Towards Knowledge Infrastructure for Highly Variant Voltage Transmission Systems

Mathias Uta and **Alexander Felfernig**

Abstract. The high voltage transmission business uses very mature technical solutions to transport electrical energy over large distances. New developments in the information technology sector are now promising opportunities to revolutionize the traditional processes within the business. In this paper the opportunities to implement a sophisticated knowledge infrastructure to improve the efficiency and quality of high voltage product manufacturers will be outlined. Therefore, possible solutions have been assessed to establish a non-redundant data structure and create an advanced database system architecture with respect to business specific requirements, considering in particular product configurators. Based on the proposed master data system, the possibility to create as well as integrate a knowledge system has been evaluated. Accordingly, the introduction of a global knowledge manager is proposed to organize inquiries of product configurators to expert systems and introduce a company-wide framework for rules and constraints. To assure communication between all parts of the software architecture, the implementation of a universally understandable format is discussed. Finally, possibilities to integrate recommendation system mechanisms into the suggested system architecture are highlighted.

1 INTRODUCTION

The market of high voltage transmission business is highly competitive. The differentiation through up to date processes, high quality and fast response times is highly desirable to manage the increasingly complex and fast changing customer requirements. This effect is intensified by general technical developments in measurements initiated by the energy transition from fossil to renewable energy sources.

Due to an increasing share of renewable energy into the grid and their decentralized physical arrangement compared to huge power plants, an extension of the electrical grid has to be provided. Additionally, the weather dependency of the new energy sources intensifies this effect. As a result, the utilities are confronted with a fast changing energy market and need to react to the new energy mixture [1]. Not only a transformation has to be managed on the supplier side but also consumers are starting to adopt new technologies with huge electrical energy demand like electrical cars. Therefore, the only solution to assure a secure energy supply in the future seems to be an extension of the transmission grid [2].

The structure of the electrical grid and the main components are shown in Figure 1. The renewable and non-renewable based power plants convert several energy sources to electrical energy and are connected via transformers to the grid. The overhead lines and the cables transport the energy over large distances to consumers. The switchgears integrate many of these lines on one conductor (busbar) and build nodes in the grid. In case of failures circuit breakers separate the faulty part of the grid to prevent blackouts. Additionally, current transformers and voltage transformers give transmission utilities the possibility to measure the transported energy. The collection of all main parts concentrated in one physical spot is called electric power transmission substation.

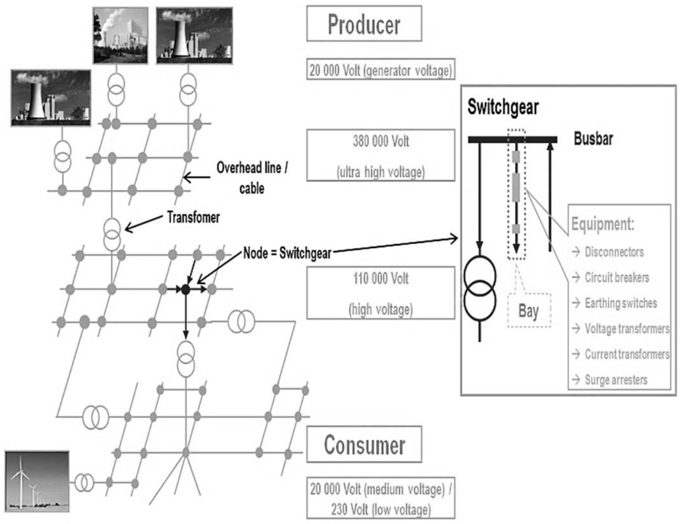

Figure 1. Electrical components in the energy grid [3]

Since each of the main modules has to correspond with many diversified customer requirements, the big suppliers in the business such as Siemens, ABB or Alstom have separated their energy transmission business in many smaller segments [4] [5]. Each of these segments is concentrated on one of the main components in the grid. Customers expect suppliers to offer solutions for specifications including a complete electric power transmission substation. As a result, very complex projects with many internal stakeholders have to be managed. This includes the decentralized data storage of in many cases identical information. In each of the involved business segments a unique tool infrastructure is installed and requires varying processes. In particular it is possible that, each business segment uses one or more individual configuration solutions to determine the bill of material derived from individual configuration rules and constraints. These configuration rules are cross segmental on a higher level and do not differ until a finely granulated level of configuration is reached. Furthermore, individual pricing and quotation tools are used which are sometimes integrated with the configuration tool, resulting in so called configuration prizing and quotation solutions (CPQ). As a consequence, very complex processes and redundant actions are decreasing competitiveness in a cost driven market environment. Additionally, data and knowledge maintenance cause disproportional effort leading to human and hardware resource binding.

To tackle the above described challenges, the introduction of a well-designed data and software infrastructure is necessary. A completely integrated data and software landscape which provides necessary information via interfaces to all quotation related tools should provide synchronization of all involved stakeholders and improve the maintainability of data and knowledge.

In an environment which utilizes several ERP (enterprise resource planning), PDM (product data management) and CRM (customer relationship management) tools, in future one company-wide integrated database should be implemented to deliver one data source for all configuration, pricing and quotation tools. The goal is to enter the information only once into the system so that the characteristic is passed on to all connected business segments and their CPQ tools. In consequence, all related objects rely on the same data input. This assures cross segmental consistency of data in all offer documents, technical descriptions and technical calculations.

Moreover, the integrated representation of data delivers the opportunity to implement consistent connections between objects in form of rules and constraints. This leads to the extension of data to knowledge and should support users by preventing false configurations and incorrect data input, i.e., to contract relevant documents. These relations between the objects need to be non-redundant as per the data itself. Supplemental to the rules and constraints describing simple relationships, interfaces to expert systems capable of very specific and complex calculations need to be established. Mechanical calculations due to earthquake requirements or ferroresonance calculations of the grid are two examples of these expert systems.

Finally, based on the achieved integrated data and knowledge, an analysis of the utilized materials, parameter characteristics and all other objects during the configuration, prizing, and quotation process can act as a basis for a self-learning system to improve the CPQ-tools. This system should be able to recommend technical solutions and parameter input based on previously chosen solutions in distinctive situations as customer specific requirements or special environmental conditions. The user in this scenario still has the opportunity to neglect the recommended solution which is consequently used to improve the recommendation logic.

Overall, a complex system composition consisting of domain specific and integrated databases with direct interfaces to configuration, pricing, and quotation tools has to be established. Further, the user of the quotation tools is supported by knowledge bases inheriting logical connections between relevant parts and an adaptive artificial intelligence which analyses user decisions.

The remainder of this paper is organized as follows. In section 2 we discuss the possibility of a holistic database and point out that different use cases lead to contradicting database types. Conducted from this perception we propose system architecture assuring data consistency for the high voltage transmission business. Based on this data consistency section 3 emphasizes the advantages of knowledge integrity. This is followed by a general survey on an approach to implement a global knowledge manager on basis of the proposed system architecture in section 4. Section 5 suggests a concept of a recommendation system in which the knowledge base is constantly extended using the results of configuration applications used by experts.

2 DATABASE EVALUATION

Data modelling plays an important role. To create a ubiquitously applicable database, one has to consider a data structure which allows multiple views and possibilities to manipulate the stored data via interfaces from manifold applications. Graeme C. et al. characterize this situation as follows 'The data model is a relatively small part of the total systems specification but has a high impact on quality and useful life of the system' [6]. Data modelling is concentrated on achieving completeness, non-redundancy, and reusability in databases. Additionally, stability, flexibility, and performance have to be taken into account. The implemented database should be capable of modifications and extensions to integrate requirements arising in the future. Nonetheless, the data modeler has to start with the given information to create a so called 'conceptual schema' which can be accomplished by different course of action like the process-driven, data-driven, or object-oriented approach to name only the most important ones. Based on the 'conceptual schema' a 'logical schema' can be deducted by using entity-relationship modelling (ERM) or unified-modelling language (UML) approaches. Finally, a physical design of the database can be created.

Typically used database types are relational databases (RDBs), object-oriented databases (OODBs) and a hybrid form which is called object-relational database (ORDB). Consistency in RDBs is achieved by using normalization methods for tables as proposed in [7]. The objective of OODBs is to create an abstract view on the reality which is easily adaptable to object oriented programming languages. Dependencies between objects are provided by pointers which allow m:n relationship representations. Therefore, OODBs are often adapted for computer aided design (CAD) or technical calculation programs (expert systems) since these programs deal usually with high complexities and many variants [8] [9]. Unfortunately, OODBs lack query performance compared to RDBs, which is why they are still not able to replace traditional relational databases for query focused applications [10]. ORDBs try to combine both approaches to achieve an improved performance and are often adapted by CPQ-tools. Michael Stonebreaker evaluated all three options in [11] and classified them as depicted in Table 1. The table points out ORDBs combine the advantages of RDBs and OODBs in comparison of the most important properties for databases, 'fast queries' and 'complex data management'. But ORDBs implicate some issues as well which is most importantly a low performance in web applications [10].

Table 1. A classification of database management systems (DBMS) [11]

	Simple Data	Complex Data
Query	Relational DBMS	Object-Relational DBMS
No Query	File System	Object-Oriented DBMS

A very detailed analysis of the given requirements is precondition for implementation of the most appropriate database and data structure. Since full integration of the whole value chain, including

PLM (product lifecycle management), ERP (enterprise resource planning), CRM (customer relationship management), configuration, pricing, quotation and technical calculations with expert systems, has to be realized, a solution with one centralized database seems with respect to deviating requirements difficult to implement. Table 2 shows adapted from Alejandro Vaisman's et al. a holistic comparison between these tools and their resulting utilized database types [12].

Table 2. Requirement comparison between ERP, CRM, PDM, CPQ and expert systems

Aspect	ERP / CRM / PDM	CPQ	Expert system
User type	Operators, office employees	Customer, sales employees	Engineers
Content	Current, detailed data	Current, detailed knowledge	Current, detailed knowledge
Data organization	According to operational needs	According to operational and analysis needs	According to operational needs
Data structures	Optimized for small transactions	Optimized for complex queries	Optimized for complex queries
Access frequency	High	High	Medium
Access type	Read, insert, update, delete	Read, insert, update, delete	Read, insert, update, delete
Number of records per access	Few	Many	Many
Response time	Short	Short	Can be long
Concurrency level	High	High	From medium to low
Update frequency	High	From medium to low	From medium to low
Data redundancy	Low (normalized tables)	Medium (normalized tables and objects)	Can be high (no normalization methods and inheritance)
Resulting database type	**Relational database (RDB)**	**Object relational database (ORDB)**	**Object oriented database (OODB)**

Whereas ERP, CRM, and PDM tools are designed for fast access to distinctive data to satisfy operational queries, CPQ and expert systems are focused on very complex queries that require combinations between many data sources. Additionally, CPQ-tools are facing a high access frequency with expected short response times which leads to an even higher level of complexity in the data structure.

In conclusion, based on these manifold requirements it seems more likely to implement a sophisticated system architecture including several customized databases for each application instead of establishing one ubiquitous database. The system architecture of such a conglomeration of databases is closer examined in the next chapter.

3 SYSTEM ARCHITECTURE

The data used along the value chain can be described most likely as 'master data' of a company. Master data is defined as 'data held by an organization that describes the entities that are both independent and fundamental for that organization, and that it needs in reference in order to perform its transactions' [13]. By other means, master data is all the data stored about customers, employees, products, materials, suppliers. The management of this data is called master data management (MDM) and can be provided by mainly four different architectures as shown in Figure 2.

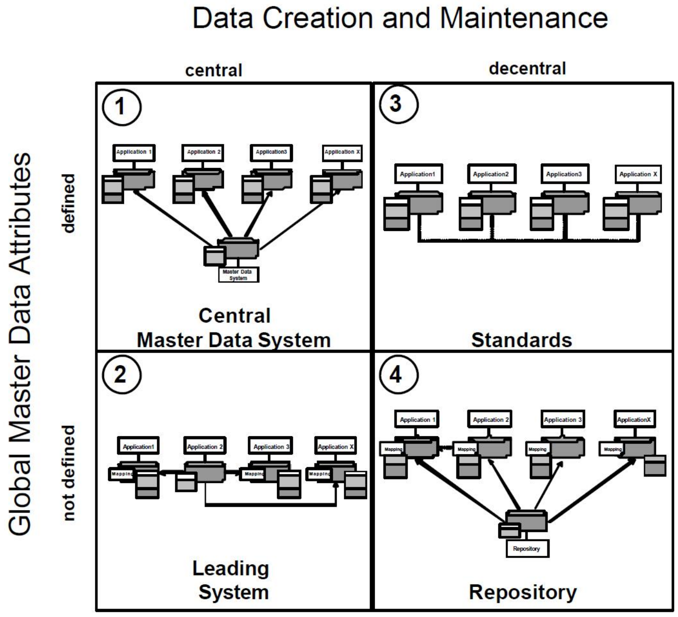

Figure 2. Classification of architecture approaches [14]

The architecture in this approach is classified by two major categories:

- Are global master data attributes defined or has each involved application its own master data definitions?
- Where is data created and maintained – in a centralized system or decentralized in each connected application?

In the centralized master data system (1) all global master data attributes are created and maintained in one database. Associated applications receive these attributes directly from this centralized system via identical primary keys whereas additional specific attributes can be defined individually in the single applications. The leading system methodology (2) instead has no global master attributes but a system which is responsible for data creation and maintenance. These attributes are transmitted to connected applications where the primary key of the attribute is mapped to the corresponding attribute in the application.

Decentralized data creation and maintenance approaches on the other hand rely on company-wide standard definitions (3) or an implementation of a repository (4). Company-wide standards do not assure consistency in each dataset of the applications. Instead they focus on the ability to create the same understanding of each attribute which leads to easier manageable comparisons between those datasets. A physical connection between applications is in this approach not mandatory. The installation of a repository follows the idea of metadata storage to connect applications with each other whenever an interaction is necessary. Therefore, primary keys of all involved applications for a certain attribute are stored in the repository and matched for cross-program query events with each other.

In case of the high voltage transmission business, an implementation of one integrated MDM solution seems, equivalent to the result of the second chapter for databases, not feasible. Two reasons are mainly responsible for implementation of a more complex structure. First of all the structure of the business includes, as already pointed out, several different independent departments responsible for distinct parts of the substation. These departments are not necessarily located at the same location but are distributed over the world to assure an increased satisfaction of locational requirements. Consequentially, one centralized system located, for instance, in Europe will lead to performance issues when it comes to queries from an application located in Asia. Live connections over long distances and many servers are due to the congestion control of the transmission control protocol not advisable [15]. A system which synchronizes independent and asynchronous to operational tasks is therefore preferable. Secondly, each department has sometimes not only one factory but several production sides each with their own portfolio. The main structure of the portfolio is same as per the department. The differences can be found on a more detailed layer of the products as, for instance, in the differentiation of high-end products and cost-optimized products which serve generally the same technical requirements but are differentiated in more sophisticated configuration options as better operation monitoring or increased maintainability of the electrical component. From these considerations follows, a solution incorporating combinations of the previously presented options seem more feasible to encounter the high complexity in the high voltage transmission business and causes the introduction of more than one MDM layer.

Figure 3 shows an option where a centralized master data system is implemented storing master data of one manufacturing location whereas the centralized database is connected to other locations via a repository to assure a synchronization of all business relevant data independent from daily operational tasks. Furthermore, standards across the whole company are defined to realize a homogenous creation of new data since each master data system is allowed to create and maintain its data by itself.

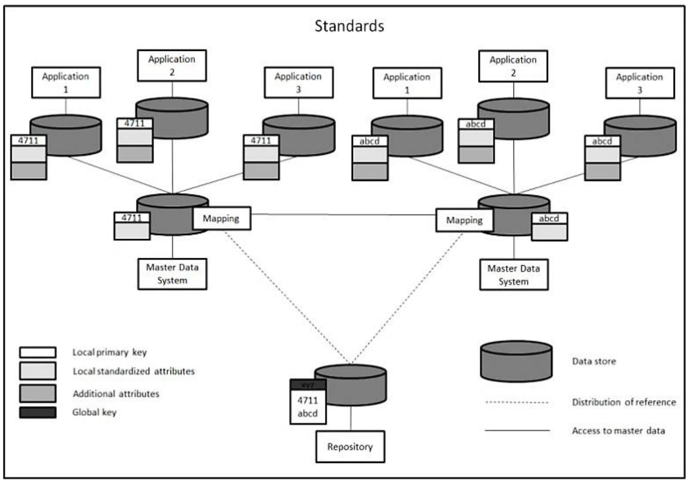

Figure 3. Three layer repository - master data system – standards approach

In another scenario a company-wide leading system located in at a distinct place is established to create and maintain all master data. This leading system is additionally used as basis for connected central master data systems installed in distributed factories of the company. These centralized master data systems receive the company-wide master data from the leading system and can be extended with manufacturing side specific master data as basis for connected applications.

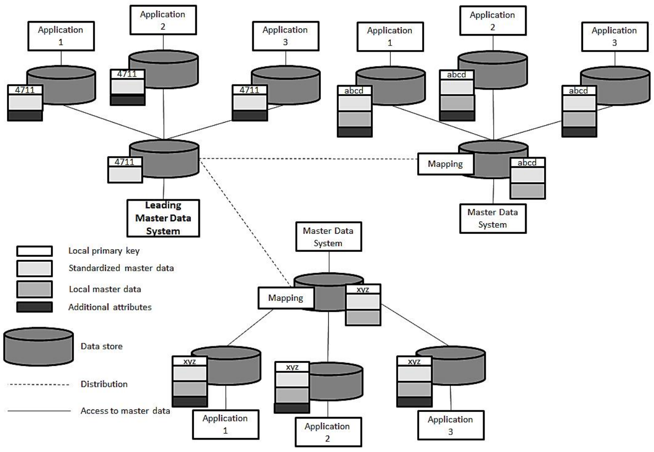

Figure 4. Two layer leading master data system approach

Further combinations are imaginable as well and have to be evaluated based on several reference values in further publications. Based on these results and taking all requirements for the high voltage transmission business into account finally a decision on the system architecture can be reached. Another important decision is the selection of the tool used as master data management system. The use of the ERP, CRM or PDM tool is imaginable since the most data is created in these databases. Additionally, the adaption of the CPQ-tool as main input tool for data might also be a reasonable choice.

Up to that point we focused on data structure and system architecture neglecting the value which can be created out of integrated, non-redundant data. The following chapters will discuss the possibilities given by such a dataset.

4 KNOWLEDGE INTEGRITY

A well-known use case for master data management systems is the implementation of data warehouses. Operational data is put through an ‘extraction, transformation, integration, and cleansing process (to) store the data in a common repository called data warehouse’ [13]. Data in the data warehouse is in reference to W.H. Inmon saved ‘subject-oriented, integrated, time-variant, nonvolatile (to) support management’s decision making process’ [16]. This means an OODB is implemented that saves data over a long time to enable analysis of developments in the observed business. Besides the data itself metadata – data about the stored data – is stored in the data warehouse. In big organizations, data is not directly used by queries to the data warehouse. Instead, data marts are introduced as an additional layer between the user and the data warehouse. Similar to the already analyzed central master data systems data is transmitted from the data warehouse (central master data system) to the data mart (application database) which fits to the requirements of a specific department. The user of the

data mart is finally able to create reports, perform data analysis and to mine data.
The restriction of this concept is already given in the definition of the data warehouse. The data warehouse should 'support management's decision making process' [16]. The area which besides the management decisions is widely neglected in centralized data and knowledge storage is the configuration and engineering sector. While huge effort is put into preparation of management reports and supporting management decisions, configuration tasks, technical calculations and dimensioning decisions are left completely to engineers and expert departments. This leads to a complete dependency of the company on their team of experts. The problem worsens if the dependence is only on a single expert. Attempts to consolidate the knowledge of engineering experts to more than one employee usually collapse because of budget and time restrictions. The experience of the expert cannot be transferred to other colleagues within in short period of time [17]. It sometimes requires years to reach the level of an expert and could result in loss of knowledge in case the expert retires.
Another issue which is more and more visible to companies arises with the implementation of more than one configuration software. Since the beginning of the new millennium a rising number of companies implement product configurators to treat a phenomenon called 'mass customization' [18]. 'Mass customization' is an oxymoron referring on the one hand to the growing production charges of companies to serve the market and on the other hand the increasing amount of individualization requirements by customers. The configuration technology relies on a subject-oriented, integrated, time-variant, nonvolatile dataset similar to the data warehouse. Combinations of entities and their attributes are configured according to combination and configuration rules to create a producible bill of material.
This technology was immediately adopted by the high voltage transmission business since the modular design of its products is predestinated to realize product configurators. Unfortunately, the degree of standardization in the business does not allow a full switch from CAD (computer aided design) applications to product configuration solutions. Special-purpose solutions define the high voltage transmission market up to fifty percent. The result is an increasing redundancy in the configuration data and knowledge. While the CAD system inherits engineering data accomplished by rules and constraints for the product with a high degree of freedom to support the engineering expert, the product configuration application is designed for sales people or even for the customer and includes a separated engineering dataset complemented by even more sophisticated rules, constraints and methods to prevent a wrong configuration. Nevertheless, a redundancy of engineering data and knowledge is introduced; leading to a high maintenance effort and complicated processes whenever the research and development department (R&D) releases technical innovations or the product lifecycle management department (PLM) disables the use of certain parts of the product.
Based on these evaluations, a centralized engineering database has to be implemented as foundation for all product configurators. PDM systems fortunately inherit all relevant product information and can, as the major tool for PLM and R&D departments, be utilized as a centralized engineering database. The result is a database valid for all configuration applications which are no longer maintained decentral by each product configurator but are centrally maintained by the product responsible departments. The technical realization of this requirement can be established by introducing interfaces between the product configurators and the PDM system. However, with this step an integrated data administration can be introduced but the knowledge administration is still redundantly managed in each configuration tool.
To achieve knowledge integrity - a major target formulated in the visions at the beginning of this paper - we propose steps connatural to the actions necessary to reach data integrity [7] [8] and which are also related to frameworks as formulated in [19]. First a 'conceptual schema' of the knowledge base has to be created using the object-oriented approach since several connections between the given objects have to be considered. Data in the PDM system is most likely stored, as analyzed in chapter 2, in RDBs and has to be addressed to objects in an OODB or ORDB. These objects have to be connected via rules and constraints on a level applicable for all product configurators to implement a 'logical scheme'. Therefore, normalization methods comparable to data normalization need to be formulated. In other words, a framework for product configurators needs to be established 'making the common parts common' [19] and making general rules and constraints applicable for all configuration tools to prevent redundant code implementation. Due to the generalization of rules and constraints on a higher level, a non-redundancy of the knowledge in the separated specialized configuration solutions is achieved and therefore, a centralized maintainability of the knowledge. Redundancy detection algorithms as proposed in [20] can consequentially be applied under these conditions to assure continuous knowledge integrity. Furthermore, the knowledge of experts can be transferred in incremental steps into the knowledge base avoiding redundancy in several tools and decreasing dependency on single persons in the company. Finally, a 'physical scheme' of the system architecture has to be created and is investigated in more detail in the next chapter.

5 A GLOBAL KNOWLEDGE MANAGER

An initial approach to set up a system architecture analog to the requirement formulated above was made in the paper 'On Knowledge-Base System Architectures' and is illustrated in Figure 5 [21]. The paper introduced a unit called global knowledge manager (GKM) to centrally handle inquiries by other knowledge-based systems and data type processors (e.g. traditional databases). To assure common semantics a translator or interface was proposed. A request incoming to the GKM is first scheduled and optimized by using the GKM's knowledge base and information delivered by the source of the request. The result is an access plan stored in the GKM's internal database. The monitor/interpreter uses this plan to process the request by applying the GKM's knowledge base rules, constraints and methods. The outcome is returned to the original source using common semantics.
This approach presumes a fully centralized inference of all inquiries. Any information in the system is gathered in the GKM, scheduled and interpreted. In consequence, a centralized expert system is introduced responsible to create and maintain all business relevant rules, constraints and methods irrespective of the complexity and domain of the request. A configuration task to calculate the mechanical forces on a circuit breaker in case of an earthquake would as well be processed as a simple request for earnings before interest and taxes (EBIT) calculation. The consequence would be a very complex set of knowledge

representations neglecting all domain specific requirements. Therefore, a master data and master knowledge approach assuring consistency between all connected tools might be a more feasible solution.

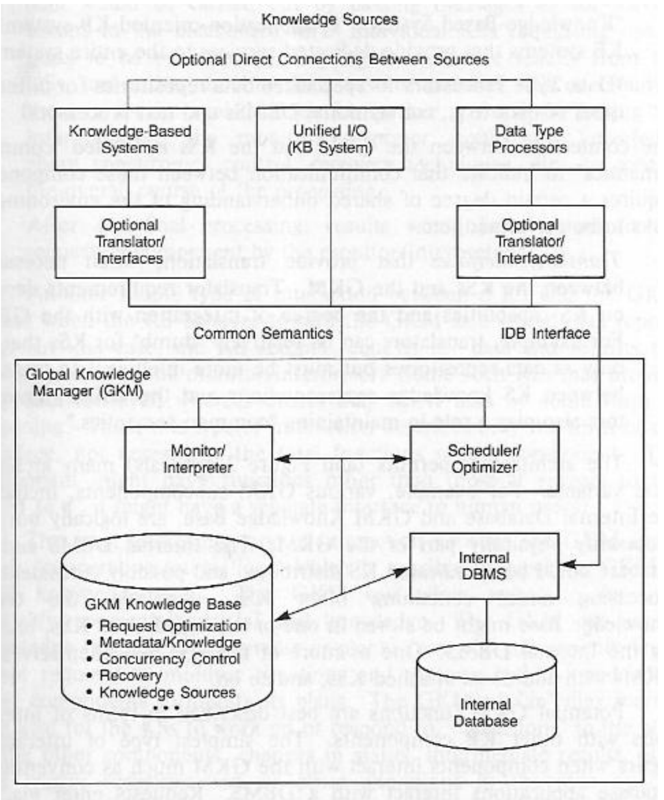

Figure 5. Architecture with global knowledge manager [21]

Figure 6 shows, based on the system architecture presented in Figure 3, an overview of the system architecture we propose to provide the requirements of an easy maintainable, integrated system for the high voltage transmission business. A company-wide standard or framework for creation of data, rules and constraints is defined to assure semantic consistency between all locations. The master data system is accomplished by the global knowledge manager to facilitate data and knowledge integrity in one location. While the master data management part is responsible for the data integrity of all databases, the global knowledge manager is comparably responsible for knowledge integrity in all connected configuration applications and expert systems. Additionally the global knowledge manager organizes special calculation inquiries by CPQ-tools to expert tools. Therefore, a universally understandable format (e.g. extensible markup language – XML) is introduced to enable communications between the GKM and all connected tools. Knowledge integrity during operation is assured by redundancy detection algorithms considering all knowledge bases of the software infrastructure. Additionally, a repository is tracking comparable data, rules and constraints of the separately operating locations by defining company-wide primary keys. Changes in the master data and global knowledge management system are synchronized asynchronous to operational inquiries to maintain consistency between all local data and knowledge sets by utilizing the repository primary keys. The behavior of the system is illustrated in more details by the two following examples.

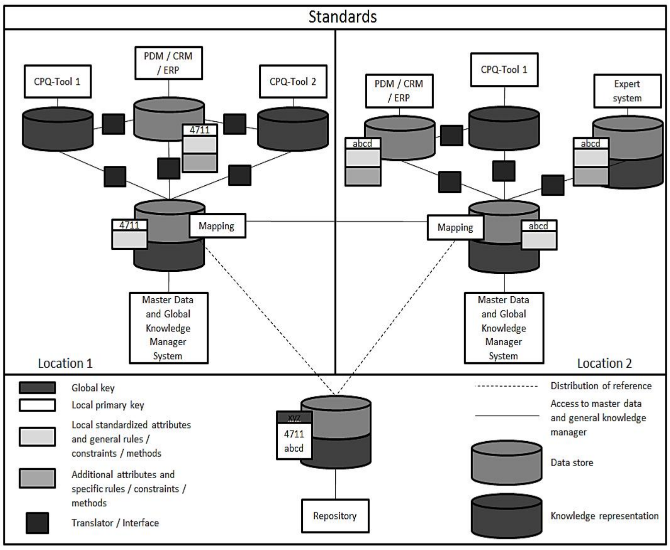

Figure 6. Three layer repository – global knowledge management – standards approach

In Location 1, which is in shown in Figure 7, a system architecture including two CPQ-tools and the usual databases is established. CPQ-tool 1 is responsible for very standardized configuration tasks while CPQ-tool 2 is an expert tool to configure specialized customer requirements. Simple if-then relationships which generally describe a part of the substation are maintained in the GKM. For instance: 'If product A is chosen then the cross section of the rectangular conductor cannot exceed 2400 mm².' The information about the product is received via an interface to the PDM database and is maintained in a fast accessible RDB database. In the GKM two objects (product A and rectangular conductor) are described by parameters (cross section, ambient temperature, body temperature, rated current) matched to the information delivered by the PDM database and are connected via the above mentioned relationship. This general rule is accomplished by a more specific rule in CPQ-tool 1 by the following relationship: 'If the rated current exceeds 2500 A at a ambient temperature of 35 °C then a rectangular aluminum conductor needs at least a cross section of 2400 mm² to not exceed 65°C body temperature.' [22].

A user of CPQ-tool 1 is in consequence not allowed to choose product A if the rated current exceeds 2500 A and an ambient temperature of 35°C is given. Contradicting to this rule, an expert in CPQ-tool 2 is not restricted by this constraint. An expert could decide to use product A if the current exceeds 2500 A at a ambient temperature of 35 °C knowing that the conductor is allowed to exceed 65 °C if the conductor is not touchable by humans. This expert knowledge includes besides product knowledge as well knowledge about spatial constraints and human interactions with the product and is with its complexity not easily describable in a non-expert configuration tool. Nevertheless both configuration systems will lead to a correct solution. CPQ-tool 1 will lead to the costlier product B but is operable by the customer herself while CPQ-tool 2 with a higher degree of freedom will deliver a cheaper solution of the configuration task but needs expert knowledge. Therefore, two options for the customer are established, a slightly costlier but very fast configuration or a very accurate solution with a more time consuming configuration.

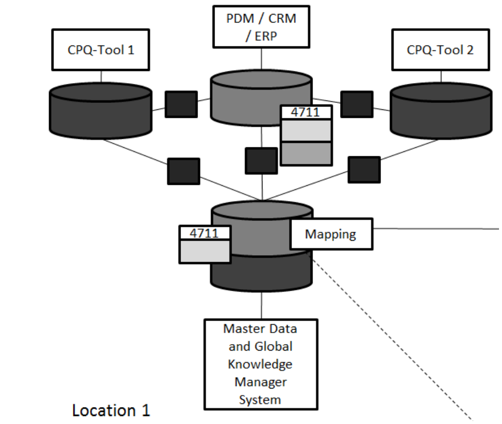

Figure 7. Multiple CPQ-tool system architecture

Location 2 is in detailed visible in Figure 8 and includes the standard databases, CPQ-tool 1 and an expert system. The previously mentioned configuration task to calculate the mechanical forces on a circuit breaker in case of an earthquake is given. This calculation task requires information which is not stored in the PDM database or any other standard available master data. Additionally, very specific calculation methods have to be applied not handled in the CPQ-tool. An expert system is necessary to solve the calculation task. The calculation inquiry is sent from the CPQ-tool via the universally understandable format to the GKM that interprets schedules and finally processes the request to the correct expert tool. In consequence, the GKM inherits besides generalized rules and constraints also information about the expert tools in the system and administrates access to them. The inquiry is handled by the expert system using the information received via the interface and by utilizing additional information like specific knowledge on the occurring earthquake forces in the concerned region to calculate the forces on the circuit breaker. The result is processed back to the GKM and further to the CPQ-tool where the resulting values trigger rules and constraints to decide which circuit breaker has to be chosen.

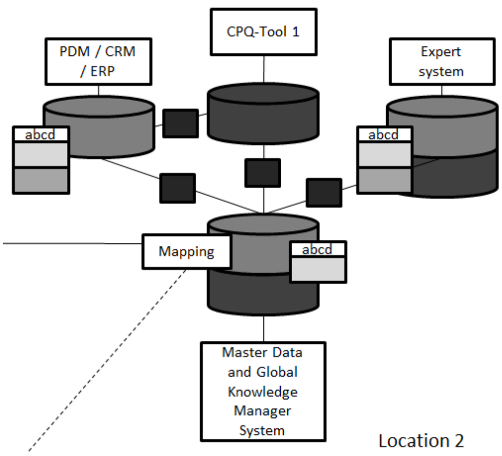

Figure 8. CPQ-tool – expert system architecture

These two simplified examples are only supposed to give an overview about the working-principle to be established with the introduction of a GKM and need to be further researched in following publications. But since the vision formulated at the beginning of this paper also includes recommendation mechanisms and possibilities to implement self-learning algorithms the proposed system architecture needs to be accomplished by continuative considerations.

6 CONSTRAINT-BASED RECOMMENDER SYSTEMS

Recommender systems are since the 1990s an increasingly used service in mainly e-commerce applications to recommend simple products to users [23]. The user gets recommended products based on previously taken buy decisions (individual information), social background information and a knowledge base that proceeds given user input in form of determined attribute and using concerned domain and contextual knowledge. But according to Felfernigs et al. definitions recommender systems could be used in a much more general way: 'Any system that guides a user in a personalized way to interesting or useful objects in a large space of possible options or that produces such objects as output.' [24].

Besides the possibilities given for e-commerce sellers, a second very large field of application opens with the usage in expert systems. Mainly product configuration applications are adopting recommendation techniques which can be classified in:

1. collaborative recommendation – relying on the choices made previously by other users with the same social and demographical background
2. content-based recommendation – relying on choices made previously by the user herself
3. knowledge based recommendation – relying on user requirements and domain knowledge
4. hybrid forms – try to combine the other types to avoid disadvantages of each stand-alone solution

With respect to the high voltage transmission business the only reasonable choice for recommender systems is the knowledge based recommendation technique since electric power transmission substations do not rely on choices or preferences of single users but on very detailed requirement specifications delivered by the customers (utilities). Even if the requirement specifications of utilities do not change frequently and a content-based recommendation could be possible in terms of parameter input to a configuration application an important circumstance, a unique feature of substations has to be taken into account – the environmental conditions. Unlike other typically configurable products like cars or kitchens substations are crucially impacted by the locations they are assembled.

For example, the Russian ministry of electricity announces in a tender a high voltage transmission substation with the same topology (electrical circuit diagram) and electrical requirements as five years before. One might conclude that the same electrical requirements lead to the same bill of material and same physical topology of the substation. But taking the new location near the Baltic Sea compared to the previous one in Novosibirsk into account this paradigm is false. The environmental conditions including temperature, air pollution, earthquake requirements and altitude of side have an essential impact on the physical arrangement of the substation. This set of conditions to be considered can be depicted as a very complex conglomeration of constraints and is in conclusion applicable for recommendation systems. However, the complexity of the configuration task is probably too high to be provided with all features in one product configuration application which is why the example in the previous chapter has been chosen.

To achieve the best configuration solution many specialized systems have to give input to the configuration process. Mechanical calculations, waste heat, ferroresonance and further calculation applications are delivering valuable information and need to be addressed to receive a feasible solution. The expert systems outcome is involved as input for the existing recommendation constraints to create a solution for the given configuration task. The user of CPQ-tool 1 in the example illustrated for Figure 7 as a non-expert gets one feasible solution recommended without any repair mechanism. The solution will be rather conservative, but is in line with the given rules and constraints. On the other hand the expert using CPQ-tool 2 in the example has the possibility to neglect the recommended solution and choose a more efficient one. The changes are recognized by the system and adapted to the existing constraints in the GKM. A self-learning system as formulated in the vision at the beginning of this paper is created using expert knowledge to continuously improve the general knowledge manager as a basis for all product configurators. Basis for this mechanism is the system architecture and universally understandable format to provide communication and non-redundancy between all parts of the system as proposed in this paper.

7 CONCLUSION AND OUTLOOK

The paper discussed the currently arising challenges and opportunities given in the high voltage transmission business and highlighted the current situation of the information technology in the business. Insulated solutions due to complexity of the products and the world-wide distribution of the associated business segments lead to redundant data storage and the development of several different product configuration solutions. On the basis of the state of the art methods to normalize data and to integrate master data solutions into the system, considerations concerning the resulting system architecture have been made. Furthermore, the need to create a non-redundant knowledge set has been emphasized in addition to a normalized and integrated data set. Deducted from this condition, the proposed system architecture was extended by introducing a global knowledge manager on the same level as the master data system to create a framework for configuration, pricing and quotation rules and constraints. Additionally, a universally understandable format was proposed to enable communication between all parts of the system and to create the possibility to integrate expert systems via inquiries scheduled and processed by the global knowledge manager. The working principle of this system was outlined by two examples. Finally, the opportunity to extend the system by recommendation techniques and self-learning algorithms was pointed out.

Since this paper was supposed to give only a first overview over the problem statement, several future prospects arise from the evaluations presented. These future prospects are the following:

- Data integrity – how can product configurators be enabled to use data from centrally maintained databases instead of using encapsulated exclusive data sets resulting in redundancy?
- Knowledge integrity – how should knowledge normalization methods look like to build the base for a global knowledge manager and a framework for rules and constraints?
- System architecture – which system architecture allows the highest performance, best maintainability, and security with respect to the requirements given by a fully integrated knowledge system?
- Global knowledge manager – how should the global knowledge manager be designed?
- Universally understandable format – how should a format look like which is processible by all databases and programs in the system, including expert systems and product configurators?
- Recommender system – how can rules and constraints be improved by analyzing decisions of experts in a product configurator to improve the recommended solutions to non-experts?

REFERENCES

[1] European Commission, *Energy - Country datasheets,* https://ec.europa.eu/energy/en/data-analysis/country, (2018)

[2] Katrin Schaber, Florian Steinke, Pascal Mühlich and Thomas Hamacher, *Parametric study of variable renewable energy integration in Europe: Advantages and costs of transmission grid extensions,* Energy Policy, Volume 42, 498 – 508, (2012)

[3] Mathias Uta, *Development of a product configurator for highly modular and modifiable products,* Figure 1, (2017)

[4] Siemens AG, *High-Voltage Products,* https://www.siemens.com/content/dam/internet/siemens-com/global/products-services/energy/high-voltage/high-voltage-switchgear-and-devices, 2016

[5] ABB, High Voltage Products – Business snapshot, http://new.abb.com/high-voltage, (2017)

[6] Graeme C. Simson and Graham C. Witt, *Data Modeling Essentials*, Third Edition, p. 10, (2005)

[7] E. F. Codd, *A Relational Model of Data for Large Shared Data Banks*, Communications of the ACM., 13 (6), 377–387, (1970)

[8] Alan Radding, *So what the Hell is ODBMS?*, Computerworld. 29 (45), p. 121–122, (1995)

[9] Frank Manola, *An Evaluation of Object-Oriented DBMS Developments*, GTE Laboratories technical report TR-0263-08-94-165, (1994)

[10] Ramaknth S. Devrakonda, *Object-relational database systems – The road ahead*, Crossroad magazine, Volume 7, p. 15-8, (2001)

[11] Michael Stonebraker, *Object-Relational DBMSs: The Next Great Wave*, Morgan Kaufmann Publishers, p. 12, (1996)

[12] Alejandro Vaisman and Esteban Zimányi, *Data warehouse systems*, Springer, (2014)

[13] ISO 8000-2:2012, *Data Quality*, Part 2: Vocabulary, (2012)

[14] Cornel Loser, Christine Legner and Dimitrios Gizanis, *Master Data Management For Collaborative Service Processes*, International Conference on Service Systems and Service Management, (2004)

[15] Habibullah Jamal and Kiran Sultan, *Performance Analysis of TCP Congestion Control Algorithms*, International Journal of Computers and Communications, Issue 1, Volume 2, (2008)

[16] William H. Inmon, *Building the Data Warehouse*, John Wiley & Sons, p. 31, (1996)

[17] Robert R. Hoffman, *The Problem of Extracting the Knowledge of Experts from the Perspective of Experimental Psychology*, AI Magazine, Summer Edition, p. 53-67, (1987)

[18] S. M. Davis, *Future Perfect: Mass customizing*, Addison-Wesley, (1987)

[19] Aparajita Suman, *From knowledge abstraction to management*, Woodhead, p 87-109, (2014)

[20] Alexander Felfernig, Lothar Hotz, Claire Bagley and Juha Tiihonen, *Knowledge-Based Configuration – From Research to Business Cases*, Elsevier Inc., Chapter 7 and 12, (2014)

[21] Frank Manola and Michael L. Broadie, *On Knowledge-Base System Architectures*, On Knowledge Base Management Systems, Springer, p. 35-54, (1986)

[22] Henning Grempel and Gerald Kopatsch, *Schaltanlagen Handbuch*, Cornelsen, 11th Edition, p. 603, (2008)

[23] Michael D. Ekstrand, John T. Riedl and Joseph A. Konstan, *Collaborative filtering recommender systems*, Foundations and Trends in Human-Computer Interaction, Vol. 4, p.81-173, (2011)

[24] A. Felfernig and R. Burke, *Constraint-based recommender systems: technologies and research issues*, In Proceedings of the 10th International Conference of Persuasive Technologies, ICEC '08, ACM, p.1, (2008)