How much Knowledge is in Knowledge Graphs? A Knowledge Management Perspective

Ivanova, Rositsa; Kerschbaum, Clemens; Neumaier, Sebastian; Kaiser, Alexander; Polleres, Axel

Published: 01/10/2020

Document Version:
Publisher's PDF, also known as Version of record

How much Knowledge is in Knowledge Graphs? A Knowledge Management Perspective

Rositsa V. Ivanova[1], Clemens Kerschbaum[1], Sebastian Neumaier[2], Alexander Kaiser[1] Axel Polleres[1]

[1] Vienna University of Economics and Business, {firstname.lastname}@wu.ac.at
[2] St. Pölten University of Applied Sciences, {firstname.lastname}@fhstp.ac.at

Arbeitspapiere zum Tätigkeitsfeld
Informationsverarbeitung, Informationswirtschaft und Prozessmanagement
Working Papers on Information Systems, Information Business and Operations

Nr./No. 02/2020
ISSN: 2518-6809

Herausgeber / Editor:
Department für Informationsverarbeitung und Prozessmanagement
Wirtschaftsuniversität Wien · Welthandelsplatz 1 · 1020 Wien
Department of Information Systems and Operations · Vienna University of Economics and Business · Welthandelsplatz 1 · 1020 Vienna

How much Knowledge is in Knowledge Graphs? - A Knowledge Management Perspective

Rositsa V. Ivanova[1] Clemens Kerschbaum[1] Sebastian Neumaier[2] Alexander Kaiser[1] Axel Polleres[1]
[1] Vienna University of Economics and Business, {*firstname.lastname*}@*wu.ac.at*
[2] St. Pölten University of Applied Sciences, {*firstname.lastname*}@*fhstp.ac.at*

Abstract

Managing and preserving knowledge in the best possible way has always been a key to the success for organisations, long before the term "Knowledge Graph" has entered the stage. However, the understanding of what exactly knowledge is, how it is represented and organised, and how knowledge is created often varies between different research communities. To this day, the scientific discipline of Knowledge Management is trying to capture the process of knowledge creation as converting implicit, i.e., tacit, knowledge into explicit knowledge. In this paper, we first summarize the idea of the Knowledge Management perspective on knowledge creation, and second discuss how Knowledge Graphs can contribute to solve the issue of making implicit knowledge explicit. We empirically survey the use of Knowledge Graphs in enterprise environments and discuss concrete use cases from a Knowledge Management viewpoint by selecting three concrete examples.

1. Introduction

The term and concept of a "Knowledge Graph" (KG) has recently been happily embraced by the Semantic Web [1] research community. It has opened new directions beyond Web standards for semantic links between URI-identified concepts alone (i.e., Linked Data), towards organizing data around semantic concepts and their relationships in more general terms. This said, we note that other core computer science sub-communities have either used the term "knowledge" much longer already, or have recently started to adopt it.

The Knowledge Representation and Reasoning (KRR) community, for example, works under the "*fundamental assumption [...] that an agent's knowledge is explicitly represented in a declarative form, suitable for processing by dedicated reasoning engines*"[1]. Yet, the "Knowledge Graph" metaphor seems to be broader than the KRR understanding, also spilling over to other communities who recently pick up the term. This is for instance demonstrated with articles carrying the term "Knowledge Graph" in their titles in top conferences on the topics of Databases (from the last SIGMOD and VLDB conferences) [2, 3, 4], Machine Learning (from the last NIPS conference) [5, 6, 7], or Natural Language Processing (from the last ACL conference) [8, 9, 10]. These references deliberately only reflect a current snapshot of communities and approaches, that may not have focused on considerations of explicit knowledge in the past, yet now recognise knowledge itself as something that could be "created" or "completed" by automated means.

Thus one could argue that while the core KRR community understands knowledge as something that is already explicit, formalised, and as something that can be operationalised by automated agents, the new KG view is also working with sub-symbolic, and as one could claim implicit forms of knowledge. In particular, the core novelty seems to be the combination of explicit and implicit knowledge, which is potentially leveraged through KGs. This new perspective has fueled the hope for an *artificial but still explainable* intelligence in machines [11], where such "explainabilty" should be understood as again being explicitly communicable to humans.

Interestingly though, many research works in the area of and around KGs do not precisely define the term "knowledge" itself. Yet such a definition is the very basis for assessing to what extent "knowledge" can be created by a machine. We address this lack of definitions by applying recognized knowledge-definitions from the field of *Knowledge Management (KM)* to the discourse around KGs. This will not only help to better understand how machines can support human knowledge creation but – perhaps even more importantly – support managers in evaluating the amount of knowledge that is stored in an organizations KG. To this end we present a theoretical argument about the knowledge contained in KGs and provide three selected illustrative examples of

[1] Web page of the 17th International Conference on Principles of Knowledge Representation and Reasoning (KR2020), `https://kr2020.inf.unibz.it/`, last accessed: 9 October 2020

organizations KGs.

For our considerations we follow the knowledge-based view on organizations that builds on Penrose's resource-based theory of the firm [12]. According to the knowledge based view, knowledge is especially crucial, because it enables the use of all other tangible resources in order to create value. The better an organisation exploits given resources through knowledge, the more likely it is to acquire competitive advantages in the long run [13, 14, 15]. However, the problem is that knowledge in a firm is hard to grasp – there is nothing like a "catalogue of knowledge". In fact, most knowledge is invisible as it is often embedded in organisational routines, carried out as organisational culture or simply bound to individual employees [16]. Usually, in relation to all existing knowledge in a company, only a very small amount of knowledge is codified and made explicit. The majority of the knowledge is considered to be implicit knowledge. Still, implicit knowledge is the important source of competitive advantage for organisations [17]. Therefore, it can turn out difficult to access a firms knowledge. First, we do not know what is all there. Second if we knew what was there, we would not know where it was. Third, it would eventually be hard to make sense of it without further knowledge of the context around a specific chunk of knowledge. To tackle those problems, literature points out that organisations should strategically manage their knowledge [16].

Following this claim, it quickly becomes apparent that technology could prove as a valuable means to exploit knowledge [18]. This could be especially in terms of storing already codified knowledge, but also to connect individuals in an organisation with experts in order to leverage their tacit knowledge on a certain subject [16, 19]. However, as Alavi and Leidner point out "*it is less the knowledge existing at any given time per se than the firms' ability to [...] create new knowledge*" that lead to competitive advantage [16]: it is this *generation of new knowledge* that really creates value for the organization.

A widely appreciated explanation for the process of knowledge creation is Nonaka's SECI model[2] [20]. Following this model, the evolution of knowledge is a circular process involving the recurring transformation of tacit knowledge into explicit knowledge, and explicit knowledge back into tacit knowledge that leads to the creation of new knowledge [20]. The process of knowledge transformation along the SECI model particularly emphasises the social component of knowledge exchange, rather than technical means.

[2] The model consists of the four phases Socialisation, Externalisation, Combination and Internalisation in which knowledge is converted from implicit to explicit and vice verca

Previous literature has explored how tacit knowledge can be used in organisations and how it can be transformed into explicit knowledge [21]. Nevertheless the question whether modern technical means – in particular KGs – are suitable and supportive for this task still remains largely unanswered. In order to close this gap, in the present work we focus on the question:

> *How can knowledge graphs foster the conversion of implicit into explicit knowledge and thus support the creation of new knowledge in organisations?*

We herein aim to examine how KGs could support knowledge creation in organisations and how they can provide us with possibilities to identify, evaluate, combine, and make sense of existing (internal) organisational knowledge. The remainder of this paper is structured as follows. Sections 2 and 3 introduce key aspects of KM and KGs, respectively. Section 4 discusses whether and how existing knowledge can be managed and new knowledge can be created using KGs. This is done by presenting a literature survey, followed by three relevant case study examples. We summarise our findings in Section 5 touching upon the discussed topics, the related potential issues, and yet to be answered questions.

2. Knowledge Management

Discussions on what is "knowledge" are getting more relevant due to the aforementioned trending technology areas, Other disciplines have dealt with this question from a much broader perspective. In particular, the socio-technical perspective of the concerned domains is often overlooked in recent discussions. While such a perspective is often regarded as a rather philosophical one, there is pragmatic relevance to it. It discusses questions such as *"What does a robot need to know in order to open a safe, and how does it know whether it knows enough to open it?"* [22]. The answers to such questions include the knowledge of *multiple actors within groups*. Therefore, we further regard the term knowledge both on a personal and on an organisational level.

An organisation is typically regarded as a group of people. On an organisational level it holds knowledge, consisting of the constantly repeated routines of the organisation - called *"organisational knowledge"* [23]. On an individual level, all members of the group are holders of individual knowledge, contributing to the creation of new and the forwarding of existing knowledge. At the same time, the inarguably important role of automated means and information systems needs to be considered, as they continuously provide people with new information in this knowledge creation process.

2.1. What is Knowledge?

From a KM perspective the term “knowledge” and its definition date back to the classical philosophers of the Greek era and are still actively discussed in recent publications. However, the term “knowledge” has not yet been defined in terms of its use in *Knowledge* Graphs by the Semantic Web community. Therefore, we adopt existing knowledge definitions for the purpose of this work.

In Theaetetus by Plato (369 BC), Socrates introduced the traditional definition of knowledge as *“justified true belief”* (i.e., subjective personal perspective) [24, 25]. Rather than a simple compilation of facts, the creation of knowledge is characterized by an individual making sense of a new situation by holding a true belief with an account [26]. Later on, this understanding is challenged by the renowned publication of Edmund Gettier [27]. The philosopher presents two counterexamples, illustrating that even justified true belief could turn out to be false. Consequently, a cognitive perspective would argue that the knowledge is dependent on the individual making sense of it [28].

Another definition of knowledge views it as an enabling activity, a *“capacity to act”* [29]. Sveiby focuses on the action element and thus on the dynamic properties of knowledge. He argues that knowledge is only useful, when it is available to individuals and increases their capacity to act. All those and further perspectives on knowledge have lead to multiple definitions on an individual and an organisational level.

Data, information, knowledge The literature on knowledge management frequently distinguishes between data, information, and knowledge. Faucher et al. give an overview of the definitions of those terms [24]. While there does not seem to be a consensus, similarities in the understanding can be found.

- *Data* is primarily seen as unprocessed raw representation of reality.
- *Information* builds on it by providing a context and thus a semantic meaning.
- Once the gained information is utilised pragmatically for the introduction of new actions, new meanings, etc. we speak of *knowledge*.

Faucher et al. also discuss the term *wisdom* in detail; we exclude the concept of wisdom in our discussions as it lies out of the scope of the paper.

Knowledge is carried through multiple entities (e.g. organisational routines, culture), is context specific and dependent on a particular time and space [30, 25]. Without those characteristics, it simply becomes information. Various publications in the past have stressed upon the importance of differentiating between knowledge and data or information. In contrast to information, knowledge is about beliefs and commitment, as well as about action and lastly, about meaning [31]. If knowledge is not different from data or information, it could be described as a stock rather than a flow and an opportunity for the generation of something new or interesting in organisational terms [32]. Additionally, there are two aspects of information - the *syntactic* and the *semantic* one. The syntactic aspect merely observes the volume of the information flow, while the semantic aspect conveys a meaning, which is useful for the creation of knowledge [31].

Implicit and Explicit Knowledge Taking a different viewpoint, the literature on knowledge management has agreed upon the separation of knowledge in two yet interconnected concepts - namely, *explicit* knowledge and *tacit* knowledge [33]. On the one hand, explicit knowledge can be expressed by individuals in formal and systematic language in forms such as text and sound, and is thus easily communicated, stored and processed. On the other hand, tacit knowledge rests in individuals’ intuitions, actions, bodily experience and also in organisations’ routines and habits. In contrast to explicit knowledge, tacit knowledge is not easily expressible [31, 25]. In terms of KGs this means that explicit knowledge is stored in already created elements (e.g. nodes, edges) of graphs, whereas tacit knowledge could eventually be derived from the meaning of the graphs as a whole (i.e. its functionality). This results from the ability of the graph to provide information about things that are not contained as explicit elements, but can be accessed through deliberate inquiry. This could be for example the use of the Google KG for answering a particular question such as “Is COVID-19 deadlier than the flu?”. Although the KG will probably not contain the answer to this question in an explicit format, it will still be able to derive an answer from its existing explicit content - a hint to potentially implicit knowledge.

As pointed out by Virtanen [34], one may find multiple definitions of tacit and explicit knowledge throughout the KM literature. In this paper we adopt the definition from Polanyi’s theory without discussing the philosophical foundation it builds upon. It describes explicit and tacit knowledge as two different kinds of awareness, but rather in form of a continuous spectrum than as separate or even independent concepts. In this context, explicit knowledge refers to focal awareness, tacit - to subsidiary awareness. Those two are mutually exclusive, this means that the awareness cannot be both simultaneously. Nevertheless, when the attention is put on the tacit knowledge, it turns from subsidiary to focal (i.e. from implicit to explicit). Lastly, subsidiary awareness serves as an enabler of focal awareness through its various clues, elements and processes [34]. Next, and in order to understand the importance of the distinction between explicit and tacit knowledge at the core of our argument, it is beneficial to outline how new knowledge is created.

2.2. On Knowledge Creation

The creation of organisational knowledge can be separated into two important steps - the process of knowledge creation by individuals and the integration of individual knowledge into the organisation [17]. Nonaka and Takeuchi's SECI model (*Socialisation, Externalisation, Combination, Internalisation*) identifies four steps of the knowledge creation process [33]. From a two-dimensional perspective an organisation converts tacit knowledge into explicit knowledge and then back to tacit knowledge. In the *Socialisation* mode tacit knowledge is shared and created by individuals through their actions and observations. Next, dialogue, reflection, metaphors are used to make tacit knowledge explicit. This is referred to as *Externalisation*. In the third mode - the *Combination* - explicit knowledge is structured and applied. Lastly, the explicit knowledge shifts to tacit by being simulated, applied and thus embodied in during the *Internalisation* phase [31, 25]. Ultimately, new knowledge is created through the permanent exchange of explicit and implict knowledge on all ontological levels.

In addition to the those two dimensions of the SECI model, the process contains an ontological level, which turns it into a spiral. Beyond the differentiation between explicit and tacit knowledge, each phase of the model captures the interaction between different combinations of entities - individual, group and organisation. While the Socialisation phase takes place only between individuals, each of the remaining phases involves a different pair combination of the three entities (in the given order) [31, 25].

2.3. Technical Support in the Process

The management of existing and the creation of new knowledge is often supported by information systems [18]. In their work on how organisations manage knowledge, Davenport and Prusak touch upon the ways in which computers may help "*transform data into information by adding value to it*" [35]. They recognise a total of five relevant methods – *Contextualized, Corrected, Categorized, Calculated, and Condensed* – noting that the latter three methods must be performed by humans, as only an individual could add meaning to data [35].

Considering the progress of technology over the last decades, however, it may be arguable, whether there is still a necessity for frequent human intervention in any of those above-mentioned five methods. Therefore, it may also be arguable whether information systems only contain information. However, it is important to consider that most so-called knowledge bases (a term in between "database" and "knowlegde"), in particular most KGs, have often been created by individuals (in companies or crowdsourced) and thus store their collective explicit knowledge. The stored knowledge is then made use of by organisations for their individual purposes. In the following section, we illustrate how KGs can serve as an approach for storing, structuring and organising knowledge bases in a graph.

3. Knowledge Graphs

The idea of representing knowledge has been around for long. There are plenty of methods to represent knowledge in order to conserve it, make it more comprehensible, accessible or meaningful (e.g. mind maps or concept maps) [36, 37]. At their core, these methods share basic principles with a "Knowledge Graph" as they connect chunks of information, thereby giving it structure and meaning.

Indeed the term "Knowledge Graph" is not an invention of the recent past. While the term has been appearing in the literature already for decades (at least since 1974 [38]), recent uptake stems from industry. Google announced their Google Knowledge Graph in 2012, initiating several subsequent announcements and developments of KGs, (e.g., at Airbnb, Amazon, Facebook, Uber) [39]. The uptake in industry was followed by many scientific publications, including articles on the definition of KGs [40], on construction and refinement techniques [41], and entire books [42, 43, 44]. We aim to give an idea and understanding of what is covered by the term by first discussing some definition approaches and second analyzing the underlying principles and techniques of modeling and creating KGs.

Definitions and Graph Models The conceptual idea followed by KGs in the enterprise setting and in academia is to represent information in a graph abstraction. Existing graph models offer a number of benefits compared to relational models and NoSQL alternatives. Hogan et al. [39] propose the following very inclusive view on KGs: "a graph of data intended to accumulate and convey knowledge of the real world, whose nodes represent entities of interest and whose edges represent relations between these entities". A more concrete definition is given by Ehrlinger and Wöß [40], who view a KG as a graph that "acquires and integrates information into an ontology and applies a reasoner to derive new knowledge". This definition distinguishes a KG from an ontology by its ability to generate new knowledge through reasoning. However, we can see that these definitions of a KG do not emphasise on the term knowledge as such, but rather focus on the technical representation and data model.

There are different graph-based data models for building a KG, such as directed edge-labelled graphs, and property graphs [39]. The *directed edge-labelled graph* model is defined as a set of nodes and a set of directed edges, where both nodes and edges are labelled. Typically, the nodes represent entities and the edges relations between them. A standardised model for directed edge-labelled graphs is RDF [45] and the corresponding query language SPARQL [46].

The RDF model was developed in the context of the Semantic Web. By naming entities using Internationalised Resource Identifiers (IRIs) the resources, such as cities or events, can be uniquely identified and the graphs can be interlinked. The underlying idea – inititally proposed by Berners-Lee[3] – is to build a "Web of Data", which makes the interlinked content not only available for human consumption (e.g., as HTML documents), but also for computers to allow automated processing (e.g., combining, reasoning, or querying the data). Eventually, such a Web of Data consists of graphs published on individual web pages. Following links to other nodes from one graph (i.e. performing an HTTP lookup on an IRI of the graph) potentially leads to another graph elsewhere on the Web with further content to retrieve.

Creation of Knowledge Graphs The techniques to create and enrich KGs range from automated methods, to human collaboration and crowd-sourcing. The methodology heavily depends on factors such as the domain of data, the application of the KG, the involved actors, and the availability of underlying data sources.

Some popular existing KGs have been constructed from human collaboration. For instance, the Wikidata project[4] is a graph of manually curated and updated entities and relations across different domains and languages. Human collaborators collect and edit the information in Wikidata. There is an interface to issue structured queries. The information from Wikidata is then re-used and linked in various other sources on the Web, (e.g., in Wikipedia articles).

Complementary, the popular DBpedia project [47] is based on an automated extraction of the structured information in Wikipedia. The extraction constructs a graph based on the links (i.e. relations) from one article (i.e. entity) to another within Wikipedia, leveraging the template structure of infoboxes to label relations. The resulting KG is further automatically enriched by linking to further structured resources (e.g. GeoNames[5] and WordNet [48]). Alternative KG creation techniques are based on entity linking and relation extraction from text sources, as well as mapping and relation extraction from Web tables, etc. [39].

For a comprehensive survey covering various aspects including the construction, models, ontology and entailment techniques we refer to the survey by Hogan et al. [39]. For further reading on KG refinement techniques we refer to the survey by Paulheim [49]; Regarding applications and use cases we refer to the books by Fensel et al. [44] or Janev et al. [43].

4. Management and Creation of Knowledge using Knowledge Graphs

Apart from use cases on the (public) Web, KGs are also increasingly used within companies and organisations to structure data and manage their knowledge. In order to explore how organisations can leverage KGs for actual knowledge creation in the field, we surveyed conference publications published by companies. The goal of this survey is to identify how KGs help to manage organisational knowledge, and to eventually highlight cases where KGs help to make implicit knowledge explicit and thereby generate new knowledge.

4.1. Literature Survey

In order to assess the development of KG applications, we focused on relevant publications in the International Semantic Web Conference (ISWC) as a top ranked conference in the field of Semantic Web and as one of the main venues with a dedicated industry track for publications related to KGs.

For our survey, we considered all available publications of the ISWC industry track, which include the years 2019,[6] 2018,[7] and 2017[8] containing 16, 14, and 18 papers respectively. In total, we surveyed 48 papers in order to obtain insights in how far KG technologies reportedly contribute to knowledge creation within organizations.

[3] `https://www.w3.org/DesignIssues/LinkedData.html`

[4] `https://www.wikidata.org/wiki/Wikidata:Main_Page`

[5] `https://www.geonames.org/`

[6] `http://ceur-ws.org/Vol-2456/`

[7] `http://ceur-ws.org/Vol-2180/`

[8] `http://ceur-ws.org/Vol-1963/`

Categorisation & Metrics For each of the papers we collected a set of relevant features and characteristics.

- **term KG:** Firstly, we scanned the papers for the explicit use of the term "knowledge graph". Our rationale behind this is to highlight these papers, however, we also observed relevant publications using a different terminology.
- **data sources:** Secondly, we categorise the source of the described/covered data into *internal data*, *external data*, or *both*. Internal data refers to previously existing data or such produced within the company. In contrast, external data is data that has not previously been integrated in the company's work. We distinguish between these cases in order to highlight, whether internal or external sources were used to create the KG and thus for the potential creation of new knowledge.
- **open KGs:** In case the paper explicitly describes the integration/use of external sources, we further consider, whether it is explicitly stated that the external data comes from an open KG (such as DBpedia or Wikidata).
- **knowledge creation:** We are particularly interested in papers that describe a process of making knowledge explicit through the use of KG techniques.

In Table 1 we quantitatively summarize the surveyed ISWC industry track papers:[9] In 20 out of the 48 papers the term "knowledge graph" was used. This indicates a strong focus on the use of KGs in industry in recent years at ISWC. Most of the reviewed publications discuss the use/integration/representation of data from *internal* sources. We only observed 12 papers (in the *external* and *both* category) which mention the use/integration non-internal data sources; eight out of these papers use openly available KGs (e.g. Wikidata, or DBpedia). Of particular interest to our research were the 16 papers (category *knowledge creation*) that explicitly discuss how KGs helped to make knowledge explicit in an enterprise context.

In addition to the selected metrics, we collected metadata from the surveyed papers such as the keywords provided by the authors, the company names for the corresponding paper and the research area of the published paper. While those metrics are not directly needed for the analysis in this paper, we find them helpful as they offer a better overview of the applications of KGs in industry. The details about the individual papers and the metrics can be found in the spreadsheet provided online[10].

[9]Note: The sum of the data sources in the table (i.e., internal, external and both) does not correspond to the number of papers in the years 2017 and 2019, since in some of the papers, this categorisation did not appear to be applicable or it turned out there was no explicit description of the used data sources.

4.2. Case Studies

To explore the process of "knowledge creation" (i.e. gaining new *explicit* knowledge) using the idea and techniques of KGs, we discuss a selection of three of the surveyed papers as exemplary case studies: (i) In Strötgen et al. [50] the sources of knowledge (i.e. relational databases and unstructured documents) are organized in a KG to generate new knowledge and eventually improve search functionalities. (ii) In Cotroneo et al. [51] the already structured information from several sources is integrated into the company's KG, again to improve and increase the search scope. Finally, (iii) in Ireson and Ciravegna [52] existing open KGs are integrated: Wikidata and DBpedia are used to enrich the initial domain knowledge base and to more accurately solve a prediction task. We consider those three to be representative, in the sense that the other articles we surveyed resemble these three uses, i.e., (i) *knowledge generation by restructuring*, (ii) *knowledge generation by integration*, and (iii) *knowledge generation by enrichment*.

(i) Strötgen et al., 2019. Towards the Bosch Materials Science Knowledge Base [50]

In the context of manufacturing, there is a wide variety of materials used in the creation of innovative products. Due to the rapid introduction of new papers, patents and regulations in the area, the retrieval of relevant information may be challenging. Often, complex queries are necessary to find the appropriate answers (e.g. *"Find anode materials in Intermediate Temperature Solid Oxide Fuel Cells (IT-SOFC) that produce high power density"* [50]). The authors state three required goals for such a query answering system: (i) integration of different sources into a unified KG, (ii) complex query functionalities such as aggregation and multi-hop reasoning, and (iii) provenance for query results.

Combination of Data Sources The paper describes the use of both internal and external sources that get integrated in a graph. The system integrates internal relational databases using Ontology-based access methods. Further, relevant text-based publications and patents are processed and integrated using NLP techniques.

Creation of new Knowledge The KG is built using entity resolution: using the graph structured (i.e. the established links) the system allows advanced queries

[10]`https://short.wu.ac.at/ISWC-ind-track-2017-2019-KGs`

Table 1. Surveyed ISWC industry track publications.

year	term KG	internal	external	both	use open KGs	knowledge creation
2019	7	9	0	3	4	7
2018	8	10	1	3	1	5
2017	5	10	4	1	3	4
total	20	29	5	7	8	16

using domain-specific entities and extracted facts. The authors argue that only the combination and integration of the sources in the KG allow such functionalities. This means only the KG makes the knowledge explicitly available via their search system.

(ii) Cotroneo et al., 2018. From Data Search to Data Showcasing: The role of Semantic Technologies in a New Service [51]

Research institutions (e.g. universities) have been struggling with finding available datasets relevant for their research. The paper discusses the creation of a research data management system, which combines cross-organisational metadata to support the search for data. The challenge in this work is that current research data repositories do not provide complete metadata, and if available, the metadata is only available as free-text. For instance, the corresponding authors' institutions cannot directly be linked due to varying spellings, etc. Using a KG of institution, publications, and datasets the free-text metadata can be disambiguated and linked. Eventually, the graph can be used to set up a search engine over the indexed datasets.

Combination of Data Sources Elsevier's previous datasets search engine is based on incomplete free-text labels (of institutions). In order to improve the available metadata and thus the search for datasets two kinds of semantic technologies are made use of. The paper describes the integration of metadata about publishing institutions, publications, and datasets from several external sources (e.g., DataCite) into the internal Elsevier KG "Scopus".

Creation of new Knowledge Through the integration of the cross-organisational metadata new links between entities are created. Through the combination of explicit links between entities, existing in the individual databases, previously disconnected nodes in the newly created graph are connected. That means, in the KG e.g. institutions, publications, and datasets are connected, which generates this new knowledge, and now allows to search for e.g. publications and datasets from certain institutions. The individuals are therefore able to utilize the knowledge stored in the KG for the purpose of finding relevant datasets.

(iii) Ireson and Ciravegna, 2017. FootballWhispers: Transfer Rumour Detection [52]

Football Whispers uses social media activity (e.g. Twitter) to predict the likelihood of transfers of players between football teams. The authors rely on rumours on social media to get an indication of player transfers. However, a KG is used to identify players and teams in the noisy, informal social media posts. Available (crowdsourced) open KGs, such as DBpedia and Wikidata, are a rich source of players, teams, and even their transfer history.

Combination of Data Sources Football Whispers is based on the integration of three external sources into their KG: the knowledge base "Opta Sports", which provides football related domain knowledge, Wikidata, and DBpedia. The entities available in "Opta Sports" are mapped to the open KGs in order to gain alternative labels (e.g. nicknames, multilingual spellings) for players, teams, etc.

Creation of new Knowledge By adding alternative labels (e.g. nicknames) for the entities the social media activity (e.g. tweets) can be filtered and analyzed based on players, teams, etc. more accurately. The collected social media posts are then analysed and used as evidence for potential transfers.

5. Summary & Conclusions

In this paper we have showcased how KG technologies can indeed support organisations and companies in the process of *knowledge creation*. The KM community has extensively researched how knowledge is created. We have provided a brief overview of definitions of knowledge and discussed the importance of explicit and implicit knowledge for organisations. Additionally, we have given a definition and short introduction to KGs.

On that basis, we have surveyed different implementations of KGs and analysed a set of research papers along various dimensions. In particular, we have focused on papers that describe the process of new knowledge creation through the use of KG techniques, in a company environment. To this end, we have surveyed all available publications of the industry track at the past ISWC conferences – a total of 48 papers. Overall we could identify three different categories of knowledge generation:

C1 **Knowledge generation by restructuring**: The first knowledge generation process, which we observer, was already to be found while companies restructure information in a graph structure. The re-organisation and restructuring (e.g. by using new class/type hierarchies) allows us to apply graph-based techniques such as entailment, and already potentially makes new knowledge explicitly available (which was unavailable before).

C2 **Knowledge generation by integration**: In various applications we saw that company's resources were combined and interlinked (using techniques such as named-entity resolution and relation extraction). These newly discovered links between the resources enable applications such as entity-based search or recommendation.

C3 **Knowledge generation by enrichment**: A common application of KGs is the use of external sources to enrich the company's resources. To this end, openly available KGs such as DBpedia or Wikidata are used. These graphs provide rich, crowd-sourced, common knowledge. Entities that already exist in the company's KG can be enriched by adding additional properties (e.g. adding players' nicknames, labels, numbers, etc., cf. Section 4.2) available in the external sources.

From a KM perspective, it can be argued that KGs can foster the externalisation process of implicit knowledge in organisations. According to Nonaka [25], externalisation plays an important role along the knowledge generation process. Based on the idea that knowledge is defined as the capacity to act [29, 53], an improved externalisation expands the capacity to act as the level of the explicit knowledge increases.

Still, for other definitions of knowledge (e.g. the justified true belief) it is questionable, whether a machine can truly generate knowledge in these terms. This is because some definitions require a human (action) to make sense of a situation or information to be provided in order to be considered knowledge. However, even following those human mind centered definitions of knowledge, there is no contradiction with our argument that KGs can facilitate the externalisation process of knowledge (i.e. making tacit knowledge explicit). Our understanding of knowledge only influences the point in time, for which we can say that new knowledge has been created. Based on a human centered definition we acquire knowledge at the time when a human makes sense of the information externalised through the KG. However, following the capacity to act approach, we acquire knowledge as soon as it is made explicit and provides the basis for action.

While our current literature study is limited in terms of case studies and surveyed publications, we identified current reported uses of KGs to fall into three main categories: (i) *knowledge generation by restructuring*, (ii) *knowledge generation by integration*, and (iii) *knowledge generation by enrichment*. In this paper we can only focus on what companies published (in short industry track papers). Therefore, we may not claim to give a full overview of current applications of KGs in organisations. Yet, we can conclude that these applications are limited to a narrow view on how KGs can address Knowledge Creation. That is, the use cases suggest that KGs – while contributing *to some extent* to knowledge creation and transfer – still offer more potential, when this question is considered in a more holistic manner. Indeed, there is hardly any work, which reports and investigates how the applications of KGs have increased the *capacity to act*.

We claim that herein a KM perspective, as we have introduced it in the first part of the present paper, may be worthwhile to explore. To this end, we close with a set of three directions to be taken into account in further research about Knowledge Graphs:

D1 **Actionable Knowledge:** The current focus in KG research still lies mostly in organizing rather static, factual "knowledge". In this regard we may question in how far it actually represents all facets of knowledge: what could a converging definition of knowledge between communities look like and, particularly, in how far do KGs address the capacity to act? In order to answer this question, we particularly lack definitions and examples of *actionable* knowledge graphs, which also encode process or behavioural knowledge, norms, etc. We note that the Semantic Web or also the Multi-Agent Systems community[11] have made various contributions in this respect in the past (e.g. in terms of attempting to define and researching ontologies on processes, norms and policies, or on communicative action models, as well as simulations of social behaviours under considerations of social norms, etc.). Yet, little of this work has been transferred to the "paradigm" shift to KGs, nor has yielded practical applications that could claim to have contributed to the respective knowledge creation at scale.

D2 **Embodiment of Knowledge:** The role of humans in the loop and the extend to which technological

[11]Example venues that have published articles in this area include the ongoing Conference series on Autonomous Agents & Multiagent Systems (AAMAS), `https://dblp.org/db/conf/atal`, but also seemingly discontinued (though hopefully not forgotten) venues such as the workshops in Computational Logic in Multi-Agent Systems (CLIMA), `https://dblp.org/db/conf/clima/`, Cooperative Information Agents `https://dblp.org/db/conf/cia/`

means (incl. KGs) can actually assist humans is not central enough, or is largely restricted to "human computation", rather than perceiving human cognition and communication as a central entity of knowledge creation itself. Several works emphasise on humans in the loop for KG creation [54]. However, we need further work on how created and collected knowledge can be collected and interfaced between machines and humans to make knowledge generation more effective. To this end, we need to better understand how human knowledge workers communicate with and about data, and what kinds of information (often dubbed background "knowledge") and signals (again, social behaviour and processes) they leverage to achieve mutual understanding.

D3 **Organisational vs. Individual Knowledge:** Referring back to our examples, we see that the creation and transfer of knowledge needs to consider an organisational dimension. All three prototypical cases – (i) the integration of knowledge within an organisation, (ii) making knowledge explicit, accessible and actionable within and across communities and organisations, or (iii) enrichment of organisational knowledge with external/common knowledge – centrally consider community boundaries for knowledge sharing. It is therefore crucial to look at the role of KGs as an enabler of the sharing and generation of knowledge not only at the individual level, but especially at the process of knowledge sharing between organizations/communities. In future works, we will have to investigate more closely, at which of these ontological levels KGs could be used.

To conclude and give a pointed answer to our research question, we argue that there can in fact be "real" knowledge in a KG. However,the exact form of it still requires further research on the identified categories of knowledge. Overall the combination of KM and KG appears as a promising approach to depict an organizations knowledge through the use of a KG and to foster organisational knowledge creation.

References

[1] *Knowledge Graphs: New Directions for Knowledge Representation on the Semantic Web (Dagstuhl Seminar 18371)*, vol. 8, 2018.

[2] A. Mohamed, G. Abuoda, A. Ghanem, Z. Kaoudi, and A. Aboulnaga, "Rdfframes: Knowledge graph access for machine learning tools," *Proc. VLDB Endow.*, vol. 13, no. 12, pp. 2889–2892, 2020.

[3] F. Akrami, M. S. Saeef, Q. Zhang, W. Hu, and C. Li, "Realistic re-evaluation of knowledge graph completion methods: An experimental study," in *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*, SIGMOD '20, (New York, NY, USA), p. 1995–2010, Association for Computing Machinery, 2020.

[4] L. Bellomarini, E. Sallinger, and G. Gottlob, "The vadalog system: Datalog-based reasoning for knowledge graphs," *Proc. VLDB Endow.*, vol. 11, no. 9, pp. 975–987, 2018.

[5] Y. Zhang, Q. Yao, and L. Chen, "Searching recurrent architecture for path-based knowledge graph embedding," in *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, 6-12 December 2020*, 2020.

[6] H. Ren and J. Leskovec, "Beta embeddings for multi-hop logical reasoning in knowledge graphs," in *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, 6-12 December 2020*, 2020.

[7] Z. Zhang, J. Cai, and J. Wang, "Duality-induced regularizer for tensor factorization based knowledge graph completion," in *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, 6-12 December 2020*, 2020.

[8] M. Bevilacqua and R. Navigli, "Breaking through the 80% glass ceiling: Raising the state of the art in word sense disambiguation by incorporating knowledge graph information," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 2854–2864, Association for Computational Linguistics, July 2020.

[9] H. Zhang, Z. Liu, C. Xiong, and Z. Liu, "Grounded conversation generation as guided traverses in commonsense knowledge graphs," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, (Online), pp. 2031–2043, Association for Computational Linguistics, July 2020.

[10] A. Saxena, A. Tripathi, and P. Talukdar, "Improving multi-hop question answering over knowledge graphs using knowledge base embeddings," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, (Online), pp. 4498–4507, Association for Computational Linguistics, July 2020.

[11] I. Tiddi, F. Lécué, and P. Hitzler, eds., *Knowledge Graphs for eXplainable Artificial Intelligence: Foundations, Applications and Challenges*, vol. 47 of *Studies on the Semantic Web*. IOS Press, 2020.

[12] E. T. Penrose, *The Theory of the Growth of the Firm*. Oxford University Press, 2009. 4th edition, originally published 1959.

[13] G. D. Bhatt, "Knowledge management in organizations: examining the interaction between technologies, techniques, and people," *Journal of knowledge management*, 2001.

[14] R. M. Grant, "Toward a knowledge-based theory of the firm," *Strategic management journal*, vol. 17, no. S2, pp. 109–122, 1996.

[15] J.-C. Spender, "Making knowledge the basis of a dynamic theory of the firm," *Strategic management journal*, vol. 17, no. S2, pp. 45–62, 1996.

[16] M. Alavi and D. E. Leidner, "Knowledge management and knowledge management systems: Conceptual foundations and research issues," *MIS quarterly*, pp. 107–136, 2001.

[17] I. Nonaka and G. Von Krogh, "Perspective—tacit knowledge and knowledge conversion: Controversy and advancement in organizational knowledge creation theory," *Organization science*, vol. 20, no. 3, pp. 635–652, 2009.

[18] S. Y. Choi, H. Lee, and Y. Yoo, "The impact of information technology and transactive memory systems on knowledge sharing, application, and team performance: A field study," *MIS quarterly*, vol. 34, no. 4, pp. 855–870, 2010.

[19] T. H. Davenport, D. W. De Long, and M. C. Beers, "Successful knowledge management projects," *Sloan management review*, vol. 39, no. 2, pp. 43–57, 1998.

[20] I. Nonaka, R. Toyama, and N. Konno, "Seci, ba and leadership: a unified model of dynamic knowledge creation," *Long range planning*, vol. 33, no. 1, pp. 5–34, 2000.

[21] K. Venkitachalam and P. Busch, "Tacit knowledge: review and possible research directions," *Journal of knowledge management*, 2012.

[22] R. Fagin, Y. Moses, J. Y. Halpern, and M. Y. Vardi, *Reasoning about knowledge*. MIT Press, 2003.

[23] R. R. Nelson, *An evolutionary theory of economic change*. Harvard University Press, 2009.

[24] J.-B. P. Faucher, A. M. Everett, and R. Lawson, "Reconstituting knowledge management," *Journal of knowledge management*, 2008.

[25] I. Nonaka, R. Toyama, and N. Konno, "Seci, ba and leadership: a unified model of dynamic knowledge creation," *Long range planning*, vol. 33, no. 1, pp. 5–34, 2000.

[26] G. Von Krogh, K. Ichijo, I. Nonaka, *et al.*, *Enabling knowledge creation: How to unlock the mystery of tacit knowledge and release the power of innovation*. Oxford University Press on Demand, 2000.

[27] E. L. Gettier, "Is justified true belief knowledge?," *Analysis*, vol. 23, no. 6, pp. 121–123, 1963.

[28] K. L. Schuh and S. A. Barab, "Philosophical perspectives," *Handbook of research on educational communications and technology*, pp. 67–82, 2008.

[29] K.-E. Sveiby, "A knowledge-based theory of the firm to guide in strategy formulation," *Journal of intellectual capital*, vol. 2, no. 4, pp. 344–358, 2001.

[30] M. Alavi and D. E. Leidner, "Knowledge management and knowledge management systems: Conceptual foundations and research issues," *MIS quarterly*, pp. 107–136, 2001.

[31] H. Takeuchi and I. Nonaka, *Hitotsubashi on knowledge management*. Wiley, 2004.

[32] L. Fahey and L. Prusak, "The eleven deadliest sins of knowledge management," *California management review*, vol. 40, no. 3, pp. 265–276, 1998.

[33] I. Nonaka and H. Takeuchi, *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. Oxford University Press, 1995.

[34] I. Virtanen, "In search for a theoretically firmer epistemological foundation for the relationship between tacit and explicit knowledge.," *Electronic Journal of Knowledge Management*, vol. 11, no. 2, 2013.

[35] T. H. Davenport, L. Prusak, *et al.*, *Working knowledge: How organizations manage what they know*. Harvard Business Press, 1998.

[36] T. Buzan and B. Buzan, *The mind map book*. Pearson Education, 2006.

[37] J. D. Novak, D. Bob Gowin, and G. T. Johansen, "The use of concept mapping and knowledge vee mapping with junior high school science students," *Science education*, vol. 67, no. 5, pp. 625–645, 1983.

[38] E. Marchi and O. Miguel, "On the structure of the teaching-learning interactive process," *International Journal of Game Theory*, vol. 3, no. 2, pp. 83–99, 1974.

[39] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. de Melo, C. Gutierrez, J. E. L. Gayo, S. Kirrane, S. Neumaier, A. Polleres, R. Navigli, A. N. Ngomo, S. M. Rashid, A. Rula, L. Schmelzeisen, J. F. Sequeda, S. Staab, and A. Zimmermann, "Knowledge graphs," *ACM Computing Surveys (to appear)*, vol. abs/2003.02320, 2021.

[40] L. Ehrlinger and W. Wöß, "Towards a definition of knowledge graphs," in *Proceedings of the Posters and Demos Track of the 12th International Conference on Semantic Systems (SEMANTiCS 2016)*, vol. 1695 of *CEUR Workshop Proceedings*, 2016.

[41] H. Paulheim, "Knowledge graph refinement: A survey of approaches and evaluation methods," *Semantic Web*, vol. 8, no. 3, pp. 489–508, 2017.

[42] M. Kejriwal, C. Knoblock, and P. Szekely, *Knowledge Graphs: Fundamentals, Techniques, and Applications*. Adaptive computation and machine learning series, MIT Press, 2021.

[43] V. Janev, D. Graux, H. Jabeen, and E. Sallinger, eds., *Knowledge Graphs and Big Data Processing*, vol. 12072 of *Lecture Notes in Computer Science*. Springer, 2020.

[44] D. Fensel, U. Simsek, K. Angele, E. Huaman, E. Kärle, O. Panasiuk, I. Toma, J. Umbrich, and A. Wahler, *Knowledge Graphs - Methodology, Tools and Selected Use Cases*. Springer, 2020.

[45] R. Cyganiak, D. Wood, and M. Lanthaler, "RDF 1.1 Concepts and Abstract Syntax," w3c recommendation, W3C, Feb. 25 2014.

[46] S. Harris, A. Seaborne, and E. Prud'hommeaux, "SPARQL 1.1 Query Language," w3c recommendation, w3c, Mar. 21 2013.

[47] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. van Kleef, S. Auer, and C. Bizer, "Dbpedia - A large-scale, multilingual knowledge base extracted from wikipedia," *Semantic Web*, vol. 6, no. 2, pp. 167–195, 2015.

[48] G. A. Miller, *WordNet: An electronic lexical database*. MIT Press, 1998.

[49] H. Paulheim, "Knowledge graph refinement : a survey of approaches and evaluation methods," *Semantic web*, vol. 8, no. 3, pp. 489–508, 2017.

[50] J. Strötgen, T.-K. Tran, A. Friedrich, D. Milchevski, F. Tomazic, A. Marusczyk, H. Adel, D. Stepanova, F. Hildebrand, and E. Kharlamov, "Towards the bosch materials science knowledge base.," in *ISWC Satellites*, pp. 323–324, 2019.

[51] P. Cotroneo, W. Haak, G. Oscares, E. Presani, A. Rohatgi, and P. T. Groth, "From data search to data showcasing: The role of semantic technologies in a new service.," in *International Semantic Web Conference (P&D/Industry/BlueSky)*, 2018.

[52] N. Ireson and F. Ciravegna, "Football whispers: Transfer rumour detection," in *CEUR Workshop Proceedings*, vol. 1963, CEUR Workshop Proceedings, 2017.

[53] N. Stehr, "Practical knowledge," *The Blackwell Encyclopedia of Sociology*, 2007.

[54] M. Sabou, E. Simperl, E. Blomqvist, P. Groth, S. Kirrane, G. de Melo, B. Mons, H. Paulheim, L. Pintscher, V. Presutti, J. F. Sequeda, and C. M. Shimizu, "Human and social factors in knowledge graphs," [1], pp. 81–84.