An approach for ontology-based research and recommendation on systems engineering projects

Mouna El Alaoui, Vincent Chapurlat, Souad Rabah, V. Richet, R. Plana

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 225 (2023) 1350–1359

Procedia Computer Science

www.elsevier.com/locate/procedia

27th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2023)

An approach for ontology-based research and recommendation on systems engineering projects

M. El Alaoui[a]*, V. Chapurlat[a], S. Rabah[a], V. Richet[b], R. Plana[c]

[a]*Laboratory for the science of risks, IMT Mines Ales, Ales, France*
[b]*Assystem STUP, 77A, Shaheed Ripon Katyal Marg, East Block, Sector 18, Gurugram, Haryana 122002*
[c]*Assystem Energy and Operation Services (AEOS), Courbevoie, France*

Abstract

Recommendation systems (RS) have dramatically evolved these past few years, generating more and more accurate results and developing innovative filtering methods. A huge breakthrough was the integration of ontologies into their recommendation process. It integrated the domain's knowledge into its reasoning process, thus overcoming the limitations of the conventional RS. The main purpose of this paper is to explore the application of ontology-based RS within critical infrastructure projects implementing the systems engineering or model-based engineering approach.

Peer-review under responsibility of the scientific committee of the 27th International Conference on Knowledge Based and Intelligent Information and Engineering Systems

Keywords: Recommendation systems; ontology-based recommendation systems; knowledge-based recommendation system; ontology-based information retrieval systems; systems engineering, model-based systems engineering

1. Introduction

In the recent years, critical infrastructures' engineering projects such as nuclear facilities engineering projects have been progressively adopting the Systems Engineering (SE) approach. SE promotes processes-oriented organization (ISO 15288:2015) and advocates the role and importance of modelling and giving sense of a model-based systems engineering (MBSE) approach. According to the SE Handbook (INCOSE, 2015) : "*SE emerged as an effective way to*

* Corresponding author. Tel.: (+33) 0760305590.
E-mail address: mouna.el-alaoui@mines-ales.fr

1877-0509
Peer-review under responsibility of the scientific committee of the 27th International Conference on Knowledge Based and Intelligent Information and Engineering Systems
10.1016/j.procs.2023.10.123

manage complexity and change. As both complexity and change continue to escalate in products, services, and society, reducing the risk associated with new systems or modifications to complex systems continues to be a primary goal of the systems engineer". The nuclear domain found a great interest in SE justified by the domain's key characteristic: a significant level of complexity to consider and manage. Due to their nature and their stringent safety, security, environmental, and dismantling requirements, a significant part of nuclear projects are 'megaprojects'. 'Megaprojects' refer to projects that exceed $1 billion budgets and involve a significant level of complexity and innovation (G. Locatelli, 2018). More particularly, these projects are considerably long (in the order of decades) and must take into consideration various and evolving infrastructures complexity factors and critical needs. This complexity is particularly reflected by the set of requested and handled data shared and managed by projects' stakeholders. Thus, this article focuses on data, information and knowledge management, including storing, structuring, accessing, filtering and even retrieving and sharing, of larger and larger, evolutive, heterogenous and even questionable amount of data all along projects. As an example, a common saying is that in a project, 80% of required data, information and/or knowledge already exist somewhere, reflecting the experience and the culture of the enterprise. Therefore, it has to be accessible, reusable or at least it has to constitute a source of inspiration for the project's stakeholders based on their relevance and quality.

Therefore, to be the most effective, industrials should focus on developing methods that aim to gain as much time as possible during the process of research, retrieval and reuse of Data, Information, and Knowledge (DIK). This paper presents and illustrates the so called DIK approach in the case of a Nuclear Infrastructure. It starts by presenting the observed issues and problems that motivated the identified research questions. It then proposes a state-of-the-art in accordance with the announced field of work. The contribution of this paper is then presented and part of it is illustrated through two recent nuclear projects before concluding.

2. Problematic statement, research question and proposed contributions

Data, Information, and Knowledge (named DIK) constitute more than ever the main challenges that the industries are facing. In 2006, Clive Humby, a UK mathematician and data scientist once said in an interview: "*Data is the new oil. It's valuable, but if unrefined it cannot really be used*.". It is consequently crucial to acknowledge the difference between DIK in order to manage each of them in the most appropriate way. "*Data can be defined as a sequence of raw unprocessed items (i.e., signals, codes…), that haven't been processed to make sense of what it represents. On the other hand, information can be considered as a 'processed data' or data that has been interpreted in a given context to make sense of it"* (El Alaoui et al., 2022). For instance, (Terra & Angeloni, 2003) gathered 3 main definitions of information: "*Organized data*", "*Data endowed with relevance and purpos*e", and "*Interpreted data*". Knowledge's definition however is much more challenging, it mainly relies on human cognitive capacities which makes it a concept that it hardly captured, represented, evaluated, shared, and more generally managed. (Terra & Angeloni, 2003) concluded that "*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*".

These DIK are hereafter characterized as synthetized in (El Alaoui et al., 2022) through the set of essential properties (5V): Volume, Variety, Variability, Velocity, and Veracity. The 5V-related issues are particularly presented and observed in large engineering projects that involve critical infrastructures due to their high complexity and their important needs in terms of safety and project performance – any error or delay can cost millions of dollars. Moreover, the considerably long projects' duration only adds to the complexity of DIK management, since the project' stakeholders might leave the project before its schedule ends. Therefore efforts must be deployed to retrieve, recover, formalize, and reuse the deliverables, progress, knowledge, and expertise from these actors. Another difficulty specific to those projects is precisely the large number of involved, concerned, and/or impacted stakeholders, since each of them has his own perspectives, skills, habits, practices, objectives, and roles/responsibilities. They will have different needs in terms of DIK in order to be able to conduct and successfully complete their assigned activities and report and justify their results and deliverables.

Furthermore, the stakeholders' activities, especially V&V-related, require complete traceability of the existing - static and dynamic - dependencies and correlation relationships between the DIK. To conduct these activities, the stakeholders also need to effectively identify and compare the DIK that are adapted to the processes and activities, the given moment, the context, the nature and complexity factors of the system of interest, as well as the maturity and reliability level of the said DIK. Finally, it is crucial to consider the need to filter, structure, recommend, and present DIK all whilst considering their 5V characteristics outlined below.

The origin of the DIK-related features presented above can be justified by the challenging interpretation of terms due to their semantic meanings, which are quite subjective and differ for the exact same term based on the context, the interpreter (its background, profile, domain, and current activities) and the moment. It is also widely observed that there's a lack of traceability and accountability on the DIK in terms of the dependencies between them as well as the actions and operations made on them. This can be related to the use of non-interoperable tools either at the same stage or in different stages of the project. Overall, difficulties are observed in the querying and researching of DIK in terms of rapidity and effectiveness due to the insufficiency of DIK structuring solutions resulting from the fact that they're unable to handle their diverse sources and the multitude of types and formats they come in. Furthermore, there are limited means of DIK representation crossing and covering multiple domains and professions in a common way that is accepted by all the stakeholders according to their uses and/or domain. In this context, ontologies were conceived to address this particular issue. This work focuses on data requested for large systems engineering projects driven by SE processes. However, in the SE domain, several ontologies exist and each of them address different aspects, but none actually covers the integrity of the concerned areas in SE. Last, means and methods allowing to assist the stakeholders in DIK selection and tracking remains limited when considering for instance stakeholders' roles, and 5V characteristics of the DIK at the given moment.

The addressed research question is the following: how to proceed (search, retrieve, analyze, categorize, classify, structure efficiently) the available DIK, then suggest and make recommendations to stakeholders about the relevance of these DIK?. Within this scope, this work's main objective is to develop a method that optimizes the organization, management, and exploitation of data, information, and knowledge, in order to guarantee their accessibility during the right stage of the project, for the right activity, to the right role, and in the right operable configuration and format. To this issue, this paper focuses on an operational approach and a gradual construction process of a reusable and generic *DIK Research and Recommendation System*. To this issue, the following state of the art explores some solutions allowing recommendation and research/retrieval of DIK.

3. DIK Research and Recommendation: State of the art

The recommendation process orchestrates many activities of which data retrieval, indexation, organization, comparison, evaluation, and presentation. The first step is to identify the stakeholder's needs according to specific criteria considered in a recommendation model, then to search and retrieve the DIK from the different data sources using a DIK retrieval model avoiding extracting and storing them to prevent duplication and update issues.

Recommendation systems Content recommendation has long been used in business and commerce, the demand of recommendation systems (RS) has particularly increased with the expansion of online shopping and its related explosive growth of the amount of information and number of visitors. RS are able to predict whether a particular user (and therefore a potential customer) would be interested in an item or not based on his profile. Recommendation systems have also proved to improve decision making process and quality. There are several types of RS, they each differ in the recommendation filtering techniques, the main ones are presented below:

- *Content-based filtering,* is based on the analysis of items' attributes/characteristics. In this case the recommendations are made based on the user's profile using features from the previously evaluated items, a matching is then made between them and the similar items to the highest ranked ones by the user are suggested. It has the ability to adapt quickly to the users' change of taste/preferences. This method however knows certain limitations such as the fact that it relies on rich description of items and well-organized user profile before recommendations can be made, also known as the cold-start problem (Isinkaye et al., 2015).
- *Collaborative filtering,* is based on the similarity of taste and/or opinion between the users. The idea is that if two users have shown interest in the same items in the past, they will most likely be interested in the same items later on. The principle of collaborative filtering is the computation of similarity between users, that implies looking for other users with similar ratings to the current target user and suggesting their highly rated items or similar items. (Tarus et al., 2017). The main disadvantage of this method is that it relies on the availability of users' ratings on each item to make recommendations.
- *Ontology-based filtering,* is a type of knowledge-based filtering process that uses ontologies to represent

knowledge concerning the users and the items and bases its recommendations on these ontologies. Studies show that the use of ontologies enhances the accuracy and quality of recommendations and mitigates the difficulties of the conventional recommendation methods such as cold-start problems and rating sparsity problems (Tarus et al., 2018).

Information Retrieval (IR): DIK retrieval consists of indexing the DIK and storing only their index and information about their actual location. There are two main types of IR models: Keyword-based IR models (e.g. Boolean Model), and concept-based IR models.

Keyword-based Information Retrieval Model, these models are mainly used in commercial search engines. They are based on keyword indexing systems and Boolean logic queries and are sometimes enriched with statistical methods such as introducing the frequency of occurrence of a keyword or some proximity constraints (Haav & Lubi, 2001). The keyword-based IR models rely completely on a keywords list to describe the contents of information objects, the problem however is that it does not take into consideration the semantic relationships that may exist between the keywords nor the meaning of the terms or phrases. "*A word for this model is only a sequence of binary codes representing a word*" (Haav & Lubi, 2001). A major disadvantage of such models is that they can be misled by the ambiguity of terms (e.g., homograph) and miss out on the semantic relationships among terms (e.g., synonym, hypernym) (Baziz et al., 2005).

The Boolean model is among the first data recovery models and among the most criticized (James & Kannan, 2017). In this model, several terms are used to represent the data containers, these terms then take a Boolean value for each container, and the queries are expressed using Boolean operators such as AND, OR and NOT. The container is therefore either classified as relevant or irrelevant based on the comparison of the terms associated to it and the terms of the query. The main limitation of this model is its inability to rank the resulting information objects.

- *Concept-based Information Retrieval Model,* are based on general or domain conceptual structures on which the terms are represented, that implies that sets of words, names, noun-phrases, terms, etc. will be mapped to the concepts they encode (Haav & Lubi, 2001). Conceptual structures are created manually or automatically, and include dictionaries, thesauri (Word-net, UMLS) or ontologies (e.g., Gene Ontology) (Sy et al., 2012). The survey of (Haav & Lubi, 2001) presents the main types of conceptual structures used in concept-based information retrieval systems among which is found ontology-based IR models.

Ontology-based information retrieval model, are characterized by their use of ontologies as a concept structuring method for the terms. An ontology is a conceptual representation of the entities, events, and the existing relationships that describe a specific domain. Two main relationships are abstraction (subsumption) and composition ("part-of" relationship) (Haav & Lubi, 2001). Historically, (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*". An ontology is therefore a knowledge representation method widely used to overcome the limitations of keyword-based search and it has been considered one of the motivations of the Semantic Web ever since its emergence in the late 90's. The idea is to use the ontologies that specify the current domain to enhance the understanding of the query and the information objects considered/searched, in order to generate a more precise and relevant search result. At the end, the quality of the results will be directly related to the quality of the ontology, Therefore, important efforts must be deployed in the phase of creation and evaluation of the ontology.

In this specific type of IR models, the ontologies will also be used in the query processing and expansion stage by mapping the query terms to the concepts in the ontology, identifying synonyms and related terms, and resolving any ambiguities or inconsistencies. In this context (Sy et al., 2012) present in their paper a semantic similarity measurement method that improves the query expansion process.

Based on the above, both the recommendation and the information retrieval systems considered in this work are ontology-based due to their below-listed advantages compared to conventional systems. Therefore, an essential component is the ontology built and adapted for the intended use: a SE-oriented ontology.

Ontologies in SE: SE's focus on ontologies goes a long way back. Indeed, many attempts were made to create a SE ontology.. (Yang et al., 2019) in their paper present a state of the art that displays in a table that compares, among other things, the references, the scopes, the key concepts, and the key properties of the existing SE ontologies. The conclusions drawn are that in the SE and MBSE scopes, many attempts were made to come up with a functional common and mutual ontology for SE uses, however, these initiatives were met with the following issues:

- The ontologies are not complete: a quick overview of Yang et al.'s work shows that each specific ontology addresses a certain aspect of SE, implying that none of them actually covers all the concepts and aspects of SE,

- They are not semantically aligned: some ontologies use different designations for the same concepts, most of the time without necessarily establishing a corresponding link to assure the semantical alignment, which goes against the basic principles of ontologies.

4. Contribution: DIK Research and Recommendation System

4.1. Items definition

In the following, the term 'item' is used to describe a semantic object or entity, meaning that it is an object that holds a semantic content which can be subject to an interpretation that may differ depending on the context, the involved discipline, and the interpreter himself and his current state of mind in the given situation and during that specific moment. Those items are by definition tangible; they have a physical and perceptible aspect that can be represented or described by a logical representation. This implies that it has a digital form that can be formalized, generated, used, made accessible and shared between different stakeholders in order to serve the specific objective of irrigating an engineering process through activities, tasks, operations, and/or subprocesses. The items naturally exist in containers that meet accessibility and security specifications and therefore can undergo specific actions of creation, addition, manipulation, processing, and verification & validation actions. Each item is characterized by a DIK status that categorizes it as data, information, or knowledge according to the stakeholder who created it and/or is using it and his current activities, and the degree of maturity and trust that the stakeholder places in the item for a specific use.

4.2. DIK Research & Recommendation System (DIK-RRS)

The mission of DIK Research and Recommendation System (DIK-RRS) is: (i) to proceed a user's research request and return a list of relevant items and their characteristics (e.g. position, relevance, or maturity and credibility level for the stakeholder), and (ii) to make more global recommendations, not request-driven, to the stakeholder based on his/her actual activities and project, roles, needs, and expected deliverables. The DIK-RSS is therefore composed of 3 subsystems shown in Fig. 1, of which a recommendation system and an information retrieval system, both ontology-based to capitalize and integrate the existing knowledge and enhance the results and an evaluation and display systems that ranks and presents the results to the user. The idea is to define and to dispose of 2 kinds of ontologies: (i) domain-specific ontologies, developed and validated by domain experts, and (ii) one generic ontology called 'pivot' that assures the interoperability between all the domain-specific ontologies, through, among other things, integration of the different semantic relationships and the compliance or equivalence between different concepts of the domain-specific ontologies. That way each group of domain experts can work on their specific ontology to represent their appropriated concepts and use the terms they are familiar with, without the constraints of building a unique ontology and its never-ending debates between experts from different domains. This approach implies that efforts must be deployed to make the domain ontologies interoperable by identifying, manually or automatically, the semantic relationships between certain concepts. In this context, some existing contributions such as (He et al., 2018) explore ontology interoperability through semantic alignment and propose operating tools that implement it. During the research and recommendation process of this contribution, first the 'pivot' ontology is consulted in order to reach all the corresponding concepts of the considered terms in the domain-specific ontologies. The query is then expanded and provided to the information retrieval system that interrogates all the information objects.

To conduct its mission and meet its objectives, the DIK-RRS relies on an ontology management system (e.g., Protégé (He et al., 2018)), that browses and manages the ontologies. These ontologies are stored in a project Repository of Expertise and Knowledge (REK) defined by (Bourdon et al., 2023). Two REKs co-exist. The project-REK is dedicated to a specific project. It gathers structures, and gives access to the whole sets of items that are related to this project, managing different types and formats. The Enterprise-REK is unique, it stores, shares and gives access to items that could be shared between various projects-REKs (e.g., best practices, patterns, reference models, etc.) on the enterprise's level. Thus, these items represent a part of the culture of the enterprise. The collection of project-REKs and Enterprise-REK is administrated and managed by, respectively, the project manager and the Information Systems manager.

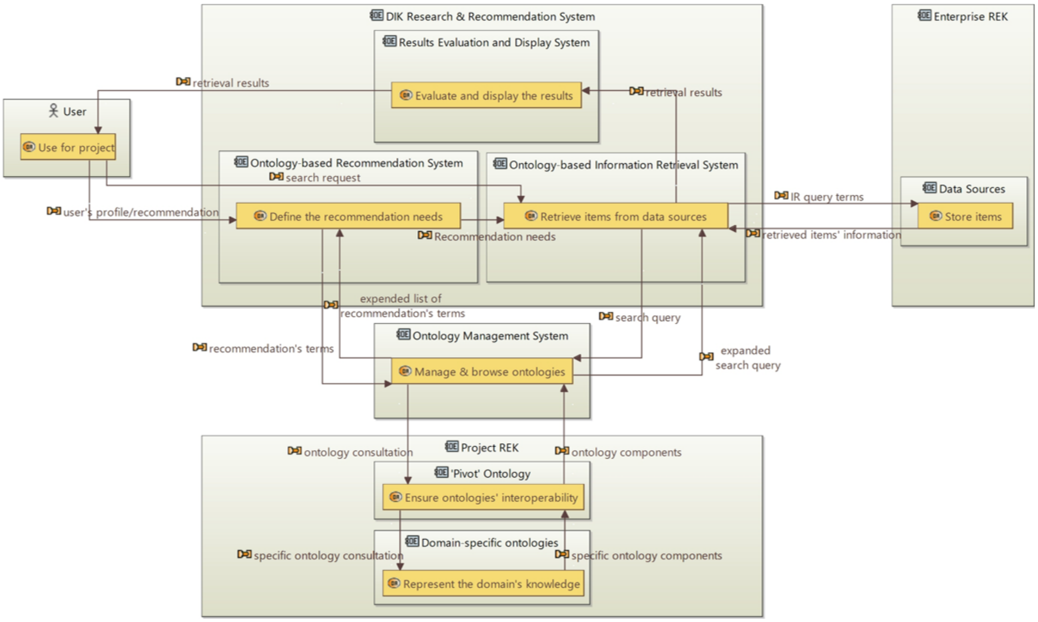

Fig. 1. Representation of the DIK-RRS's operational context.

4.3. The recommendation process

The so-called "global recommendations" are specifically adapted to a user's profile. To achieve that, the system is based on a global assessment of the functional, organizational, and informational sequence of the processes defined (ISO 15288:2015), an extract of this assessment concerning the "*business or mission analysis*" process is shown in table 1. This assessment defines the *activities* and *roles* involved in this process as well as the *documents* and *items* expected as inputs and outputs. This is a generic study done for each of the processes promoted in (ISO 15288:2015).

Activities	***Detailed activities***	***Roles involved***	***Documents (SE Handbook)***		***Items***	
			Inputs	***Outputs***	***Input***	***Output***
- prepare for business or mission analysis - define the problem or opportunity space - characterize the solution space - evaluate alternative solution classes - manage the business or mission analysis	- define the SOI, its mission, objective, scope - define the initial life cycle - identify the stakeholders - develop preliminary operational concepts - distinguish environmental conditions and constraints that bound the solution space	- legislator/ regulator - financing / owner -sponsor / program manager - modeler - business analyst - project director - project manager - technical coordinator - technical coordinator - discipline / domain experts - stakeholder representative	- Organization strategic plan - ConOps - Source documents - Life cycle constraints - Project constraints - Stakeholder requirements traceability	- Business or mission analysis strategy - Major stakeholder identification - Preliminary life cycle concepts - Problem or opportunity statement - Business requirements - Alternative solution classes -Preliminary validation criteria - Preliminary MOE needs - Preliminary MOE data - Business requirements traceability - Business or mission analysis record	Context, organization's mission, organization's vision, organization's values, policies, rules and regulations, environment properties (weather...), long-range goals, short-range objectives, projects REX	Mission, purpose, objective, lifecycle (implicit & explicit), stages, milestones, project planning, project stakeholders, system of interest, upper-system, operational context, operational modes, transitions, operational scenarios, functioning principles, behaviour, interfaces,

Table 1. Extract of the informational, organizational, and functional assessment of the ISO15288 processes.

The construction of this table required the confrontation of the different components of the processes, it was obtained by consulting the norm (ISO 15288:2015) and SE Handbook (INCOSE, 2015) in order to achieve a certain complementarity. Then the result was presented to a selected group of stakeholders to validate that it is consistent with what they encounter in their project activities and potentially complete it.

This assessment was realized with the objective of achieving a global view that reveals in detail what happens within each process. The obtained table therefore establishes the links between the roles of the actors involved in the project's processes, the activities that they conduct and the *items* they manipulate and the ones they are expected to produce. These information are essential in the recommendation process, they are the reference that helps identifying the initial needs of the considered stakeholder. This first list of needed items will then be confronted to the available ontologies to expand it as explained below.

At the start of each project, certain information first need to be collected. The project manager is responsible of specifying this information to the system. Within this information are the project name, objectives, domain, and disciplines. The system will then propose to the project manager to choose the processes that are planned in the project from (ISO 15288:2015). The next step consists of allocating an actor to each identified role.

Based-on the allocated roles and activities, and the available assessment, the system identifies the need of each user in terms of *items* inputs and outputs, the next step is to retrieve these items from relevant projects that have the most in common with the user's actual project.

Later, the 'pivot' ontology is consulted with the identified list of required items in order to come up with an expanded list of terms corresponding to the actual need.

Once the required items are properly identified and formulated into a query, the projects at hand are all compared to the project considered based on 9 characteristics:

- *The system of interest (SoI),* it's the collective set of all elements of any system considered by a lifecycle (SEBoK, 2016),
- *The stakeholders,* or organisations involved in the projects,
- *The SoI's objectives,* in terms of a technical, financial, human, environmental, safety, and performance point of view,
- *The project's objectives,* in terms of a technical, financial, human, environmental, safety, and performance point of view,
- *The domain,* in which the project evolves (e.g., nuclear, defence…),
- *The disciplines* involved and constituting the project (e.g., maintenance, decommissioning…),
- *The professions*, operating in the project (e.g., safety engineer, finance expert…),
- *The processes*, as defined in (ISO 15288:2015),
- *The deliverables*, expected of the project (e.g., report, bill of materials…).

These characteristics are used to calculate the similarity degree (σ) between two projects, this similarity degree is a score that compares two elements based on the 9 above-presented characteristics. It is the sum of the product of each characteristic's similarity score (S) and its importance score (I), as shown in eq. (1). The characteristics have been assigned importance weights (I) at their definition stage, following a consultation with a representative group of people. These importance weights are static and common for all the comparisons and represent the importance of the characteristic to the study. The projects are compared two by two, and a similarity score is given to each characteristic based on a semantic comparison and domain knowledge provided by experts. It ranges from 0 to 1, 0 meaning there is no common point between them, and 1 representing two identical characteristics.

$$\sigma = \sum_{k=1}^{9} S_k . I_k \tag{1}$$

Where k identifies the comparison characteristic referred to.

Since the sum of the importance weights is equivalent to 1, the similarity degree (σ) will range from 0 to 1 following the same principle as the characteristics' similarity scores.

From there, the most similar projects are selected, and the required items extracted from them. The final step is the evaluation of the results and display them to the user.

5. Description use case

Two projects are presented hereafter to illustrate the contribution.

Spare Parts project (SP), concerns nuclear power plants' spare parts (SP) management. The objective was to manage these SP all along their lifecycle, from procurement to on-site commissioning, it therefore concerns 3 major processes:

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

At the time, the number of registered SP in the catalogs was over 200.000, each one featuring several characteristics that need to be monitored and kept up to date. Managing these parts required a lot of efforts, resources and therefore budget, however it was essential; due to the critical nature of the infrastructures, their maintenance is closely monitored and is constantly a topic of high interest. the goal is to ensure the availability of the right parts at the right time and place, to prevent unnecessary delays, whilst also avoiding the storage of obsolete and/or excess SP, which can be very costly. This implies an effective management of the SP and a reliable quality and accuracy of the information concerning them.

The implemented approach consisted of assessing the integrity of the SP's lifecycle processes, through existing documents and experts' interviews, and representing it through models that played an important role in facilitating and structuring the communication between the different actors of the project. These operations mobilized, for nearly 6 months, experts of two distinct yet collaborative organizations, two modelers that were each attached to one of these organizations in order to streamline the exchanges, a modeling expert that was responsible of the coordination and consistency of the produced models, and a project manager.

The resulting digital model described (i) the actions realized on the SP, (ii) the actors performing these actions and their respective organization, and (iii) the interactions between them. In the short-term, having a global representation of such a complex and rich process helped identify several improvement opportunities and practices to optimize the SP management.

Moreover, these models were later used as a base to conduct a data survey study. The objective was to fix some entry data, that was misspelled during the manual entry. First a data model was initialized to represent how the data is organized, stored, and manipulated. Then these representations were used to properly configure and implement an IR algorithm. This algorithm generates ranked propositions of the required attribute of the part based on its other characteristics by going through the available data sources and reasoning on the frequency of the propositions.

The used algorithm is a keyword-based IR method. Using a considerable amount of input and training data from the data sources, the algorithm was successfully implemented and generated estimations along with their percentage of reliability. However, these percentages can be improved by integrating the domain's knowledge to the IR process. In analogy with this paper's presented approach (fig. 1), a keyword-based IRS is used instead of an ontology-based one, and instead of using a recommendation system, the experts define themselves the terms to be searched based on the SP models (replacing the ontology in the presented approach). Therefore, building an ontology is very time-consuming and didn't fit into the project's timeline. It is however an opportunity to explore, and the generated models will constitute a strong base to initiate the ontologies' construction.

Considering the data complexity of this project, the prior global representation of SP has saved time and efforts in the project and domain's acquaintance and comprehension. Therefore, more efforts were deployed in the deployment of the solution and its optimization.

SIMOM project, aims to achieve an "augmented" maintenance by creating an intelligent system that supports the maintenance operations of nuclear power plants. In other words, this system is expected to integrate and support all the aspects of the maintenance, go beyond the mere operations, and therefore develop what would be called "maintenance 2.0". in the nuclear domain, maintenance is essential to maintain the installations operational and safe during their defined operating time. However, due to the high complexity of the systems, the nuclear maintenance is a rich and elaborated process; each aspect is a major work focus managed and carried by an expert team.

A preliminary step of the project was to properly identify these aspects and clearly define the modeling strategy that would help to determine the SoI's perimeter and tame the complexity of the subject. There were 5 identified main maintenance aspects were: spare parts management, HR management and planification, budget management, material resources management and planification, and technical operations planification. The primary basis of this work was the available documents (including the SP project's models) and experts' interviews, and the adopted strategy consisted of addressing the modeling through 4 intertwined levels:

- Level 1: high-level representation of the different types of maintenance considered (predictive, corrective, conditional and components/systems modification),
- Level 2: representation of the different processes that compose the components of level 1 including the actors, the elementary activities and actions performed, the manipulated data and the used systems. These processes are organized as: upstream planification, performance of the maintenance operations, post intervention,
- Level 3: representation of the difficulties encountered during the level 2 processes, highlighting the nature of the problem as well as the location of its occurrence according to the previous levels processes,
- Level 4: representation of the different solutions developed by the enterprise to address the difficulties identified in the previous level with the intention to find a way to make them interoperable and collaborative.

The choice was carried out on a modeling tool that would facilitate the implementation of links/relationships between elements created in different levels, and this project's work demonstrated that traceability is the greatest weapon to face data complexity. The adopted procedure presented above helped converge relatively faster and more efficiently and ensure to provide a thorough perimeter delimitation and description.

Moreover, through the realized models, the main objective was to effectively connect the existing solutions and tools. These models describe, among others, the data models of the work environment, and contribute to prepare the data sources to improve the retrieval of *items* from them for the contribution by creating a suitable environment for the execution of the IR part of the process. The extensive content of the models also provides a strong basis to initiate ontology development work to extend and apply the entire suggested approach of the DIK-RRS.

Fig. 2 illustrates the application of the projects' comparison method presented in §4.3 to the SP project and SIMOM. in this table, each project's characteristics are extracted, compared two by two, and a similarity score is allocated for each characteristic.

Project Characteristics	SP Project	SIMOM Project	similarity score (S)	Importance (I)
SOI	SP management system	Maintenance 2.0 Management System	0,75	0,05
Stakeholders	Assystem, EdF	Assystem, EdF	1	0,1
SOI Obj.	T: master and manage nuclear maintenance PDRs F: control and manage stocks and requests H: X EX S: Ensure the quality of the PDR P:X	T: master, encompass and manage all aspects of nuclear maintenance F: plan, control and represent the financial aspect (budget) H: plan, control and represent HR management E: represent and analyze the environmental impact S: Ensure the quality of maintenance operations	0,75	0,1
Project Obj.	T: represent and map the nuclear maintenance PDR management process F: Guarantee the maintenance completion time - Optimize inventory management H: Respect the HR schedule EX S: ensure the good operation of the equipment P: guarantee the availability of equipment	T: represent and map all aspects of nuclear maintenance F: guarantee of maintenance completion time H: guarantee the availability and competence of HR E: Minimize the environmental impact of maintenance operations S: ensure the proper functioning of nuclear facilities P: guarantee the availability of nuclear	0,75	0,05
Domain	Nuclear	Nuclear	1	0,1
Disciplines	Maintenance	Maintenance	1	0,1
Professions	project manager, modeller, Maintenance eng. , Supply eng. ,	project manager, modeller, Maintenance eng. , Supply eng. , work planning expert, logistics eng., Finance dir., HR management,	0,75	0,2
Processes	business or mission analysis process - stakeholder needs & requirements definition process - architecture definition process	business or mission analysis process - stakeholder needs & requirements definition process - architecture definition process	1	0,2
delivrables	CR - model (Capella)	CR - model (Enterprise Architect)	0,75	0,1

Fig. 2. SIMOM and SP project's comparison.

Following the equation (1), the similarity degree between SIMOM and the SP project is 0,875/1. It is a considerably high similarity score that shows that these two projects have a lot in common and that it would make sense to present each project's *items* to the other project's stakeholders. In the case of multiple projects, as it is the case in enterprises, each of them will be compared to the current stakeholder's project. The projects will be ranked based on their similarity scores and the most "similar" projects will be selected and their *items* that correspond to the stakeholder's need will be presented to him/her.

6. Conclusion and perspectives

This paper presents a proposition of a data, information, and knowledge research and recommendation system. That system identifies the user's currents needs in terms of *items*, based on several parameters described in §5. It then explores the available ontologies to expand its identified need and integrate the represented knowledge in order to integrate and consider all the relevant terms (e.g., synonyms…). An IR system takes over to browse the available data sources, extract the relevant elements, and hand them over to an evaluation and presentation system that ranks the results and provides them to the user.

This DIK-RRS is expected to generate numerous benefits for the project, the collaborators, and the enterprises, starting by a significant improvement of the collaboration between the different stakeholders in terms of quality and efficiency, facilitating the DIK exchange and sharing. And most importantly, it could generate a real gain of time especially in the data research and retrieval process leading directly to a productivity and performance increase.

The upcoming focus of this work concerns the resulting items evaluation, the question that arises is, what are the criteria of *items* evaluation that consider their 5V characteristics to effectively identify the most relevant elements?

Another interesting prospect of this work is the potential integration of machine learning methods and technics to improve the quality and accuracy of recommendations and identify good practices that can be re-used in other projects.

References

Baziz, M., Boughanem, M., Pasi, G., & Prade, H. (2005). A fuzzy set approach to concept-based information retrieval.

Bourdon J., Couturier P., Chapurlat V., Plana R., Richet V., Baudouin B., Model-Based Architecting Evaluation Method for the Delivery of Complex Nuclear Projects, International Journal Systems, INCOSE Ed., 2023 (In Press)

El Alaoui, M., Rabah, S., Chapurlat, V., Richet, V., & Plana, R. (2022). An original Data, Information and Knowledge management approach for model-based engineering projects. IFAC-PapersOnLine, 55(10), 3262–3267. https://doi.org/10.1016/J.IFACOL.2022.10.135

G. Locatelli, 2018, "Why are Megaprojects, Including Nuclear Power Plants, Delivered Overbudget and Late? Reasons and Remedies" , Report MIT-ANP-TR-172, Center for Advanced Nuclear Energy Systems (CANES), Massachusetts Institute of Technology.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. Knowledge Acquisition, 5(2), 199–220. https://doi.org/10.1006/KNAC.1993.1008

Haav, H. M., & Lubi, T. L. (2001). A Survey of Concept-based Information Retrieval Tools on the Web. Proc. of 5th East-European Conference ADBIS, 2.

He, Y., Xiang, Z., Zheng, J., Lin, Y., Overton, J. A., & Ong, E. (2018). The eXtensible ontology development (XOD) principles and tool implementation to support ontology interoperability. Journal of biomedical semantics, 9, 1-10.

INCOSE. 2015. INCOSE Systems Engineering Handbook: A Guide for System Life Cycle Processes and Activities, 4th Edition. Edited by D. D. Walden, G. J. Roedler, K. J. Forsberg, R. D. Hamelin, and T. M. Shortell. Hoboken, US-NJ: Wiley.

INCOSE, B. E. B. (2016). Guide to the Systems Engineering Body of Knowledge SEBoK 2.2. In Guide to the Systems Engineering Body of Knowledge (SEBoK) (Issue May). http://g2sebok.incose.org/app/mss/menu/index.cfm

Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. In Egyptian Informatics Journal (Vol. 16, Issue 3). https://doi.org/10.1016/j.eij.2015.06.005

ISO, IEC, & IEEE. (2015). ISO/IEC/IEEE 15288:2015 Systems engineering—System life cycle processes. International Organization for Standardization, Geneva, Switzerland.

James, N. T., & Kannan, R. (2017). A Survey on Information Retrieval Models, Techniques and Applications. International Journal of Advanced Research in Computer Science and Software Engineering, 7(7). https://doi.org/10.23956/ijarcsse.v7i7.90

Sy, M. F., Ranwez, S., Montmain, J., Regnault, A., Crampes, M., & Ranwez, V. (2012). User centered and ontology based information retrieval system for life sciences. BMC Bioinformatics, 13 Suppl 1. https://doi.org/10.1186/1471-2105-13-s1-s4

Tarus, J. K., Niu, Z., & Mustafa, G. (2018). Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning. Artificial Intelligence Review, 50(1). https://doi.org/10.1007/s10462-017-9539-5

Tarus, J. K., Niu, Z., & Yousif, A. (2017). A hybrid knowledge-based recommender system for e-learning based on ontology and sequential pattern mining. Future Generation Computer Systems, 72. https://doi.org/10.1016/j.future.2017.02.049

Terra, J. C., & Angeloni, T. (2003). Understanding the difference between Information Management and Knowledge Management. KM Advantage.

Yang, L., Cormican, K., & Yu, M. (2019). Ontology-based systems engineering: A state-of-the-art review. In Computers in Industry (Vol. 111, pp. 148–171). Elsevier B.V. https://doi.org/10.1016/j.compind.2019.05.003