OntoDIVE: An Ontology for Representing Data Science Initiatives upon Big Data Technologies

Vitor Afonso Pinto[1][a] and Fernando Silva Parreiras[2][b]

[1]*Technology Department, Operational Technology for Mine, Plant and Expedition, Vale Mozambique, Tete, Mozambique*

[2]*Laboratory for Advanced Information Systems, FUMEC University, Rua do Cobre, Belo Horizonte, Brazil*

Keywords: Ontology, Data Science, Big Data.

Abstract: Intending to be more and more data-driven, companies are leveraging data science upon big data initiatives. However, in order to reach a better cost-benefit, it is important for companies to understand all aspects involved in such initiative. The main goal of this research is to provide an ontology that allows to accurately describe data science upon big data. The following research question was addressed: "How can we represent a Initiative of data science upon big data?" To answer this question, we followed Knowledge Meta Processes guidelines from Ontology Engineering Methodology to build an artifact capable of explaining aspects involved in such initiatives. As a result, this study presents OntoDIVE, an ontology to explain interactions between people, processes and technologies in a data science initiative upon big data This study contributes to leverage data science upon big data initiatives, integrating people, processes and technologies. It confirms interdisciplinary nature of data science initiatives and enables organizations to draw parallels between data science results for a particular domain to their own domain. It also helps organizations to choose both frameworks and technologies based on their technical decision only.

1 INTRODUCTION

Data is constantly created, and at an ever-increasing rate. Mobile phones, social media, imaging technologies and several other examples create new data which must be stored somewhere for some purpose (Dietrich et al., 2015). IT technologies have made all devices, equipment, and systems in automation domain intelligent, communicable, and integrated from the field level to the operation level for seamless data flow in both directions. Thus, there are multiple technologies for data acquisition, transmission, storage, modeling and so on. Organizations know that studies that were difficult to conduct in the past time due to data availability can now be carried out (Liu et al., 2016). Organizations are also aware that the timely analysis and monitoring of business processes are essential to identify non–compliant situations and react immediately to those inconsistencies (Vera-Baquero et al., 2016).

Nevertheless, even with all the progress that has been made, companies are still struggling with how to capture insights that are not obvious. It is a problem of how to discover meaningful relationships (Hurwitz et al., 2015). In a general way, organizations have difficulties to leverage big data initiatives as they may not know what exactly is involved in such initiatives. Although this concept has been implemented by many parties, there exists a number of misconceptions related to the concept from the aspect of understanding and implementation of a project like this (Abdullah et al., 2017). The lack of concepts and an increasing list of new technologies creates a fuzzy environment where organizations do not know what they exactly need to do and on the other hand consultants, technology developers, standard publishers and researchers do not know how to help organizations to achieve their goals. This condition limits the usage of advanced analytics tools, preventing the capture of potential benefits.

Main purpose of this study is to describe data science initiatives upon big data by using an ontological approach. Ontologies play a fundamental role in bridging computing and human understanding (Parreiras, 2011). They have been used in several fields as an engineering artifact with the main purpose of conceptualizing a specific object of study. The term Ontology has its origin in philosophy and denotes the philosophical discipline that deals with the nature and

[a] https://orcid.org/0000-0002-2731-0952

[b] https://orcid.org/0000-0002-9832-1501

Pinto, V. and Parreiras, F.
OntoDIVE: An Ontology for Representing Data Science Initiatives upon Big Data Technologies.
DOI: 10.5220/0009416500420051
In *Proceedings of the 22nd International Conference on Enterprise Information Systems (ICEIS 2020) - Volume 1*, pages 42-51
ISBN: 978-989-758-423-7

the organization of reality (Osterwalder, 2004). In this sense, ontology involves identifying the fundamental categories of things (Parreiras, 2011).

Although there exist many definitions of ontologies in the scientific literature, some elements are common to these definitions: a computer ontology is said to be an agreement about a shared, formal, explicit and partial account of a conceptualisation (Spyns et al., 2002). The common vocabulary of an ontology, defining the meaning of terms and their relations, is usually organized in a taxonomy and contains modeling primitives such as concepts, relations, and axioms (Staab and Mädche, 2000). Thus, ontologies constitute formal models of some aspect of the world that may be used for drawing interesting logical conclusions even for large models (Staab et al., 2010). For (Spyns et al., 2002), an ontology contains the vocabulary (terms or labels) and the definition of the concepts and their relationships for a given domain. In many cases, the instances of the application (domain) are included in the ontology as well as domain rules (e.g. identity, mandatoriness, rigidity, etc.) that are implied by the intended meanings of the concepts. The role of ontologies is to capture domain knowledge and provide a commonly agreed upon understanding of a domain (Staab and Mädche, 2000). For (Mizoguchi and Ikeda, 1998), the purpose of ontology engineering is to provide a basis of building models of all things in which computer science is interested.

This study intends to answer the following research question: *"How can we represent a initiative of data science upon big data?"*. By addressing this research question, this study presents OntoDIVE - an ontology to explain interactions between people, processes and technologies in a data science initiative upon big data. This study is structured as follows: Section 2 presents methods of this research. In section 3 research results are presented. These results are discussed in section 4. This study is concluded and future work is suggested in section 5.

2 METHODS

In order to address research question, we followed Knowledge Meta Processes guidelines from Ontology Engineering Methodology, which consists of five main steps: a) Feasibility Study, b) Kickoff, c) Refinement, d) Evaluation and e) Application and Evolution (Sure et al., 2009). For the first step, a group of seven professionals involved with data science upon big data initiatives in a large organization was interviewed in order to capture their thoughts. This group was asked the following questions: a) To what extent would the lack of vocabulary prevent data science initiatives over big data? b) To what extent could an artifact as an ontology leverage data science initiatives over big data? Insights were compiled and a technical specification was generated. Then, software Protégé Desktop version 5.5.0 was used to build an ontology as a representational instrument (Musen, 2015). DL Query plugin was installed on top of Protégé to validate final ontology. This plugin is based on Manchester OWL syntax and presents a user-friendly syntax for OWL DL.

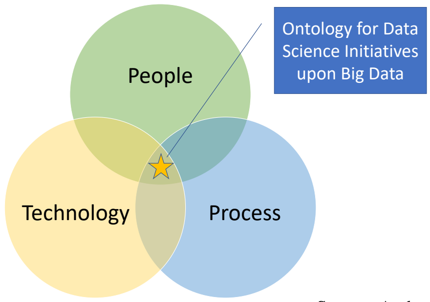
3 RESULTS

In this section we present results for each stage of this research. Subsection 3.1 presents collected insights for problems, opportunities and potential solutions. A summary of ontology requirements is presented in subsection 3.2, including competency questions supposed to be addressed by final ontology. Subsection 3.3 presents general information about the ontology designed to meet identified requirements. Subsection 3.4 presents answers to competency questions raised during requirements gathering stage.

3.1 Feasibility Study

Professionals selected for this study were asked to what extent the lack of vocabulary could prevent data science initiatives over big data. According to them, one of the major problems related to data science upon big data is that companies are still struggling to choose big data technologies and to define which data science framework to choose. Such decisions involve multiple concepts not always dominated by internal personnel and part of consultants have exclusive partnerships with technology developers or standard publishers. For them, creating an ontology capable of integrating People, Processes and Technologies would help to establish a common ground for all parts involved in the whole big data ecosystem.

Respondents were also asked to what extent an ontology could leverage data science initiatives upon big data. According to them, there is an opportunity for any kind of artifact capable of clarification of concepts. They believe business challenges can be addressed by data science applications deployed as final result of data science processes carried out over big data technologies. Nevertheless, they believe such initiatives are performed by people and if everyone involved clearly understands the concepts and termi-



Source: Authors.

Figure 1: Semi-formal Description of the Ontology.

nologies then the dissemination and usage of data science upon big data will be facilitated.

3.2 Kickoff

In a general way, the expected ontology should act as an information provider on the existing relationship between people, processes and technologies in data science initiatives upon big data. A semi-formal description of expected ontology is represented in Figure 1. Next subsections present summary of ontology requirements specification document.

3.2.1 Ontology Purpose

Purpose of ontology will be to provide a knowledge model of data science initiatives upon big data technologies.

3.2.2 Ontology Scope

Ontology will focus on relationships between people, processes and technologies during data science initiatives upon big data technologies.

3.2.3 Implementation Language

Ontology will be implemented in Protegé as it is based on Java and supports the latest OWL 2 Web Ontology Language and RDF specifications from the World Wide Web Consortium (W3C). DL Query plugin will be used for ontology validation.

3.2.4 Intended End Users

Ontology will be directed to end users below.

- **Customers:** organizations from any economic activity who want to run data science initiatives upon big data technologies in their own operations.
- **Consultants:** people or group of people who want to provide consultancy services for data science initiatives upon big data technologies
- **Technology Developers:** people or group of people who develop big data technologies and want to analyze how their products fit in real-world data science initiatives
- **Standards Publishers:** people or group of people who develop data science processes and want to analyze usage of their processes
- **Researchers:** people or group of people who want to perform comprehensive analyses of data science initiatives upon big data technologies for academic purposes

3.2.5 Intended Uses

Ontology will be designed to use cases below.

- **UC01:** Customers need to search for prior data science upon big data and also for related people, processes and technologies
- **UC02:** Consultants need to publish prior experiences related to data science initiatives upon big data
- **UC03:** Technology Developers need to publish new technologies, functionalities and upgrades
- **UC04:** Standards Publishers need to publish new standards, approaches and/or methods
- **UC05:** Researchers need to provide statistics and explain interactions of people, processes and technologies

3.2.6 Non Functional Requirements

Ontology will address non functional requirements below.

- **NFR01:** The ontology must support a multilingual scenario
- **NFR02:** The ontology must be based on de facto standards in existence.

3.2.7 Functional Requirements

Ontology will address particular functional requirements for Customers (CQ01, CQ02 and CQ03), Consultants (CQ04, CQ05 and CQ06), Technology Developers (CQ07), Standards Publishers (CQ08) and Researchers (CQ09 and CQ10).

- **CQ01:** What are the internal functional processes ready for a data science initiative upon big data?

Table 1: OntoDIVE - Object Properties Related to People.

Property	Domain	Range	Usage
affiliationOf	Affiliations	People, Processes and Technologies	This property relates affiliation (such as.: companies, communities, associations) to other classes.
havePriorExperience	People	Processes, Technologies, Functions, Positions and Softskills	This property relates people to their experiences
haveSoftSkills	People	SoftSkills	This property relates people to soft skills
haveFunction	People	Function	This property relates people to functions
havePosition	People	Positions	This property relates positions to people
locatedAt	People	Locations	This property relates locations to people
peopleLinking	People	People	This property relates an instance from class people to another instance of the same class

Source: Authors.

Table 2: OntoDIVE - Object Properties Related to Processes.

Property	Domain	Range	Usage
processOf	Processes	Frameworks	This property relates processes to frameworks
activityOf	Activities	Processes	This property relates activities to other classes
artifactOf	Artifacts	Processes	This property relates artifacts to other classes
constraintOf	Constraints	Processes	This property relates constraints to other classes. Processes may have constraints
performanceOf	Performance	Processes	This property relates performance metrics to other classes
changeEventOf	ChangeEvent	Frameworks Technologies	This property change events to frameworks and technologies
targetOf	Targets	Performance	This property relates targets to other classes
lifeCycleOf	LifeCycle	Frameworks Technologies	This property relates lifecycle to other frameworks and technologies
frameworkLinking	Processes	Processes	This class relates a particular class to itself

Source: Authors.

- **CQ02:** What are the technologies internally available for starting a data science initiative upon big data?
- **CQ03:** What are the internal people (or group of people) skilled for a data science initiative upon big data?
- **CQ04:** Which data science frameworks are recommended by a particular consultant?
- **CQ05:** Which big data technologies are recommended by a particular consultant?
- **CQ06:** What are the people skilled for a data science initiative upon big data working for a particular consultant?
- **CQ07:** Which algorithms, methods or techniques are included into a particular technology?
- **CQ08:** Which activities are expected to be performed during a data science initiative according to a particular framework?
- **CQ09:** What are known data science frameworks available in the market?
- **CQ10:** What are known big data technologies available in the market?

3.3 Refinement

The ontology built in this study was named as **OntoDIVE**, a short for Ontology for data science Initiatives upon big data technologies. The core class of OntoDIVE is **DSUponBD** which is intended to be applicable to a very broad range of big data initiatives. This class represents a data science initiative upon big data. The main purpose of this class is to represent integration between people, processes and technologies. Each instance of this class represents a single initiative and each initiative may be related to

Table 3: OntoDIVE - Object Properties Related to Technologies.

Property	Domain	Range	Usage
algorithmOf	Algorithms	Techniques, Methods and Technologies	This property relates algorithms to other classes
methodOf	Methods	Technologies, Approaches, Techniques and Algorithms	This property relates methods to other classes
techniqueOf	Techniques	Methods, Algorithms and Technologies	This property relates techniques to other classes
approachOf	Approaches	Methods, Technologies, Processes and Theories	This property relates approaches to other classes
theoryOf	Theories	Approaches and Technologies	This property relates academic theories to classes
categoryOf	Categories	Roles	This property relates categories to roles, according to taxonomy of big data
haveRole	Technologies	Roles	This property relates technologies to roles, according to taxonomy of big data technologies
haveCategory	Roles	Categories	This property relates technologies to categories, according to big data taxonomy
technologiesLinking	Technologies	Technologies	This property relates a particular class to itself

Source: Authors.

other initiatives. Class **People** is related to soft skills, positions, functions, locations, etc. People may be related to other people. Class **Frameworks** represents all processes used to support data science initiatives. This class is related to: processes, activities, resources, technologies, people, affiliations, theories, etc. Frameworks may be related to other processes. Class **Technologies** describes all technologies involved in a data science initiative upon big data. This class is related to: roles, theories, approaches, methods, techniques, algorithms, etc. Figure 2 shows all classes OntoDIVE intends to explain and their relationships. Next subsections present details of classes and properties, grouped by people, processes and technologies.

3.3.1 People Perspective

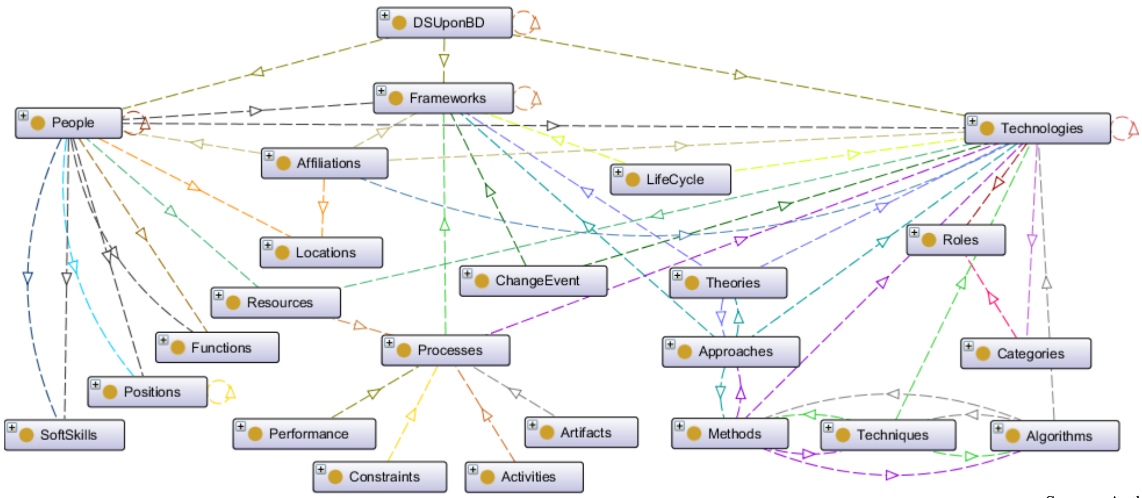
Classes below are related to people.

- **People:** This class describes all people involved in a data science Initiative. A particular initiative may require relationship between different people. A data engineer may be associated to a project leader, for example.
- **Affiliations:** This class explains affiliations such as: organization, communities, associations, companies, etc. Should a particular data science initiative requires accredited personnel from specific organizations, this class could be used.
- **Functions:** This class describes functions performed by people in data science initiatives. People may perform different functions in different initiatives. One may be data scientist in a particular initiative and a data engineer in another initiative.
- **Positions:** This class describes positions occupied by people in data science Initiatives. A particular initiative may require a hierarchical or a matrix structure. This class supports this kind of relationship. Some examples include: director, manager, staff, among others.
- **SoftSkills:** This class describes soft skills related to data science initiatives. An initiative may require certain skills, such as: logical thinking, communication, etc.
- **Locations:** This class describes physical locations where people are based at. People may be physically based in a different location of their organizations.

3.3.2 Processes Perspective

Classes below are related to processes.

- **Frameworks:** This class describes frameworks to support data science initiatives upon big data. Each initiative may adopt a different framework.



Source: Authors.

Figure 2: OntoDIVE Overview.

Examples of frameworks include: CRISP-DM, KDD, etc

- **LifeCycle:** This class represents a stage of a particular framework or technology. An initiative may use a framework during its experimental lifecycle. CRISP-DM, as an example, may evolve to CRISP-DM 2.0
- **ChangeEvent:** This class represents an event which resulted in a change of a particular framework. New version of a set of best practices is an example of this kind of event.
- **Processes:** This class describes processes associated to frameworks. Each framework has particular processes, such as: business understanding, data selection, etc.
- **Activities:** This class describes activities related to a particular process. Process "data selection" includes some activities such as: identify datasources, acquire data, etc.
- **Artifacts:** This class describes artifacts required or generated by a process. All processes generate outputs based on their inputs. Artifacts may be input or output of processes.
- **Constraints:** This class describes constraints that restrict processes. Some examples include: language to use, computational environment, etc.
- **Resources:** This class describes resources required by processes. People and technologies are some examples.
- **Performance:** This class is used to clarify the goals of end user in terms of what he wants to obtain from data.

3.3.3 Technologies Perspective

Classes below are related to technologies.

- **Technologies:** This class describes technologies of data science Initiatives.
- **Roles:** This class explains roles performed by technologies according to taxonomy of big data technologies. Few examples: data creation, data acquisition, etc.
- **Categories:** This class explains categories of technology roles. Role "data creation" may be categorized into sensors, logs, etc.
- **Theories:** This class describes all theories related to a particular technology or framework. Some examples include: information theory, automata theory, database theory, machine learning theory, etc
- **Approaches:** This class describes approaches related to theories, frameworks or technologies. A particular initiative may require an specific approach which may bring a specific set of techniques, methods and algorithms. Machine learning theory has some approaches: supervised learning, unsupervised learning, reinforcement learning, multi-task learning, etc.
- **Techniques:** This class describes techniques related to a technology or approach. Supervised learning approach may be implemented by classification or by regression.
- **Methods:** This class describes methods related to a technology or process. Supervised learning by classification may be implemented by rule learn-

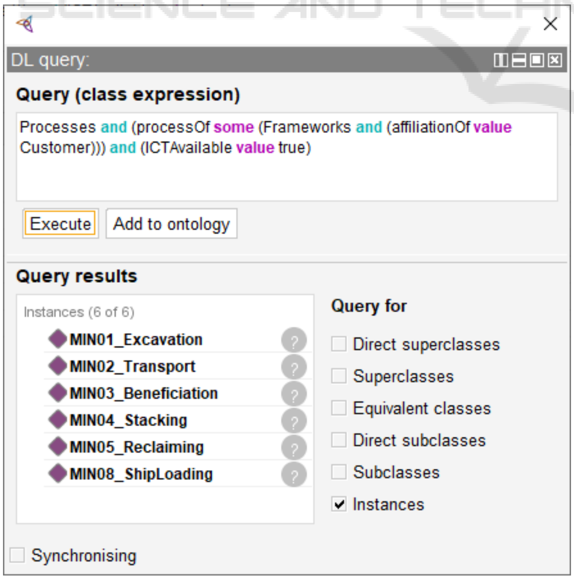
ing, neural networks, support vector machines, etc.

- **Algorithms:** This class describes algorithms. Neural networks for classification may be implemented as Radial Basis Function, Incremental Radial Basis Function, etc.

3.4 Evaluation and Application

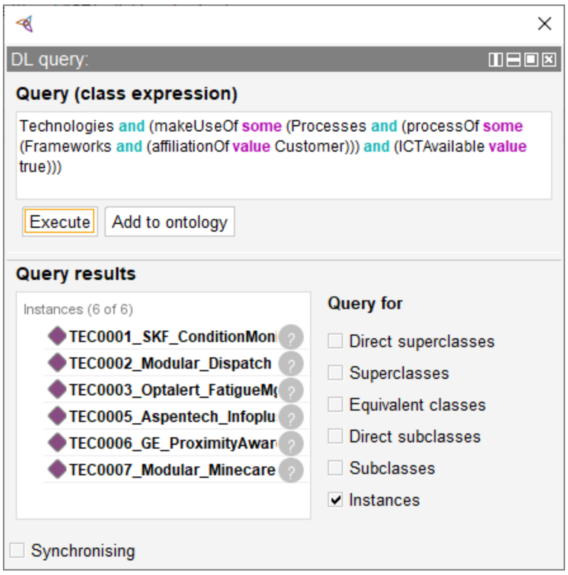
This section describes both the queries designed to answer competency questions presented in section 3.2 and the outputs provided by OntoDIVE after execution of those queries. It is important to highlight that classes and properties of OntoDIVE were populated in small scale as the purpose of this study is to build and validate an ontology capable of explain data science initiatives upon big data technologies. Although OntoDIVE has been designed to explain interactions it has not been applied in productive systems.

For each functional requirement written in natural language and described in section 3.2.7, there is a corresponding figure presenting DL Query and OntoDIVE outputs. Figures 3, 4 and 5 address competency questions from **customers**. Figures 6, 7 and 8 address competency questions from **consultants**. Figure 9 addresses competency questions from **technology developers**. Figure 10 addresses competency questions from **standard publishers**. Figures 11, 12 address competency questions from **researchers**.



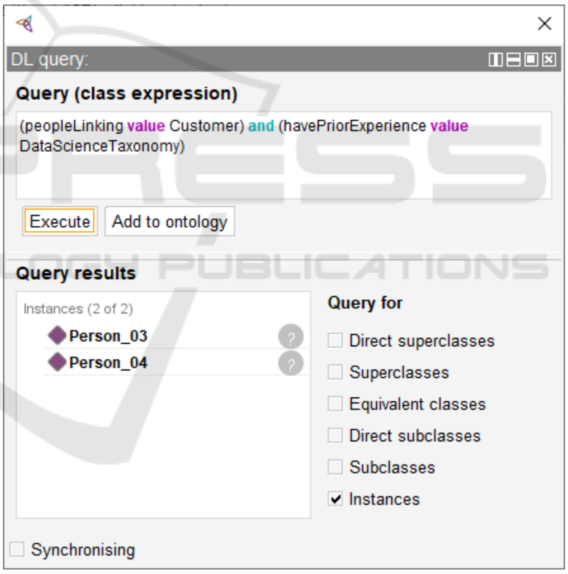
Source: Authors.

Figure 3: OntoDIVE outputs for CQ01.



Source: Authors.

Figure 4: OntoDIVE outputs for CQ02.

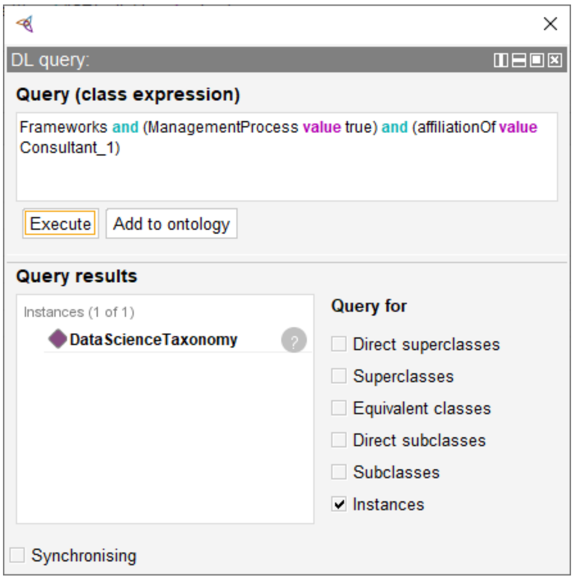


Source: Authors.

Figure 5: OntoDIVE outputs for CQ03.

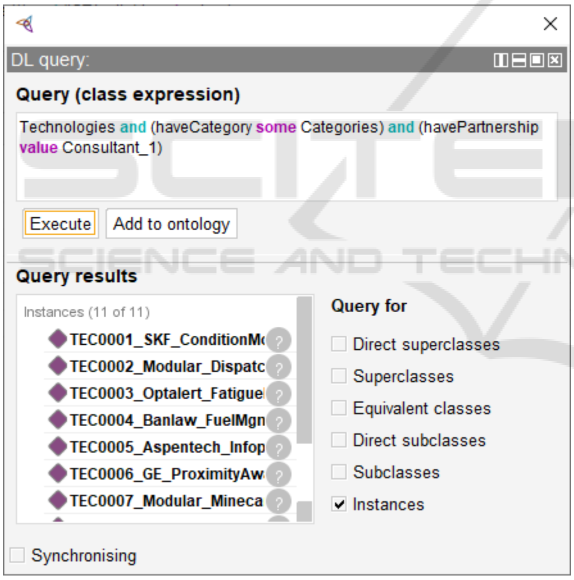
4 DISCUSSION

OntoDIVE may be used to explain interactions between people, processes and technologies in a data science initiative upon big data. The more populated OntoDIVE is the more accurate it will be to answer relevant competency questions. These answers could help organizations from all segments to leverage their data science initiatives over big data. Thus, OntoDIVE could be used as basis of a framework to sup-



Source: Authors.

Figure 6: OntoDIVE outputs for CQ04.

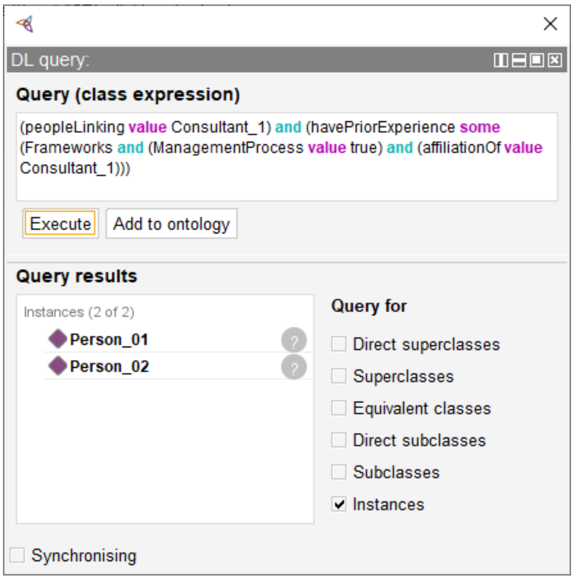


Source: Authors.

Figure 7: OntoDIVE outputs for CQ05.

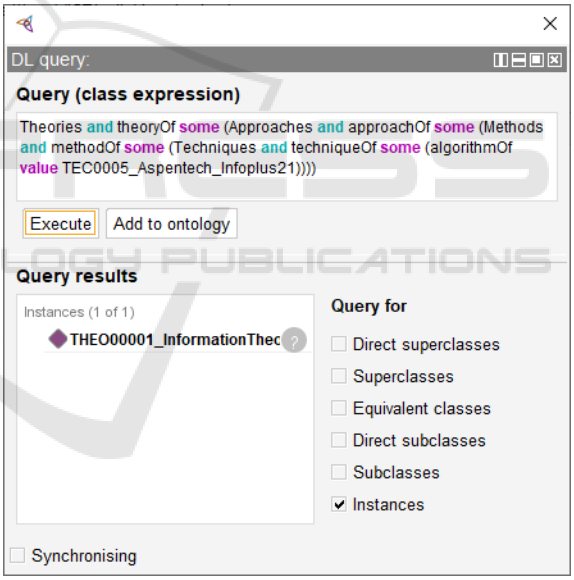
port organizations in their initiatives.

OntoDIVE may also be used to confirm interdisciplinary nature of data science initiatives upon big data technologies. Considering only instances created for this study, it is possible to see five different profiles of data scientist: domain expert, statistician expert, computing expert, business expert and communicator expert. Additionally, there are other relevant functions such as data engineer and project manager. In this study, each function was performed by different personnel with different academic background.



Source: Authors.

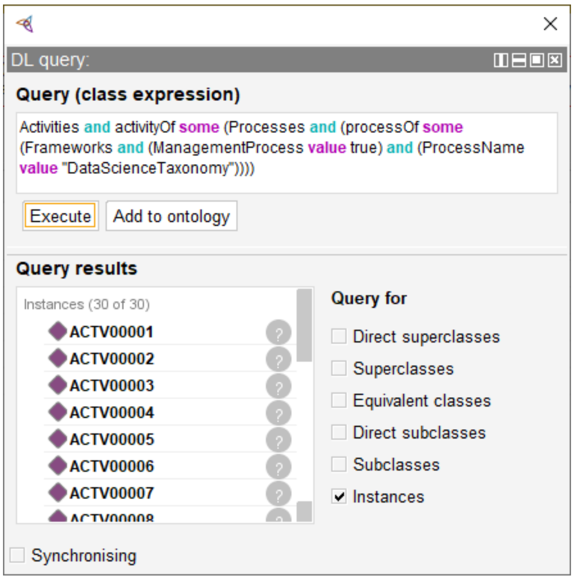
Figure 8: OntoDIVE outputs for CQ06.



Source: Authors.

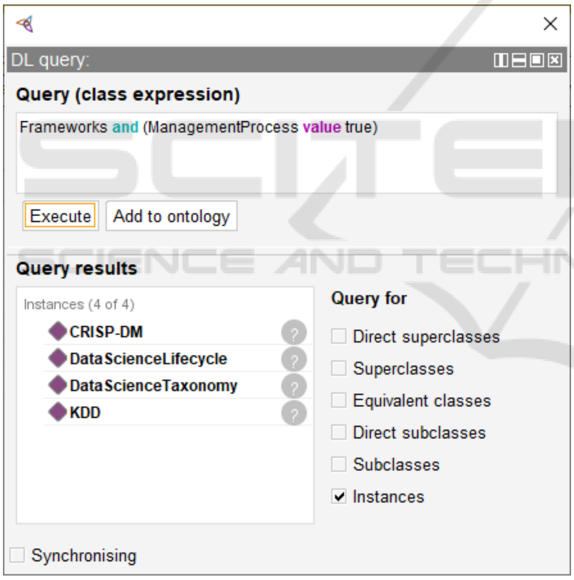
Figure 9: OntoDIVE outputs for CQ07.

Another potential usage for OntoDIVE is to make organizations capable of drawing parallels between data science results for a particular domain to their own domain. As an example, mining industry professionals could use OntoDIVE to clearly understand how data science initiatives are conducted in healthcare or transportation industries. This has a potential to enlarge possibilities of applications. OntoDIVE may be used as a tool for comparison of both data science frameworks and big data technologies. As each framework has strengths and weaknesses, it



Source: Authors.

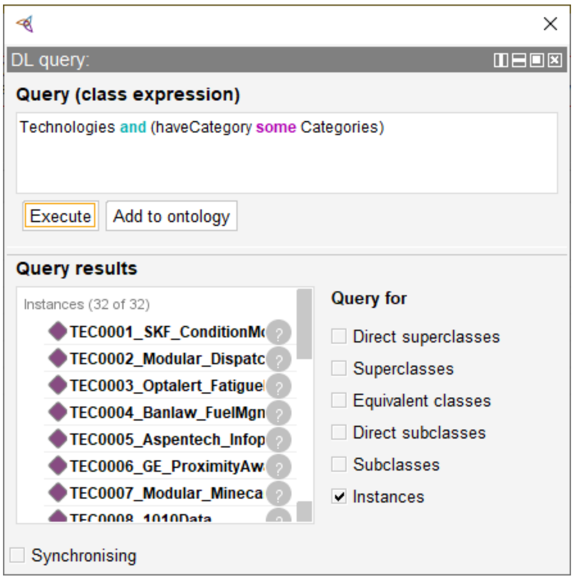
Figure 10: OntoDIVE outputs for CQ08.



Source: Authors.

Figure 11: OntoDIVE outputs for CQ09.

is important for customers to choose the more suitable for them taking into account internal restrictions. Whithin the same rationale, OntoDIVE may be used as a tool for comparison of big data technologies. OntoDIVE brings light to the fact that sometimes customers are presented to a reduced list of frameworks or technologies only because a particular consultant does not work with such framework or technology. In this regard, OntoDIVE may be used to protect customers of being subject to commercial interests depending on consultants they rely on.



Source: Authors.

Figure 12: OntoDIVE outputs for CQ10.

Main goal of mining industry is to minimize the amount of assets and resources required to run operations. There are many opportunities but, in a general way, industry seems to be focusing on cost drivers, such as: 1) increase productivity; 2) increase profitability; 3) improve assets management. In this context, data science initiatives upon big data should be focused on understanding how to reduce waste in supply chain and on finding what precisely drives fuel consumption. In the specific case of mining company object of this study, Figure 3 shows internal processes with data science readiness. This readiness is related to existence of systems that capture and make process data available. Figure 4 shows technologies internally available for starting a data science initiative. Mining industries usually have condition monitoring systems, dispatch and fatigue management systems for mining operations. A data science initiative should consider gathering data from these systems. Figure 5 shows people with knowledge or prior experience in data science initiatives. Although people may be geographically dispersed in a large company, it is relevant to identify everyone that could add value to a data science initiative.

5 CONCLUSION

This work contributes to the clarification of concepts and terminologies related to data science and big data. Data science initiatives upon big data can be analyzed from three different perspectives: people, processes and technologies. OntoDIVE ontology, proposed in this study, explains relationships among these terms and concepts in the context of a data science initiative upon big data. Proposed ontology may also be used to confirm interdisciplinary nature of data science initiatives upon big data technologies. Another potential usage for OntoDIVE is to make organizations capable of drawing parallels between data science results for a particular domain to their own domain. The ontology may be used as a tool for comparison of both data science frameworks and big data technologies and could be used as basis of a framework to support organizations in their data science initiatives upon big data.

This work has several limitations. Although OntoDIVE was designed to be a comprehensive artifact capable of explaining most part of data science initiatives upon big data, it was built based on considerations from specialized professionals that work for a single organization. Furthermore, while several ontologies within the same domain are developed independently by different communities, this study was never focused on merging OntoDIVE with existing ontologies and no method was used for ontology alignment, as proposed by (Idoudi et al., 2016). Additionally, while OntoDIVE was validated by description logics queries it has not been applied in productive system. Therefore, OntoDIVE should be considered as an initial version.

Future works could create a friendly graphical user interface to allow interaction with OntoDIVE, since Protegé interface is not understood by many who are not knowledgeable about ontologies and their editing tools. Future works coud also apply OntoDIVE on productive systems in order to collect insights and thoughts of more people and organizations focusing on evolution of the ontology.

REFERENCES

Abdullah, M. F., Ibrahim, M., and Zulkifli, H. (2017). Resolving the misconceptions on big data analytics implementation through government research institute in malaysia. In *International Conference on Internet of Things, Big Data and Security*, volume 2, pages 261–266. SCITEPRESS.

Dietrich, D., Heller, B., and Yang, B. (2015). *Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting*. Wiley, [s.l].

Hurwitz, J., Kaufman, M., and Bowles, A. (2015). *Cognitive Computing and Big Data Analytics*. Wiley Publishing, [s.l], 1st edition.

Idoudi, R., Ettabaa, K. S., Hamrouni, K., and Solaiman, B. (2016). Fuzzy clustering based approach for ontology alignment. In *ICEIS (1)*, pages 594–599.

Liu, J., Li, J., Li, W., and Wu, J. (2016). Rethinking big data: A review on the data quality and usage issues. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115:134 – 142. Theme issue 'State-of-the-art in photogrammetry, remote sensing and spatial information science'.

Mizoguchi, R. and Ikeda, M. (1998). Towards ontology engineering. *Journal-Japanese Society for Artificial Intelligence*, 13:9–10.

Musen, M. A. (2015). The protÉgÉ project: A look back and a look forward. *AI Matters*, 1(4):4–12.

Osterwalder, A. (2004). *The business model ontology a proposition in a design science approach*. PhD thesis, Université de Lausanne, Faculté des hautes études commerciales.

Parreiras, F. S. (2011). Marrying model-driven engineering and ontology technologies: the twouse approach.

Spyns, P., Meersman, R., and Jarrar, M. (2002). Data modelling versus ontology engineering. *SIGMOD Rec.*, 31(4):12–17.

Staab, S. and Mädche, A. (2000). Axioms are Objects, too - Ontology Engineering beyond the Modeling of Concepts and Relations. In Benjamins, V., Gomez-Perez, A., and Guarino, N., editors, *Proceedings of the Workshop on Applications of Ontologies and Problem-solving Methods, 14th European Conference on Artificial Intelligence ECAI 2000, Berlin, Germany*.

Staab, S., Walter, T., Gröner, G., and Parreiras, F. S. (2010). *Model Driven Engineering with Ontology Technologies*, pages 62–98. Springer Berlin Heidelberg, Berlin, Heidelberg.

Sure, Y., Staab, S., and Studer, R. (2009). Ontology engineering methodology. In *Handbook on ontologies*, pages 135–152. Springer.

Vera-Baquero, A., Colomo-Palacios, R., and Molloy, O. (2016). Real-time business activity monitoring and analysis of process performance on big-data domains. *Telematics and Informatics*, 33(3):793 – 807.