can come in, diverse sources and containers, various complex dependences with each other participates in making their structuration, organization, and exploitation more challenging.

- **Variability**: DIK are constantly changing in terms of value for instance, meaning an element can be accurate in a given moment and turn into an obsolete one that needs to be updated at any other moment, this issue leads to a variation of versions in an attempt to keep up with the variability of DIK.

- **Velocity**: another important notion is the speed of creation and evolution of DIK and/or the modification of existing ones (closely intertwined, respectively, with the ‘Volume’ and ‘Variability’ dimensions);

- **Veracity**: DIK’s veracity is perhaps one of the most important aspects that should be closely monitored and supervised since as mentioned above, it can be subject to sudden frequent changes. Its importance in engineering projects is at its highest for obvious reasons; a non-reliable or obsolete element can lead to disastrous losses (i.e., of time, budget, resources...) in attempt to address and resolve the errors or mistakes it may have caused (e.g., reworks...). The impact and gravity of these losses depends strongly on the domain of activity and the project’s phase.

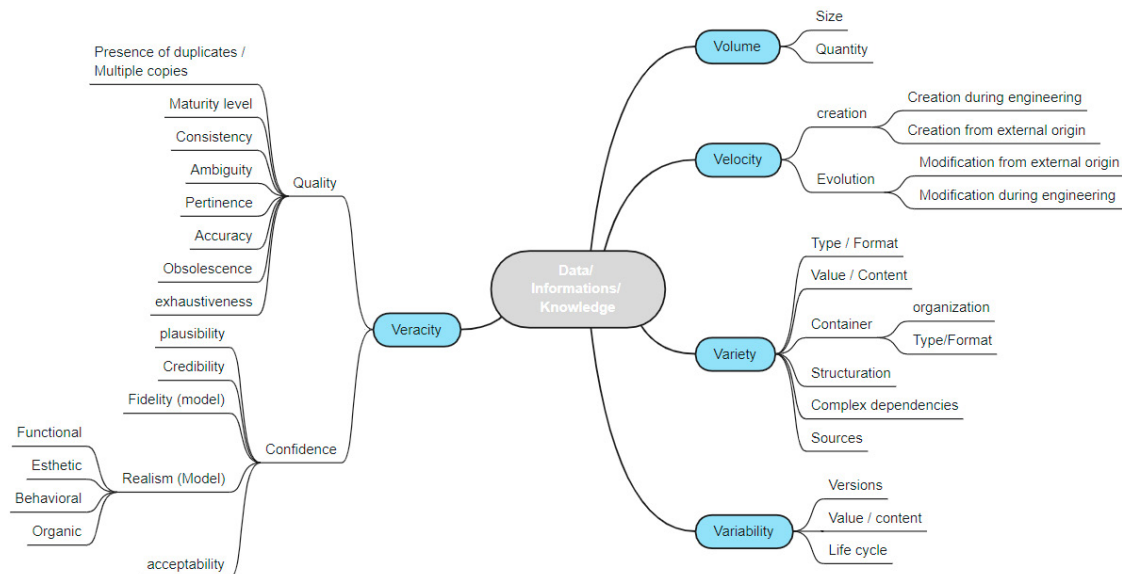


Figure 1: Representation of the proposed DIK characteristics

The ‘veracity’ includes two sides, the first one being the quality of the DIK; are we in presence of a coherent, non-ambiguous, pertinent, mature, and non-obsolete DIK? How can we make sure it is unique and that there aren’t any duplicates? Did we achieve the requested exhaustivity? The second term is the trust we can put on the DIK; how credible, acceptable, reliable, plausible, and realistic is it?

These notions are very similar to the ones present in the ‘Big

Data’ concepts as presented and developed by (Ishwarappa and Anuradha, 2015). They particularly introduced Big Data as “a collection of massive and complex data sets and data volume that include the huge quantities of data, data management capabilities, social media analytics and real-time data” which justifies fairly the relationship between Big Data’s 5V and the DIK’ 5 characteristics presented below and as it will be illustrated in §3.

- Semantic disambiguation. Proven and unambiguous Systems Engineering concepts definition remains limited, even formalized by the way of one or more coherent SE ontologies and being eventually enriched or driven by considering Model-Based Systems Engineering (MBSE) or Model-Based System of Systems (MBSSE) principles;

- Modelling Languages and models' interoperability as discussed by (Zacharewicz *et al.*, 2020) where they tried to connect enterprise modelling, enterprise interoperability, and the use of models to contribute and enrich the SE approaches;

- Tools interoperability: technical exchanges of data following exchange standards by the way of neutral files and format exist (for instance the STEP ISO 10303 Application Protocol AP233 (AP233) aims to capture the semantics of the information exchanges in SE contexts) but large projects require various tools, more or less specialized, leading to difficulties to establish and maintain the requested interoperability level;

- **Effective DIK V&V methods:** Throughout the projects' lifecycle, veracity and up to date value of DIK must be managed considering all other characteristics of volume, variability, velocity and variety. For this purpose, and prior to any use and integration of these DIK, dedicated Verification and Validation methods remain poor to face and reach these objectives;

- **Structuration/organization** of all DIK, step by step, considering both growing volume, variety (e.g., heterogeneous formats) and variability (e.g., regarding the so-called DIK life cycle) all along a project remains difficult considering expected uses (e.g., archive, traceability, transformation, computation...) and habits from projects' stakeholders;

- **Security:** although this concern is very common, there aren't clear norms and protocols dedicated to DIK security on one hand, and intellectual property management, in the context of MBSE-based projects.

3. PROJECT OBSERVATION SUPPORTING THE ANALYSIS

As previously stated, the difficulties exposed in §2.3 can be observed in most major engineering projects. An example is a recent project concerning nuclear power plant's spare parts. The objective was to manage these parts all along their lifecycle, from procurement to on-site commissioning. The required DIK concerns were:

- related to parts description and management: creation, modification, suppression, deduplication, update of catalog...;
- related to supply and procurement process from supplier to the local parts inventory: parts requests, reception, and availability update;
- related to parts delivery from local parts inventory to maintenance operations location: requests management, parts transit, and delivery.

Due to the nature of the domain of activity, the tolerance of errors and delays is very thin while the complexity is considerably high, therefore, the observed issues can be classified as seen before:

- **Volume:** at the time of this project, the number of registered spare parts in the catalogs was more than 200,000 parts where each one is characterized by various attributes that need to be kept up to date. Numerous stakeholders belonging

to different organizations and business units were involved in the project;

- **Variety:** due to the stakeholders belonging to different organizations and what that implies in terms of security reinforcements, the DIK management system was organized through six different digital tools that are not fully interoperable, making exchanges difficult without errors or mistakes. This implies to be able to deal not only with DIK of different types and formats but stored in different locations and managed by various tools and servers;

- **Variability:** Since the parts are always requested for maintenance operations, then supplied and delivered, their attributes evolve and change at any given time. In addition, other changes from the supplier could occur concerning any attribute of the parts;

- **Velocity:** Regarding the critical nature of the nuclear field, no matter what the velocity of the DIK change is, it is crucial to be prepared to deal with it and keep parts catalogs up to date for the unpredictable need of parts;

- **Veracity:** the DIK concerning the parts must always be up to date and representative of the reality.

As mentioned earlier, the tolerance in this particular domain is justifiably close to zero, yet the larger-scale issues observed in this project were that the catalogs were not always up to date, the delivery was not always on time or even conform to the demand, many parts were obsolete or duplicated in the catalogs and many efforts, time and resources are then deployed to rectify these problems.

The adopted approach consisted of tracing and tracking the parts lifecycle and describe it based on existing documents and experts' interviews. Models were then used as communication and representation means. At the end of this project, a digital model of the process was provided that describes:

- The actions made on the parts;
- The actors performing these actions, and responsibilities;
- The interactions between the actors.

Among other, this model was an excellent communication base that was understandable by the different stakeholders. As a result, several improvement opportunities were identified to address the different faced issues in the short term, such as:

- Reduce the human errors by creating bridges between the tools allowing to exchange certain DIK;
- Help master the complexity of the DIK by automating as much as possible the actions of the processes.

These solutions are only temporary fixes, since they do not take into consideration all locks previously exposed. The expected contribution of this work is therefore a more robust and durable fix that will address the problem from the source by proposing a method that suggests a new way of managing DIK and their characteristics throughout their lifecycle, and simplifies the DIK's search, exploitation, use and reuse (i.e., REX...).

4. EXPECTED CONTRIBUTION: AN ONTOLOGY-BASED METHOD FOR SE-RELATED DIK M

The proposed contribution to face these locks is a method called in the next MoDIK that aims to provide DIK management principles in the context of SE based projects. Its primary goal is to suggest a new way of structuring, organizing, storing, and more generally speaking managing

DIK systems that are generated, processed, and shared in large engineering projects. This method will take into consideration the DIK characteristics and problematics defined in §2 and is expected to contribute to the definition of a new way DIK is sought & found, exchanged, traced, and capitalized from a project to another.

The method is composed of five key elements essential for any attempt to establish an efficient method:

- **Concepts:** description and definition of all the concepts, relations between concepts and attributes of such concepts and relations, all required to model domain's and businesses' fields concerned by DIK management. This consists on the definition of a common and shareable vocabulary, and on its formalization to make sure all these elements have been identified, formalized, checked, and validated beforehand, then are used in a right and mutual way.

- **Language:** in an attempt to allow users to handle and work with the previously defined vocabulary, the method shall provide various languages i.e., Domain Specific Modelling Languages (DSML) for modelling, simulation, or analysis purposes. These DSML are basically defined by: 1) an abstract syntax defining which concepts and relations from the vocabulary are concerned, 2) one or more concrete syntaxes allowing graphical or textual rendering, conform to the abstract syntax and allowing then to provide end-users with a representation that corresponds with his/her usage and habits, and 3) an operational semantics composed or interpretation rules and properties allowing then to favorize formal analysis of built models contents and simulation. In the literature, a DSML abstract syntax is defined considering the different views and points of views that are requested by end-users e.g., requirements engineering view, behavioral view, or architectural view. Various existing DSML exist and must be then integrated in the method. Other DSML could be also defined from scratch, or get inspired from existing DSML, such as SysML for instance, and look for ways to improve and complete it in order to meet DIK management requirements.

- **Processes:** The goal is here to define, structure and organize the requested actions to follow during the method application in a real use-case. The processes need to be sufficiently clear and understandable to allow users to use step by step the constructed method by applying the previously defined language.

- **Tools:** any method must be supported by both methodological and IT tools and means for the end-users to create and elaborate models with the proposed DSML and following the specified processes.

- **Expertise and experiments:** its aim is to highlight and make available and shareable various experiments, references models, related standards, identified patterns, material course for supporting end-users for method acquisition and application.

The work done at this moment focus on the vocabulary of concepts, attributes and relations. In complement with the previous results presented in Figure 1 and Figure 2, the next steps consist of:

- Analyzing existing works about 1) SE and MBSE related concepts e.g., distinguishing SE ontologies (Lopez *et al.*, 2020) and meta models (Maleki *et al.*, 2018), DSML (Laurenzi *et al.*, 2018), model types (architectural,

requirements repository, recommendation, decision... -, and model management, and 2) Data sciences applied to DIK formalization, handling and management;

- Formalizing the whole in one or several and interoperable ontologies, gathered into a so-called MoDIK-ontology;

- Checking MoDIK-Ontology by crossing feedbacks from both classical data analysis methods (top-down approach) and workshops with end-users (bottom-up approach).

(Gruber, 1993) defined ontologies as follows, "*An ontology is an explicit specification of a conceptualization. The term is borrowed from philosophy, where an ontology is a systematic account of Existence*". In other words, an ontology is a knowledge representation method mostly used in the computer science domain and specifically in Knowledge Engineering and Artificial Intelligence. (Broekstra *et al.*, 2002) explained that a conceptualization (used to define ontologies) refers to an abstract model of a given domain that identifies and defines the representative concepts of the domain. Moreover, they stated that an ontology is:

- **Explicit:** meaning that both the concepts used and the restrictions on their use must be explicitly and unambiguously defined;

- **Formal:** ontologies must be formalized enough to be understood by the system meaning it must be able to interpret the semantic of the entered information;

- **Shared and mutual:** ontologies represent a consensual knowledge, and that it is not restricted to a few individuals but accepted by a collective.

The obtained ontology is then modelled as a knowledge graph of the domain, representing the concepts and attributes as respectively nodes and properties, and the edges as relations. The SE's interest in ontologies is no recent news, in fact, many attempts were made to create a SE ontology. In fact, (Yang, Cormican and Yu, 2019), in their state of the art proposed a complete table that displays, among other things, the references, the scopes, the key concepts, and the key properties of the existing ontologies. The conclusion that can be made from this table is that in the SE and MBSE scopes, many attempts were made to come up with a functional mutual ontology for SE uses, however, these initiatives present the following issues:

- The ontologies are not complete: a noticeable fact that pops out of Yang *et al.*'s work is that each of the existing ontologies addresses a certain aspect of SE, which implies that none of them actually covers all the concepts and aspects of SE;

- They are not semantically aligned: a number of ontologies use different designations to describe the same concepts, without necessarily creating a corresponding link to assure the semantical alignment, which goes against the basic principles of ontologies;

- There are some concepts that exist in some ontologies but lack from others or are replaced by other concepts (which is also observed in the different SE languages). It does not represent a problem itself, but it needs to be well documented and formally described.

Consequently, the specification of MoDIK-ontology is based on existing ontologies and aims to align them semantically.

For this purpose, a list of targeted stakeholders (e.g. Sponsor or Program Manager, Customer, Operator or user, and even People acting against the system), and their expectations & objectives has been described in an exhaustive study as detailed for instance in the extract shown in Figure 3 of the

creation phase of DIK for these stakeholders' usages. The result is being verified and validated by the said stakeholders in order to prevent future misunderstandings and/or errors later in the work. This requires then to populate the MoDIK-ontology by creating instances of the available classes.

What are concerned concepts?	For doing what?	When?	How to proceed?	With what or who? (material, application, and human resources)
Initial need	- Express the expectations, unsatisfactions, and problems of the project owner in their own vocabulary	- As soon as the project owner identifies one or more problems, dissatisfactions, limitations, ... and starts looking for partners to address these problems	Not concerned a priori	Not concerned a priori
Value	- Clarify and start formalizing the expectations, dissatisfactions, and problems of the project owner in his vocabulary	- As soon as the opportunity of a business and a contract arises for the contractor	Through : - Meetings - Reviews - Exchange of documents - Interviews...	- Representatives of the project owner and contractor - Premises, means of communication, exchange, creation and visualization of documents, ... - Architecting process and Business & Mission Analysis process
Expressed need	- Communicate with the contractor with the project owner's vocabulary	- As soon as the opportunity of a business and a contract becomes precise and concrete between the client and the contractor		
Perceived need	- Specify and progress in the formalization the expectations, dissatisfactions, and problems of the project owner in his vocabulary	- As soon as an agreement and a contract defining the roles and objectives of the contractor towards the project owner are defined and an owner/contractor engagement is made		
Retained need	- Communicate with the contractor with the project owner's vocabulary			
Specified need	- Formalize the repository of needs (clarify, classify, specify, detail, argue...) that will guide the future conception	- From the start of the conception	- Architecting process, business analysis and mission analysis, requirements engineering as part of System Engineering	- SE / MBSE method - IT tools - knowledge repositories acquired by the company through previous projects - ...
System need				
Needs' repository				

Figure 3: Extract of the study of DIK's concepts: the creation phase of DIK

5. CONCLUSION AND PERSPECTIVES

As detailed and presented in this paper, in the critical infrastructures scope, DIK management makes sense in most engineering projects that follow SE and MBSE principles and processes. This article of intention defines the main DIK characteristics and problematics and presents a concrete project that illustrates it, then introduces the locks to be studied. The proposed contribution is a method based on various existing DIK managements strategies and principles. The first part of the contribution focus on conceptual part of the method, aiming to explore existing ontologies then to align and eventually merge them creating a more suitable and unambiguous SE ontology, that will be populated and checked, then confronted to the reality of large engineering projects.

6. REFERENCES

- Broekstra, J. et al. (2002) Enabling knowledge representation on the Web by extending RDF Schema, *Computer Networks*, 39(5), pp. 609–634.
- Gruber, T.R. (1993) A translation approach to portable ontology specifications, *Knowledge Acquisition*, 5(2), pp. 199–220. doi:10.1006/KNAC.1993.1008.
- INCOSE, B.E.B. (2016) *Guide to the Systems Engineering Body of Knowledge SEBoK 2.2, (SEBoK)*.
- Ishwarappa and Anuradha, J. (2015) A brief introduction on big data 5Vs characteristics and hadoop technology, in *Procedia Computer Science*. Elsevier B.V., pp. 319–324.
- ISO and IEC (2016) ISO/IEC TR 29110-1:2016 Systems and software engineering — Lifecycle profiles for Very Small Entities (VSEs) — Part 1: Overview, *International Organization for Standardization*, Geneva, Switzerland .
- ISO, IEC and IEEE (2015) ISO/IEC/IEEE 15288:2015 Systems engineering—System life cycle processes, *International Organization for Standardization*, Geneva, Switzerland [Preprint].
- Laurenzi, E. et al. (2018) Towards an Agile and ontology-aided modeling environment for DSML adaptation, in *Lecture Notes in Business Information Processing*.
- Lopez, B. et al. (2020) Ontology configuration management for knowledge-centric systems engineering in industry, in *Proceedings - 50th Annual IEEE/IFIP International Conference on Dependable Systems and Networks: Supplemental Volume*, DSN-S 2020.
- Maleki, E., Belkadi, F. and Bernard, A. (2018) A Meta-model for Product-Service System based on Systems Engineering approach, in *Procedia CIRP*.
- Terra, J.C. and Angeloni, T. (2003) Understanding the difference between Information Management and Knowledge Management, *KM Advantage* [Preprint].
- Vernadat, F.B. (2007) Interoperable enterprise systems: Principles, concepts, and methods, *Annual Reviews in Control*, 31(1), pp. 137–145.
- Yang, L., Cormican, K. and Yu, M. (2019) Ontology-based systems engineering: A state-of-the-art review, *Computers in Industry*. Elsevier B.V., pp. 148–171.
- Zacharewicz, G. et al. (2020) Model Driven Interoperability for System Engineering, *Modelling*, 1(2), pp. 94–121.