Orchestrating progression of enterprise data into actionable intelligence using conceptual modeling

Vasile Ionut-Remus Iga[1], Ana-Maria Ghiran[1], Cristina-Claudia Osman[1]

[1]Babes-Bolyai University, Faculty of Economics and Business Administration, str. Th Mihali, nr. 58-60, Cluj-Napoca, Romania

Abstract

Modern enterprises underlie their business decisions on evidences and insights from collected data. Advances in data collection means and predictive models enabled a streamlined answer to many problems, which created a habitual tendency to consider an automated solution to many business decisions. The cost might be to lose traceability between data and business objectives. This paper suggests a solution based on conceptual models to enable any stakeholder (without demanding technical skills) to create a map of the relationships established between low level data and high-level business goals. This map serves not only for visual purposes, but it helps in a rapid identification of the connections that are created between the top enterprise objectives and low level data objects, due to the semantic conversion of the diagrams in a graph based structure provided by the RDF serialization.

Keywords

DIKW pyramid, graph representation, knowledge graphs, conceptual modeling.

1. Introduction

Situated between information science and knowledge management, the DIKW (Data, Information, Knowledge, Wisdom) pyramid [1]-[3]emphasizes the connection between data, information, knowledge and wisdom. Most of the time, DIKW pyramid is represented just as an abstract structure - its building blocks are constructed based on each other in a hierarchical way, with no much elucidation about their role (why it is actually useful to segregate which is simple data or which is knowledge): first, there is *data*, that is the foundation for *information*, then *knowledge* follows, and *wisdom* is right on top. For many years, the DIKW pyramid has been seen as the canon of information systems, knowledge management, information management, according to [4].

Other more recent approaches in knowledge management have simplified the layered architecture and considered two major levels: the data lake level and the insights generated from it (see a survey from Aberdeen Group that showed that companies that implemented a data lake outperformed similar companies by 9% in revenue growth [5]).

Through this paper we defend the need to still consider a four-tier architecture in the sense that it can be more intuitive to express the transformative process that data is taking starting from the pyramid base all the way up to the top, through sequential data composing. To achieve this, a "knowledge manager" should be able to fully orchestrate it, preferably with no or very little programming skills. Our proposal is to employ a modeling tool that can be used to express the "progression" of data as it becomes "actionable intelligence". We rely our proposal on machine readable serializations of the visual diagrams in RDF format [6] that can be stored on graph repositories and can be queried to gain valuable insights. In this way, a manager can have a visual representation of the traceability of the data to address the organization's objective while applying the graph-based discoverability.

BIR-WS 2023: BIR 2023 Workshops and Doctoral Consortium, 22nd International Conference on Perspectives in Business Informatics Research (BIR 2023), September 13-15, 2023, Ascoli Piceno, Italy
ionut_iga@yahoo.com (V.I.R. Iga); anamaria.ghiran@econ.ubbcluj.ro (A. M. Ghiran); cristina.osman@econ.ubbcluj.ro (C. C. Osman)
0000-0001-7890-9386 (A. M. Ghiran); 0000-0002-9706-2915 (C. C. Osman)

We evaluate our solution using some competency questions, as we apply a Knowledge Graph representation for the diagrams depicting the progression of data (and competency questions became the de facto way to evaluate Knowledge Graphs). These are considered a powerful way to assess if an ontology is an appropriate representation for a domain [7], so we use them to demonstrate that our artifact can be used to store the enterprise knowledge. Competency questions are a set of questions in natural language that support the graph repository development by setting up the requirements that can be translated into graph queries to evaluate what results are returned. Samples of these competency questions are provided when we describe the details of our solution and we illustrate a fictious scenario.

In the next section, we start with some arguments in favor of maintaining DIKW hierarchy model, while presenting some adaptations over its traditional model. We describe from a bird's eye view, the main components of the DIKW pyramid and how can we map them in a conceptual model. The next section further details our proposal and identifies the technological requirements that enable its operationalization. Then, we present some related works. The paper ends with conclusions.

2. Proposal overview

Wisdom Model is a model which has the role to capture the flow of the data in order to become wisdom, which is used to achieve an objective. Traditionally, it is described as a hierarchical model composed of rigid building blocks that are aggregated in stages [8].

However, in the current digital environments, as data takes many forms and shapes, without a systematic and a commonly adopted meaning about what is data, information, knowledge, wisdom, we might unintentionally mix up these concepts. Consequently, we might apply them interchangeably, which could raise up barriers in transmitting knowledge/information/data.

In the following, we define the concepts that are the components of the DIKW pyramid:

- Data is a collection of facts that cannot be understood either alone or without a context; often obtained via measurements in different activities.
- Information is a set of data, together with a context applied to them; something that has meaning.
- Knowledge is the result of using/combining information to achieve a specific goal.
- Wisdom is the ability to select knowledge that is consistent with and supportive of a general set of values.

Objective is the top of the pyramid (in many representations of the DIKW pyramid is not even shown) representing how each wisdom is going to be used to achieve that objective.

The relationships that connect all these concepts must be easily traceable, especially for a person that is involved in decision making. We propose the use of a modeling tool that can support not only the human factor through diagrammatic representations but also the automation of data to wisdom transformation through machine readable serializations of the models. Each of the previously described concepts could be mapped to a concept that can be employed in a model that offers an overview of the entire path between the enterprise's abstract strategic objectives and its data.

ADOxx [9] is a metamodeling platform that enables the development of a modeling tool through a metamodeling approach. One only needs to define the classes, attributes and relations that need to be instantiated in models, i.e. to define the Metamodel, and based on this, the platform generates a modeling tool that accommodates the domain-specific modeling language with the newly created concepts.

Figure 1 presents an overview of the Metamodel for the proposed domain-specific terminology, i.e. a modeling language that contains the concepts for describing the DIKW pyramid. The metamodel describes the proposed extension – e.g. concepts grouped in the Wisdom model type. However, we did not include in the metamodel all concepts from existing standardized modeling languages, e.g. BPMN [10] or Working Environment, just those that are reused in our adapted modeling language.

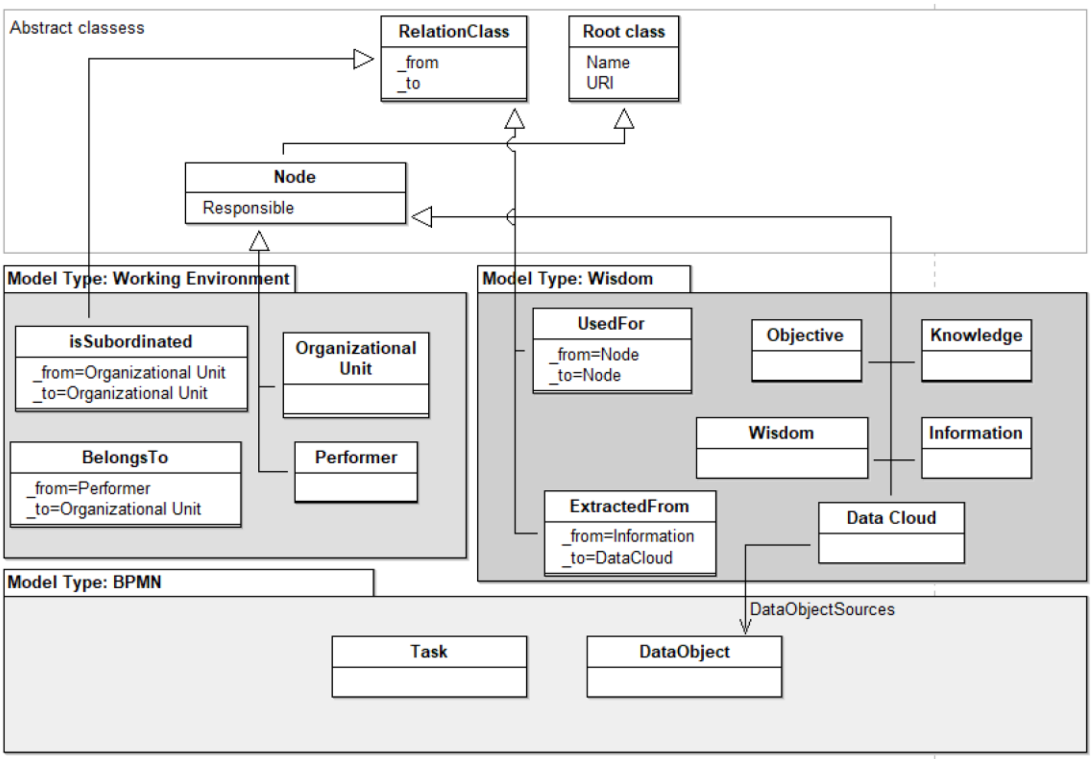

Figure 1: Metamodel for the proposed modeling language

We separated our new concepts into a model type, called Wisdom (as it incorporates the concepts from the wisdom pyramid). Besides the class concepts, there are 2 relation classes: UsedFor and ExtractedFrom and their domain and ranges are set using the _from and _to attributes.

Two other model types are visible on the picture:

- the Working Environment model type (that includes concepts connected with concepts from the Wisdom models; these class concepts are used to describe the organizational structure of the company)
- BPMN model type (it is used to depict process models that describe procedures to achieve a wisdom).

Abstract classes section contains concepts that do not have a visual representation on the modeling tool. Rather, they are used as general concepts for inheritance purposes, to allow the propagation of certain attributes to all the subclasses derived from them, e.g the Name attribute or the URI attribute (we will use this attribute to specify a unique identifier for each object in the models and impose it as an identifier to be used in our graph repository rather than the object' identifier randomly allocated by the modeling tool). Similarly, the Node concept is used as a more general concept from which 2 specialized classes are derived - Organizational Unit and Performer, both will include the "Responsible" attribute.

We also use abstract classes to specify the domain and range for relationships when we want to express a larger scope: e.g. the Node concept is used to express the origin and the target for relation Used for.

3. Proposal refinement

In this section, we detail the proposed metamodel. For each new concept, we have to decide over its notation, syntax and semantics. ADOxx platform allows us to define them gradually, in quick, iterative cycles as it generates a fast prototyping environment after each change in the metamodel. This kind of development for modeling languages and tools is Agile Modeling Method Engineering (AMME) [11] that was first introduced by [12].

Figure 2 presents the graphical notation for the main concepts and relationships used by our modeling language:

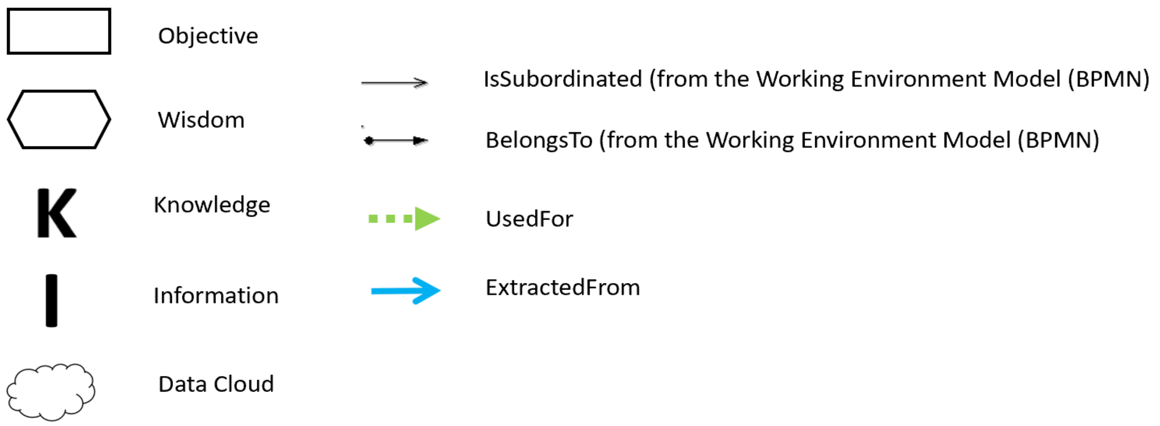

Figure 2 Graphical notation for the proposed concepts

Objective concept is used to present an objective of the business; a target to be reached. It is represented as a simple rectangle and its attributes are:

- Responsible – denoting who is the responsible person for attaining that objective
- Description – a more detailed presentation of the objective

Each objective is achieved by using some Wisdom objects.

Wisdom concept is the ability to select knowledge that is consistent with and supportive of a general set of values. Attributes:

- Model - the linked model to which the wisdom is further detailed
- Responsible – who is responsible (here a whole team/department might be more appropriate, or in some cases a role)
- Description

Each Wisdom element is achieved by combining Knowledge from different sources.

Knowledge concept is the result of using/combining information to achieve a specific goal; it is a validated form of information (information that is purposeful). Attributes:

- Responsible - who is responsible (the mapping should be at role level)
- Description

Each Knowledge is built using different pieces of Information.

Information concept is the result of applying preprocessing operations on data; an understandable form of data. Attributes:

- Responsible
- Description

Each Information is extracted from a data cloud, after applying some preprocessing.

Data Cloud concept is a representation of the group of data stored/used by the company in order to extract information out of it (now or in the future).

Attributes:

- link to a list of data objects from BPMN as source/ an ER diagram/an external link to a database/archive etc.

As relationships, there are two types:

Used for - maps the use of the concept to obtain another concept: wisdom to objective; Knowledge to wisdom; Information to knowledge.

Extracted from - maps the data cloud with the information concept.

The models that can be created with this language are used as *inventories of objectives and wisdom objects* that are needed to achieve them, together with other concepts used to build them – *knowledge, information, data.* This way, the board of a company can observe what are the main areas of focus in terms of information, if they miss something. They can also transform a flow into a know-how etc.

Figure 3 exemplifies an implementation of a scenario in which a fictious company, called Dunder Mifflin, that sells paper supplies, wants to map the achievement of the objective “Increase sales” down to its data sources.

The company wants to capture the process of reaching this objective by highlighting the important steps that each employee of the company has to do. Also, the company is interested to understand the know-hows of the processes that happen every day.

The company has identified the following competency questions:

1. Q1 the list with the objectives of the company
2. Q2 the list of required wisdom objects for each objective, or the list of knowledge associated to a wisdom, the list of information related to a knowledge and so on.
3. Q3 each wisdom could show statistics about the duration, cost (from the associated process model) and number of available employees to do it (by counting the number of employees that have a certain role)
4. Q4 each knowledge will display the name of the employee (s) that knows it (if a role was assigned to the knowledge, and if there is an employee which has that role) and similarly for information or data.

Our fiction company, in order to achieve its objective, requires 3 Wisdom elements (knowledge that contribute or add value in fulfilling the objective): about the Selling Process, the Financial Situation and the Product (figure 3).

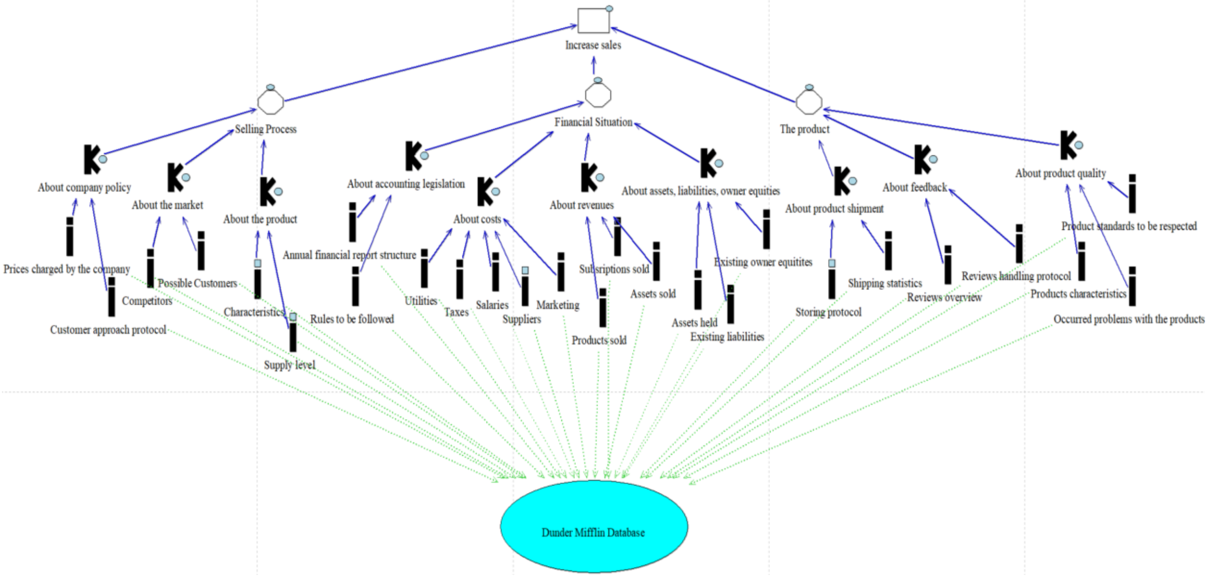

Figure 3 Modeling the Increase sales objective

Then, each Wisdom element is also related to some Knowledge, using the UsedFor relationship.

For some of the Knowledge or Information objects, we can store who provided it, by setting the Responsible attribute with the object from the organizational structure diagram (described later in figure 6).

First, the Selling Process Wisdom element relates to 3 Knowledge entities:

1. the company policies (Salesman as responsible); requires Information about the charged prices by the company and customer approach protocol;
2. the market (Marketing Director as responsible); requires Information about the competitors and possible customers;
3. and the product (Salesman as responsible); requires Information about the characteristics (QA director as responsible) and supply level (Warehouse foreman as responsible).

The Financial Situation Wisdom element relates to 4 Knowledge entities:

1. the Accounting Legislation (Head of Accounting as responsible); requires Information about the annual financial report structure and rules to be followed;
2. the costs (Accountant as responsible); requires Information about utilities, marketing, taxes, salaries and suppliers (for this the Supplier Relations Supervisor is responsible);
3. the revenues (Accountant as responsible); requires Information about products sold, subscriptions sold, assets sold;
4. the assets, liabilities and owner equities; requires Information about assets held, existing liabilities, existing owner equities.

The Product Wisdom element relates to 3 Knowledge objects:

1. product shipment (Warehouse Foreman as responsible); requires Information about Storing protocol and Shipping statistics;
2. feedback (Customer Service Specialist as responsible); requires Information about Reviews overview and Reviews handling protocol;
3. product quality (Quality Assurance Inspector as responsible); requires Information about Product characteristics, Occurred problems with the products and Product standards to be respected.

Information objects are linked to Data Cloud objects through the ExtractedFrom relationship. A Data Cloud object reunites under the Data Object Sources attribute (figure 4) references to the companies' databases:

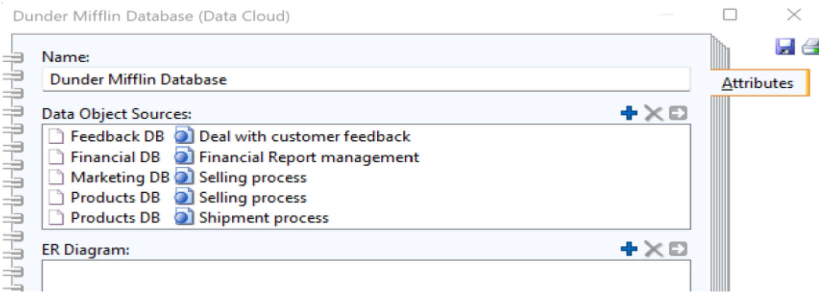

Figure 4 Data source objects in a Data cloud

All the Wisdom concepts are also related to some business processes – i.e. other diagrams (BPMN diagrams) which further describe the know how process:

1. Selling Process is linked to the Selling Process Model, which is described in a separate BPMN diagram (figure 7); the responsible will be the Sales department;
2. Financial Situation is linked to the Financial Situation management; and the responsible will be the Accounting department;
3. Product will be linked to two models, Shipment Process and Deal with customer feedback; and the responsible will be the Logistics and Customer Service departments.

Figure 5 shows how these links are set using the Model attribute. We can see the attributes for one of the Wisdom objects, the Selling Process. We can also see the attributes for the top objective, Increase sales.

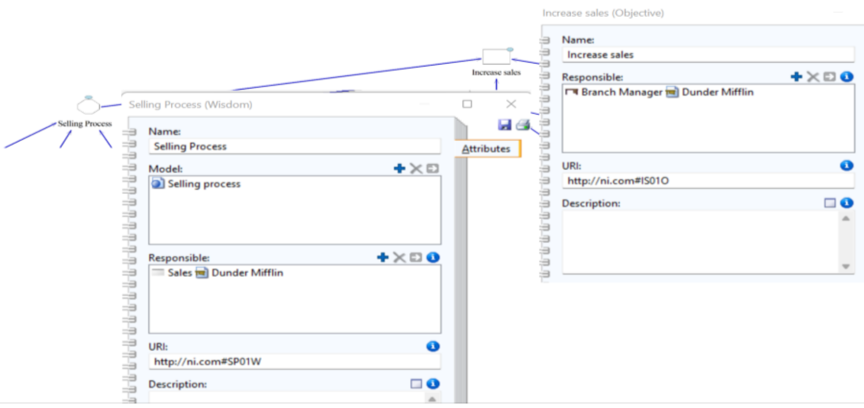

Figure 5 Attributes of the Selling process Wisdom object and of the Increase Sales objective

Figure 6 shows the organizational hierarchy that contains the objects that are assigned to the DIKW concepts.

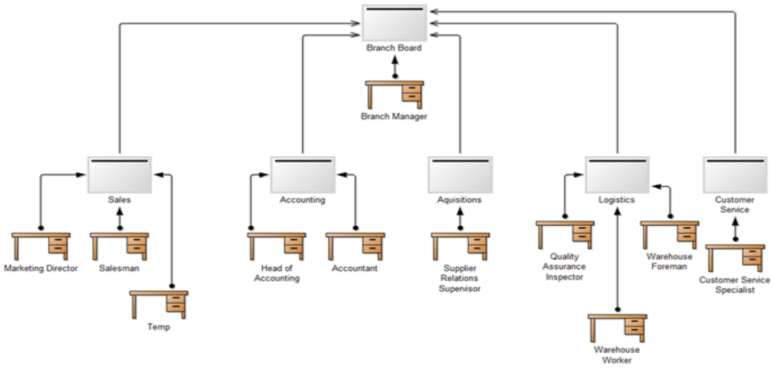

Figure 6 Organizational structure of the company

The Selling Process diagram (figure 7) shows how a sale should happen (it is linked to the Selling Process Wisdom). The Salesman checks the supply to see how much can be offered, i.e., how much is the stock (reading from the Products DB). Then she looks for possible customers (reading from Marketing DB). If the possible customers do not have a contract with another company for a product similar to the one offered by the current company, then an offer is prepared. If they already have a supplier, then the competitors' prices are analyzed (reading from the Marketing DB). A check with the company policy about prices is performed to validate a better offer. If yes, then the offer is initiated, else the selling process is stopped. After the offer is forwarded, the sale can go either successfully or not.

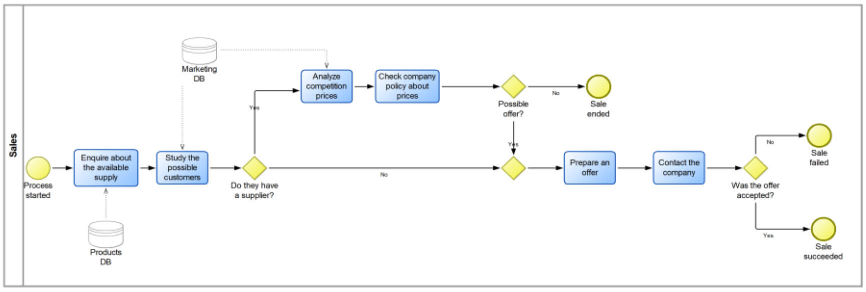

Figure 7 The Selling Process diagram

Other process diagrams (Deal with customer feedback, Financial report management, Shipment process) are not described in the paper.

We can set a unique identifier URI for the entire model: e.g. figure 8 shows the URI value we set for the Selling Process model – e.g. http://ni.com#Sellingprocessv1.

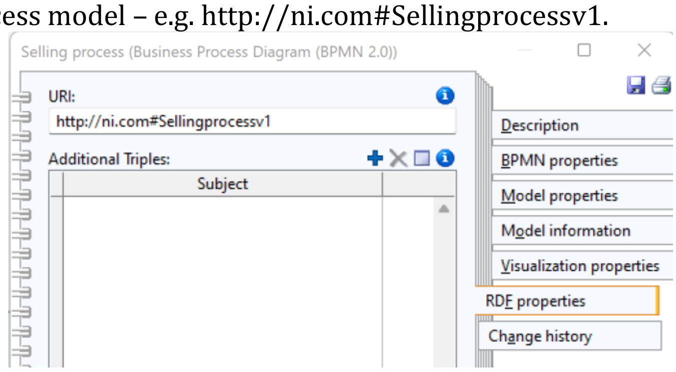

Figure 8 The URI attribute for the Selling Process model

By employing unique URIs to each object in the created models we are able to link them to information that is already in a graph repository, for instance the following type of statements could be from external sources (e.g. they could be gathered by converting information that is in the company's databases):

```
@prefix : <http://ni.com#> .
:Sellingprocessv1    :hasCost        110;
                     :hasDuration    240 .
:MichaelScott        :hasJob         :BranchManager .
```

Mappings between diagrammatic models and RDF graph structures have been discussed in [14] together with transformation patterns to guide implementations [15]. Buchmann and Karagiannis [11] exemplify these mappings in the ComVantage research project.

In order to answer the competency questions, we omit the trivial queries, e.g. Q1. The following query can be used to answer Q2 (Which are the Knowledge objects that are related to a Wisdom?) with results visible in figure 9:

```
PREFIX : <http://ni.com#>
SELECT ?k ?kName where {
    :SP01W ^:Used_For   ?k.
    ?k       a          :Knowledge;
             :Name      ?kName.
}
```

Table | Raw Response | Pivot Table | Google Chart | Download as

Filter query results — Showing results from 1 to 3 of 3. Query took 0.1s, yesterday at 18:51.

	k	kName
1	:ATM01K	"About the market"
2	:ATP01K	"About the product"
3	:ACP01K	"About company policy"

Figure 9 Results of the SPARQL query that retrieves the Knowledge objects

A bit more complex question can extract for a specific Objective all the Wisdom objects that are attached to it but also other external information (like the unit cost, unit duration, the number of available employees, based on the attached responsible role:

```
PREFIX : <http://ni.com#>
    SELECT  ?w ?wName (SUM(?cost) as ?sum)  (MAX((?duration/60)) as ?max) (COUNT(?emp) AS ?nrOfAvlEmp) where{
        { :IS010 ^:Used_For ?w.
         ?w      a          :Wisdom;
                 :Name      ?wName;
                 :Model     ?model.
        ?model  :hasCost    ?cost
        }
        UNION
        { :IS010 ^:Used_For ?w.
         ?w      a          :Wisdom;
                 :Name      ?wName;
                 :Model     ?model.
        ?model  :hasDuration ?duration.
        }
        UNION
        { :IS010 ^:Used_For ?w.
         ?w      a          :Wisdom;
                 :Name      ?wName.
        OPTIONAL {
          ?w    :Responsible/^:Belongs_to/^:hasJob ?emp.}
        }
        }
        GROUP BY ?w ?wName
```

Table | Raw Response | Pivot Table | Google Chart | Download as

Filter query results — Showing results from 1 to 3 of 3. Query took 0.1s, on 2022-06-07 at 22:37.

	w	wName	sum	max	nrOfAvlEmp
1	:TP01W	"The product"	"75"^^xsd:integer	"90"^^xsd:decimal	"4"^^xsd:integer
2	:FS01W	"Financial Situation"	"1500"^^xsd:integer	"1920"^^xsd:decimal	"3"^^xsd:integer
3	:SP01W	"Selling Process"	"110"^^xsd:integer	"4"^^xsd:decimal	"6"^^xsd:integer

Figure 10 Results of the SPARQL query that combines information from the models with external information

4. Related Work

To express the relationships that are created between enterprise data objects and its high level objectives, managers could employ Enterprise Architecture languages like Archimate [16] which includes many concepts that have a standardized meaning and but it requires a considerable time effort to learn them. Through this work we wanted to provide a much simpler and, in the same time, more domain specific alternative.

In the literature, the concepts from DIKW pyramid have been mapped to various resources, but as far as we know, not to conceptual modeling constructs.

Rowley maps different types of systems to each level from the DIKW model [17]. She declares that Expert Systems represent the wisdom, followed by Decision Support Systems (as Knowledge), Management Information Systems (as Information) and Transaction Processing Systems (as Data).

There are some approaches that propose customized graphical visualizations of DIKW pyramid. Chen et al. [18] claim that the knowledge of the user is the essential part of the visualization and proposes different types of visualizations of how data is converted to knowledge by defining information-assisted visualizations and knowledge-assisted visualizations [18]. Ontology mapping and workflow management are only few examples mentioned by [18] as knowledge-assisted visualizations.

Another diagrammatic model is proposed by [19]. They use the DIKW model to highlight the semantic of natural language content and human intention by using UML metamodel of data, information, knowledge and wisdom.

DIKW is also used in several domains like Design Thinking, Information Technology Service Management (ISTM) or Graph databases. Tomita et al. propose an extended version of DIKW as a Structured Design Thinking Framework [20]. Yang et al. use the DIKW pyramid in the context of nutritional epidemiology by building graph database [21].

5. Conclusions

The perception of the DIKW pyramid, its role, purpose and benefits might seem a continuous debate. As with many other models, it is only relevant if it finds an applicability. We believe that having a conceptual model to describe the evolution of data from observations to actionable intentions could bring compelling value for a business. Starting from the very bottom, we can use our modeling tool to add pieces of data (in perspective this could be done automatically using scripts that place the instantiable concepts on the modeling canvas) and composite them into information, knowledge and wisdom. Using conceptual models is more than simple diagrammatic representation of the relationships between objects, it is also a quarriable structure or a guide that drives the execution of systems (in a model-driven fashion).

In this paper, we propose a solution that enables any stakeholder to create a map of the relationships established between low level data and high-level business goals. We employed ADOxx platform to create a customizable modeling language and tool that can be used to express the “progression” of data as it becomes “actionable intelligence”. Our proposal relies on machine readable serializations of the visual diagrams in RDF format that can be stored on graph repositories and can be queried to gain insights. In this way, a manager can have a visual representation of the traceability of the data to address the organization’s objective while applying the graph-based discoverability.

References

[1] R. L. Ackoff, From data to wisdom, Journal of Applied Systems Analysis 16, (1989) 3–9.

[2] M. J. Adler, A Guidebook to Learning: for a Lifelong Pursuit of Wisdom, 1st ed., Macmillan, New York, 1986.

[3] M. Zeleny, Management support systems: towards integrated knowledge management, Human Systems Management 7(1), (1987) 59–70.
[4] M. Frické, The knowledge pyramid: a critique of the DIKW hierarchy. Journal of information science, 35(2), (2009) 131-142.
[5] M. Lock, Angling for Insight in Today's Data Lake, Aberdeen Group, 2017.
[6] W3C Recommendation, RDF Resource Description Framework 1.1 Concepts and Abstract Syntax, 2013. URL: https://www.w3.org/TR/rdf11-concepts/, last accessed 2023/03/13.
[7] C. Bezerra, F. Freitas, F. Santana, Evaluating ontologies with competency questions, In 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), Vol. 3, IEEE, 2013, pp. 284-285. doi: 10.1109/WI-IAT.2013.199
[8] E. Horvitz, T. Mitchell, From data to knowledge to action: A global enabler for the 21st century. arXiv preprint arXiv:2008.00045 (2020)
[9] BOC Group, The ADOxx Metamodeling platform, 2023. URL: https://www.adoxx.org/live/home. last accessed 2023/03/13.
[10] OMG, The BPMN specification page. 2023. URL: http://www.bpmn.org. last accessed 2023/03/13.
[11] R. A. Buchmann, D. Karagiannis, Agile modeling method engineering: lessons learned in the ComVantage research project. In The Practice of Enterprise Modeling: 8th IFIP WG 8.1. Working Conference, PoEM 2015, Valencia, Spain, Springer International Publishing, 2015, pp. 356-373. doi: 10.1007/978-3-319-25897-3_23
[12] D. Karagiannis, Agile Modeling Method Engineering, in: Proceedings of the 19th Panhellenic Conference on Informatics, ACM, 2015. doi: 10.1145/2801948.2802040.
[13] W3C Recommendation, SPARQL 1.1 Query Language, 2013. URL: https://www.w3.org/TR/sparql11-query/, last accessed 2023/03/13.
[14] R. A. Buchmann, D. Karagiannis, Enriching linked data with semantics from domain-specific diagrammatic models. Business & Information Systems Engineering, 58, (2016) 341-353. doi: 10.1007/s12599-016-0445-1
[15] R. A. Buchmann, D. Karagiannis, Pattern-based transformation of diagrammatic conceptual models for semantic enrichment in the Web of Data. Procedia Computer Science, 60, (2015) 150-159. doi: 10.1016/j.procs.2015.08.114
[16] The Open Group, Archimate 3.1 Specification, a Standard of the Open Group, 2023. URL: https://pubs.opengroup.org/architecture/archimate3-doc/ last accessed 2023/03/13
[17] J. Rowley, The wisdom hierarchy: representations of the DIKW hierarchy, Journal of Information Science, 33(2), (2007) 163–180. doi:10.1177/0165551506070706.
[18] M. Chen, D. Ebert, H. Hagen, R. S. Laramee, R. Van Liere, K. L. Ma, D. Silver, Data, information, and knowledge in visualization, IEEE computer graphics and applications, 29(1), (2008) 12-19. doi: 10.1109/MCG.2009.6
[19] Y. Duan, L. Zhan, X. Zhang, Y. Zhang, Formalizing DIKW architecture for modeling security and privacy as typed resources, In: International Conference on Testbeds and Research Infrastructures, Springer, Cham, 2018, pp. 157-168. doi: 10.1007/978-3-030-12971-2_10
[20] Y. Tomita, K. Watanabe, S. Shirasaka, T. Maeno, Applying design thinking in systems engineering process as an extended version of DIKW model, in: INCOSE International Symposium, Vol. 27, No. 1, 2017, pp. 858-870. doi: 10.1002/j.2334-5837.2017.00398.x
[21] C. Yang, B. D. Baets, C. Lachat, From DIKW pyramid to graph database: a tool for machine processing of nutritional epidemiologic research data, IEEE International Conference on Big Data (Big Data), 2019, pp. 5202-5205. doi: 10.1109/BigData47090.2019.9006469.