From Genesis to Maturity: Managing Knowledge Graph Ecosystems Through Life Cycles

Sandra Geisler
RWTH Aachen University
Aachen, Germany
geisler@cs.rwth-aachen.de

Cinzia Cappiello
Politecnico di Milano
Milan, Italy
cinzia.cappiello@polimi.it

Irene Celino
CEFRIEL
Milan, Italy
irene.celino@cefriel.com

David Chaves-Fraga
Universidade de Santiago de Compostela
Santiago de Compostela, Spain
david.chaves@usc.es

Anastasia Dimou
KU Leuven
Leuven, Belgium
anastasia.dimou@kuleuven.be

Ana Iglesias-Molina
Universidad Politécnica de Madrid
Madrid, Spain
ana.iglesiasm@upm.es

Maurizio Lenzerini
Sapienza University of Rome
Rome, Italy
lenzerini@diag.uniroma1.it

Anisa Rula
University of Brescia
Brescia, Italy
anisa.rula@unibs.it

Dylan Van Assche
IDLab, Ghent University – imec
Ghent, Belgium
dylan.vanassche@ugent.be

Sascha Welten
RWTH Aachen University
Aachen, Germany
welten@dbis.rwth-aachen.de

Maria-Esther Vidal
TIB Hannover
L3S Research Center
Leibniz University Hannover
Hannover, Germany
vidal@l3s.de

ABSTRACT

Knowledge graphs (KGs) play a crucial role in the integration and organization of heterogeneous data and knowledge, enabling advanced data analytics and decision-making across various industries. This vision paper addresses critical challenges in managing KGs, emphasizing their relevance in integrating information from disparate sources. We propose the concept of knowledge graph ecosystems and life cycles to systematically manage tasks, e.g., data integration, standardization, continuous updates, efficient querying, and provenance tracking. By adopting our approach, organizations can enhance the accuracy, consistency, and reliability of KGs, thus improving knowledge management, enabling the extraction of valuable insights, and ensuring transparency and accountability.

PVLDB Reference Format:
Sandra Geisler, Cinzia Cappiello, Irene Celino, David Chaves-Fraga, Anastasia Dimou, Ana Iglesias-Molina, Maurizio Lenzerini, Anisa Rula, Dylan Van Assche, Sascha Welten, and Maria-Esther Vidal. From Genesis to Maturity: Managing Knowledge Graph Ecosystems Through Life Cycles. PVLDB, 18(5): 1390 - 1397, 2025.
doi:10.14778/3718057.3718067

Proceedings of the VLDB Endowment, Vol. 18, No. 5 ISSN 2150-8097.
doi:10.14778/3718057.3718067

1 INTRODUCTION

Nowadays, sharing of high-quality data within data ecosystems is essential to fostering collaboration, efficiency, innovation, and competitiveness among ecosystem stakeholders [39]. However, data ecosystems, e.g., in healthcare and biomedical research, are highly complex, involving a wide range of stakeholders and critical information is often dispersed across multiple, disparate sources. This fragmentation complicates access to the data necessary for generating insights and enabling advanced applications [10]. In such ecosystems, data is inherently heterogeneous, multi-modal, voluminous, and sensitive, presenting challenges related to interoperability and reusability. Additionally, the knowledge needed to describe and contextualize this data is often fragmented, potentially ambiguous, and distributed across extensive ontologies and taxonomies, which often lack mappings between them. A substantial amount of knowledge often remains implicit, captured only in individual expertise and not documented [34]. Therefore, the challenge of harnessing distributed data and knowledge is significant, complex, and extends well beyond the healthcare domain [1].
Knowledge graphs (KGs) provide a robust solution by integrating data from disparate sources into a cohesive data structure providing unified knowledge and data. This integration enables comprehensive insights across data ecosystems. Many articles [27, 42, 47] and books [19, 32] have focused on KGs, providing different definitions. However, a single, universally accepted definition for KGs still does not exist. This paper considers a KG as “a graph of data intended to accumulate and convey knowledge about the real world, whose

nodes represent entities of interest and whose edges represent relationships between these entities." [27]. However, building and maintaining an effective KG is non-trivial. Considering its complete ecosystem, including data sources, ontologies, and constraints requires well-developed approaches to support the entire KG life cycle within an organization or domain. For KGs to remain accurate, comprehensive, and up-to-date, defining and implementing mature technical and organizational processes is decisive.

1.1 Requirements Analysis

The challenges outlined in the introduction are prototypical for applications and data ecosystems. To identify requirements for a general KG life cycle representation, we have gathered experiences from a multitude of projects, specifically in the healthcare domain [12, 17, 20, 22, 46, 53], and also in other domains, such as manufacturing [43], and from a seminar with leading international researchers in the field [14]. In the following, we delineate the requirements along a concrete representative healthcare example. The example describes an application of a federated analysis supporting the diagnosis of leukodystrophy [59]–a rare genetic disorder causing movement and sensory perception disturbances. This application analyzes data from three healthcare organizations using distributed analysis frameworks. The application is not limited to a specific disease but could be applied to any other illness. The frameworks are based on the idea of trains (program code for analysis) which visit stations (the health organizations), querying and analyzing the local data without transferring data outside an organization. In this scenario, elicited requirements are the following:

- **Facilitate Collaboration While Preserving Privacy.** Enable organizations to securely collaborate on data analysis without transferring sensitive patient information. This approach ensures compliance with privacy regulations and data sovereignty.
- **Achieve Interoperability and Unified Knowledge.** Ensure systems can consistently interpret diverse data formats using shared vocabularies while maintaining up-to-date knowledge that reflects evolving clinical guidelines, records, and research.
- **Ensure Transparent, Accountable, and High-Quality Data Use.** Track all data usage, updates, and analysis processes for trust, reproducibility, and compliance while guaranteeing data and metadata accuracy, completeness, and reliability.
- **Support Tailored and Efficient Actor Interactions.** Provide role-specific tools and workflows for different stakeholders (henceforth referred to as *actors*), such as clinicians, data scientists, and administrators, enabling efficient and actionable insights to support real-time decision-making in critical scenarios.
- **Discover Patterns Across Cross-Organization Data.** Enable advanced reasoning to uncover hidden patterns and relationships across distributed datasets, such as correlations between genetic markers, symptoms, and patient treatment outcomes.

Satisfying these requirements poses significant challenges, including maintaining data integrity, traceability, and accessibility over time [22]. These challenges emphasize the necessity for well-defined life cycles and structured frameworks [5]. Such an approach will enable organizations to effectively manage and maintain the quality and reliability of their KGs by formalizing these concepts and implementing clear processes for data integration, KG updates, adherence to standards, efficient querying, and provenance tracking. Additionally, this formalization facilitates the identification of diverse user needs, supporting varied interactions and decision-making workflows. Such a structured approach enhances knowledge graphs' robustness, adaptability, and transparency, while maximizing their utility and trustworthiness. Subsequently, we will discuss existing definitions of KG life cycles, the steps and tasks in a life cycle, and the evolution of knowledge graphs.

1.2 Related Work

Recent publications have explored approaches to KG development processes [51] and life cycles [15, 49, 62]. These studies outline steps for managing KGs, including construction [24, 38, 55, 60, 64], refinement [41], completion [48], quality assessment [58, 61, 63], and storing and querying [3]. For example, Yip and Sheth [62] propose a six-step life cycle model: (i) design and requirements scoping, (ii) data ingestion, (iii) data enrichment, (iv) storage, (v) consumption, and (vi) maintenance, alongside a platform to implement these steps. Similarly, Cimmino and García-Castro [15] and Simsek et al. [49] describe four life cycle phases—KG creation, hosting, curation, and deployment—with Cimmino and García-Castro [15] introducing the Helio framework based on elicited requirements.

Beyond focusing on specific KG life cycle steps, prior work has addressed Linked Data life cycles [4, 45] and even standardization initiatives [28]. However, these contributions often focus on limited tasks or address specific scenarios, leaving gaps in the comprehensive formalization of KG ecosystems and life cycles. Moreover, existing approaches lack mechanisms to comprehensively specify actors, roles, requirements, and constraints across life cycle steps.

Some works address the evolution of KGs [44, 50], categorizing them into temporal KGs (valid statements within a time range), static or versioned KGs (snapshots of KGs), and dynamic KGs (atomic changes, e.g., Wikidata [57]). Evolution studies have focused on analyzing structural changes [36], developing tools [25], and proposing methodologies [40]. Recent efforts [16, 21, 35, 56] enable incremental updates to KGs but do not provide robust methods for maintaining change provenance or ensuring consistency across interconnected, dynamic, and evolving components.

Despite these efforts, a formal and general framework for KG ecosystems and life cycles is lacking. Our work addresses this gap by proposing a comprehensive formalization that integrates life cycle management with actor roles, tasks, and constraints. Such formalization is both rigorous and practical for real-world applications. In fact, in this way, we ensure that responsibilities are explicit, making it easier to identify issues and enforce accountability. The proposed approach also incorporates mechanisms for tracing changes in order to help stakeholders understand the evolution of the system and facilitate debugging or refinement. In summary, the formalization aims to manage complexity, aligning roles with tasks, and ensuring compliance with the constraints improving the system reliability.

1.3 Research questions

Based on the requirements and existing approaches, we identify two key research questions: **RQ1)** Who are the actors in KG ecosystems, and what roles, tasks, and needs do they have regarding life cycle management? This question is addressed in section 2. **RQ2)**

What are the fundamental components within a KG ecosystem that encompass KG life cycles, and how do these components interact with each other? By defining KG ecosystems and their life cycles we tackle this question in section 3. Both conceptualizations are grounded in the co-authors' extensive experience with data spaces across domains, including health [2, 7, 10, 11, 18, 54], energy [30, 31], manufacturing [8, 23, 43], science [6], and mobility [13].

2 ACTORS, ROLES, AND TASKS

This section discusses which *roles* can take part in a KG Ecosystem (KGE). The roles are played by *actors* which comprise individuals responsible for performing or overseeing the execution of services within the KGE. We base our work on the three major roles identified by Li et al. [37]: *KG Builders*, *KG Analysts*, and *KG Consumers*. We adapt these roles to KGEs, such that they are not only restricted to KGs (i.e., *Knowledge Builders* instead of *KG Builders*), and introduce two additional actors: *Knowledge Providers* and *Knowledge Auditors*. Each actor may assume more than one role across different KG life cycle steps. Figure 1 shows which actors intervene in KGEs, what role(s) they play, their tasks, and their needs.

Knowledge Providers bring domain expertise to the KGE. They provide input not only on the KGE subject matter (i.e., as domain-knowledge experts, such as engineers in a manufacturing scenario), but also on the data, required regulations, and knowledge engineering aspects. They define the needs and tasks of what the KGE must fulfill to (i) specify the requirements for the *Knowledge Builders*, (ii) comply with the *Knowledge Auditors* requirement, and (iii) ensure that the needs of the *Knowledge Consumers* and *Knowledge Analysts* are met. These needs and requirements are defined in Section 3 as part of a life cycle step. *Knowledge Providers* then require tools for a seamless communication with the rest of the actors, e.g., communication and visualization tools, and for collecting and sharing their knowledge as input for the KGE.

Knowledge Builders are responsible for integrating and maintaining the knowledge provided by the *Knowledge Providers*, ensuring that the generated knowledge is compliant with the defined constraints of the corresponding life cycle step. This group comprises experts in KGE-related technologies, such as knowledge engineers and application developers. Their output must be up to the coverage and quality standards of the *Knowledge Auditor*, and be appropriate for its use by *Knowledge Consumers* and *Knowledge Analysts* given the needs, requirements, and constraints. Therefore, *Knowledge Builders* must report, document, and provide provenance traces for all processed and produced resources.

Knowledge Auditors assess the KGE in terms of quality and compliance with the requirements, needs, and constraints. This task is mainly performed by domain-knowledge, domain-data, and domain-regulation experts, with the intervention of *Knowledge Builders*. They define the metrics for evaluation depending on the KGE's needs, regulations, and corresponding requirements. Their efforts serve to validate and improve the KGE's quality.

Knowledge Analysts directly interact with the KGE to extract insights from it. These actors are usually data scientists, ML/AI experts, or app developers. They are not necessarily knowledge engineering experts but possess the skills to interact, extract information, and support discovery in the KGE. The output of their services is then provided to *Knowledge Consumers*, and *Knowledge Auditors* to verify that their needs are fulfilled.

Knowledge Consumers are the end-users of a KGE. They do not usually interact directly with the KGE, so they are not required to have technical skills and tend to utilize user-friendly interfaces. They need documentation, reports, and interfaces for consuming the KGE, and communicate whether the KGE meets their needs and requirements. After we have detailed which roles and actors exist in a KGE, we will subsequently define KGEs and their life cycles.

3 KG ECOSYSTEMS - FUNDAMENTAL OPERATIONAL COMPONENTS

KG Ecosystem A triple $KGE = (D, O, M, DC, KG, L)$ represents a KG ecosystem with the following fundamental components:

- *Data Sources*: D is a set of data sources, where each source has a schema ($\theta(ds)$) defining its structure and attributes, and instances ($\alpha(ds)$) representing the data organized according to the schema.
- *Ontology*: O is a logical theory that defines entities and relationships in the domain using a structured vocabulary.
- *Mappings*: M is a set of assertions linking the ontology (O) to the data sources (D), defining attributes and relations of each data source ($\theta(ds)$) in terms of concepts in O.
- *Constraints*: DC is a theory expressed in a formal language. These constraints ensure the consistency, accuracy, and quality of the data for all the components of the ecosystem.
- *Knowledge Graph*: KG can be empty or the rendering of the ontology O with individuals generated by data collected from the sources described by D based on mapping assertions in M.
- *Log*: L is an ordered list of entries ensuring traceability and auditability. Each entry includes a timestamp, the data state before and after a life cycle step, and a description of changes. The log tracks modifications, supports data provenance, and ensures transparency and accountability in the KGE [26].

Example 3.1. We employ the healthcare example of distributed analysis from the introduction. In this example, the data is coming from diverse sources at each hospital in the data ecosystem. To harmonize, enrich, and efficiently query the data, for each hospital, the transfer and integration of the data into a KG is targeted, and we will refer to this example Health KGE as $HKGE$.

Life Cycles. A KGE undergoes *life cycles*, comprising a series of ordered steps and potential sub-cycles; they manage the creation, validation, curation, maintenance, traversal, and analysis of KGE components. Each step follows a defined partial order, ensuring systematic execution and progression. A life cycle (LC) is represented as a partial order (LCS, R), where LCS is a set of life cycle steps and R is a precedence relation that is reflexive, anti-symmetric, and transitive. A *life cycle* is defined inductively as follows: A life cycle $LC = (LCS, R)$ consists of a set of steps LCS and a precedence relation R. If $LC' = (LCS', R')$ is an existing life cycle: (i) For a single step lcs, $LC = (\{lcs\}, \{(lcs, lcs)\})$, where R contains the reflexive pair (lcs, lcs). (ii) Adding a new step lcs to LC' results in $LC = (LCS' \cup \{lcs\}, R)$, where R extends R' with pairs (lcs, lcs') or (lcs', lcs) for each $lcs' \in LCS'$, indicating whether lcs' precedes or succeeds lcs. This inductive definition allows life cycle steps to be

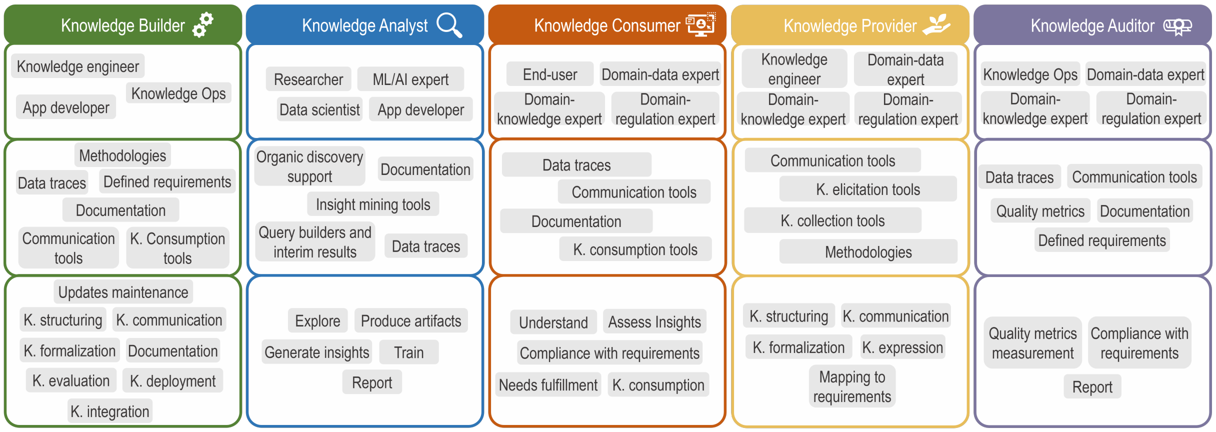
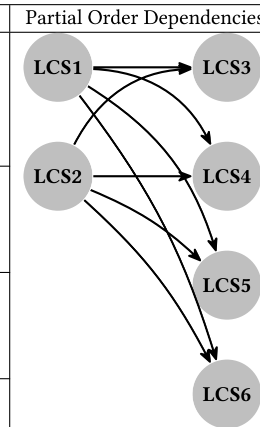


Figure 1: Actors involved in KGEs, with their corresponding tasks, needs, and roles they can play.

Table 1: The six life cycle steps and relationships between them on the example of a HKGE

Life Cycle Steps	Tasks in the HKGE	Partial Order Dependencies
Creation (LCS1)	Extract data from diverse sources (D), including structured data (e.g., relational patient master data), semi-structured data (e.g., FHIR HL7 records), and unstructured data (e.g., free text from questionnaires). Harmonize schemas, resolve conflicts in data formats, and integrate data into KG using mapping assertions (M) aligning the source data with the ontology. Ensure that the integration process preserves semantic consistency and supports subsequent reasoning and analytics tasks within the ecosystem.	
Ontology Evolution and Maintenance (LCS2)	Provide an overarching ontology (O) integrating knowledge from diverse ontologies and nomenclatures. Semantic alignment, i.e., create mappings between vocabularies, as data sources may reference conflicting vocabularies. Update ontology upon, e.g., new examination types to accommodate new concepts and domain-specific knowledge. Ensure consistency between the ontology (O) and KG.	
Validation (LCS3)	Validate potentially incomplete and inaccurate data against domain-specific constraints (DC) to ensure quality and compliance with standards. Perform data cleaning, deduplication, entity resolution, and normalization to improve quality and consistency. Harmonize syntactic and semantic representations across ontologies, mappings, constraints, and KGs in KGEs.	
Querying (LCS4)	Perform measurement analytics using KGs to assemble statistics about patient cohorts at each site, such as age distributions. Support decision-making by providing insights derived from data and knowledge, e.g., identifying cohorts, predicting outcomes, and recommending treatments.	
Monitoring and Feedback (LCS5)	Monitor performance and usage of the HKGE components, especially user access and updates to the data (provenance). Collect detailed feedback from developers, researchers, and end-users to assess the system's functionality, usability, and adaptability to evolving requirements.	
Optimization and Scaling (LCS6)	Optimize the performance of the HKGE components to handle large-scale data and complex analytics tasks. Scale the system to integrate growing data volumes and user demands while maintaining efficiency and reliability especially for reasoning and querying of the data.	

added in a systematic order, ensuring that dependencies between steps are respected and the KGE evolves consistently.

Example 3.2. For our example, Table 1 summarizes general life cycle steps, examples from the HKGE for tasks in these steps, and the dependencies between the six life cycle steps. As indicated in the figure, Steps 1 and 2 should be executed before Steps 3, 4, 5, and 6. Thus, the partial order between the life cycle steps enables the management and evolution of a KGE's different components.

Life Cycle Steps. A *life cycle step* is a tuple $lcs = (S, \langle P, Ro, C, Re, N\rangle)$, where $\langle P, Ro, C, Re, N\rangle$ comprises the contextual information that guides the execution of lcs over a KGE.

- *Service*: S implements KG operations (e.g., creation, validation, updating, querying) that modify, curate, or analyze the KGE.
- *Actors*: P is a set of individuals responsible for contributing to the execution of life cycle steps, as detailed in section 2.
- *Roles*: Ro is a logical theory defining roles (e.g., knowledge engineer) responsible for executing services within the KGE.
- *Constraints*: C is a logical theory defining conditions, such as data quality and compliance, to be satisfied during life cycle execution.
- *Requirements*: Re is a logical theory outlining desired outcomes or conditions the life cycle step aims to achieve.
- *Needs*: N tuples consisting of a set of requirements and constraints stated by an actor while playing a role.

When a life cycle step $lcs = (S, \langle P, Ro, C, Re, N\rangle)$ is executed on a $KGE = (D, O, M, DC, KG, L)$, it produces an updated ecosystem $KGE' = (D_1, O_1, M_1, DC_1, KG_1, L_1)$. This process, denoted as $\sigma(lcs, KGE)$, applies the service S to KGE while ensuring that the

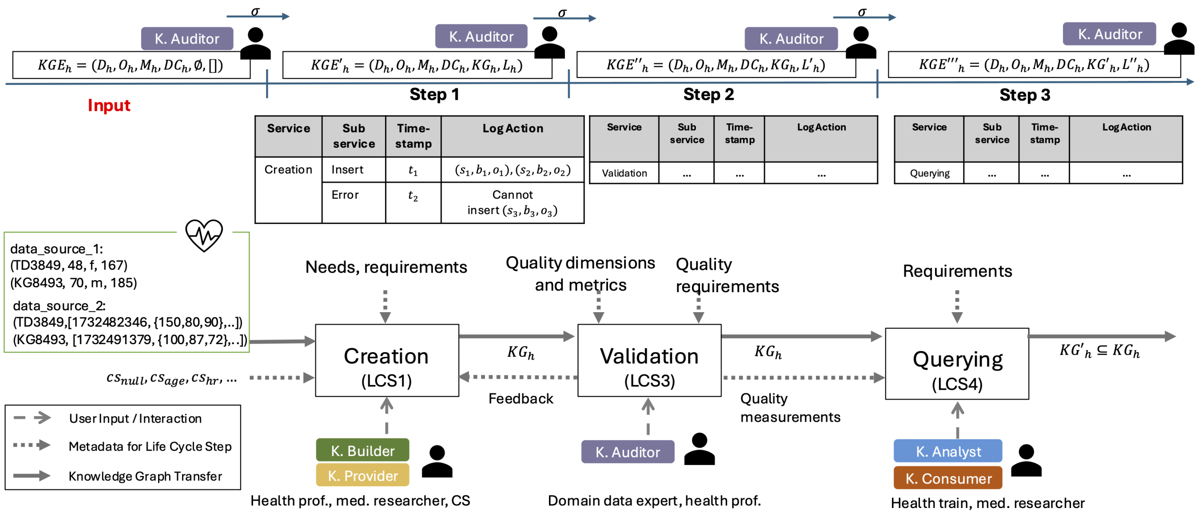


Figure 2: Three life cycle steps for Data Analysis in the example, Creation, Validation, and Querying, compose the life cycle

needs N are satisfied and the constraints C are validated.

Life Cycle Execution. For a life cycle $LC = (LCS, R)$, which consists of a set of life cycle steps LCS and precedence relations R, executing LC over KGE (denoted as $\sigma(LC, KGE)$) involves applying each step in LCS according to R. For a single step lcs, where $LC = (\{lcs\}, \{(lcs, lcs)\})$, executing LC updates the components of KGE: $\sigma(lcs, KGE) = (D_1, O_1, M_1, DC_1, KG_1, L_1)$, where D_1, O_1, M_1, and KG_1 are modified by S based on N and C. The log L_1 records the inputs, outputs, timestamps, and validation results.
If $KGE' = (D_1, O_1, M_1, DC_1, KG_1, L_1)$ is the result of $\sigma(LC', KGE)$ for $LC' = (LCS', R')$ adding lcs to LCS' forms $LC = (LCS' \cup \{lcs\}, R)$. Executing LC over KGE' produces $KGE_2 = \sigma(LC, KGE')$, with updated components D_2, O_2, M_2, DC_2, KG_2, and a log L_2 that extends L_1 with the input, output, and validation results of S. This process ensures that life cycle steps are executed in order, respects defined dependencies, and maintains a complete log for traceability.

Example 3.3. In the healthcare example, we want to create a KG at each hospital and use it as the basis for distributed data analysis. Figure 2 depicts the corresponding life cycle with prototypical steps for (1), (3), and (4) from Table 1. At the top of the figure, KGE_h and its versions along the evolution through the steps are shown. Each version must be validated by one or more KG Auditors, i.e., medical data experts. At the bottom, the life cycle steps with inputs and outputs are delineated. KGE_h is initialized with a set of data sources $D_h = \{ds1, ds2\}$, an ontology O_h, the mapping assertions M_h, a set of domain-specific constraints DC_h, an empty KG, and an empty log list L_h. ds_1 is a patient master data set and ds_2 represents a set of examination (measurement) data. O_h refers to concepts from medical ontologies, e.g., the Human Phenotype Ontology[1]. DC_h contains multiple constraints, e.g., cs_{null}, postulating that core attributes, such as sex, must not be null. KGE_h is the input to the first life cycle step - the *Creation* step.

Creation (LCS1) : Actors involved in this step are the Knowledge Builder and the Knowledge Provider. These can be domain experts, e.g., health professionals. They are in charge of the data acquisition. Computer scientists support importing data and creating the KG from it. Further, the domain experts and computer scientists jointly work on creating and curating the ontology and the mapping assertions between ontology and KG. The data is imported into the KG according to the ontology and the mapping assertions. All operations executed during creation are logged in the log list L_h as exemplified in the table below the Creation step in Figure 2. For example, at time t_1 data items $< s_1, b_1, o_1 >$ and $< s_2, b_2, o_2 >$ are inserted. At time t_2 an error is reported when inserting $< s_3, b_3, o_3 >$. The resulting KG ecosystem $KGE_h = D_h, O_h, M_h, DC_h, KG_h, L_h$ is input to the next step - Validation (Step 2).

Validation (LCS3): In this step Knowledge Auditors are involved, such as domain data experts. Inputs to this step are data quality (DQ) definitions, e.g., DQ dimensions and metrics, and the corresponding goals. In our example, we want to assess the completeness (number of null values) and plausibility (is the age below 110?). Output of the step is a KGE''_h in which all components are unaltered, despite the log L'_h, which contains new entries for the DQ measurement results. KGE''_h is input to the next step.

Querying (LCS4): In this step Knowledge Consumers are involved, e.g., the health train algorithm. They issue a query to the system maintaining KG_h. The requirements of that step could be goals for the response times or accuracy of the query. The querying step will deliver a KGE'''_h, including KG'_h - a subgraph of KG_h - representing the query results The log is extended to L''_h comprising, log entries with some query metadata, e.g., the query statement or the execution time. This creates the final ecosystem KGE'''_{dc}.

4 CHALLENGES AND FUTURE DIRECTIONS

Operationalizing KGEs presents challenges beyond traditional data ecosystems due to their reliance on semantic alignment, dynamic interactions, and diverse stakeholder requirements. Below, we summarize key challenges and strategies for addressing them.

[1] https://hpo.jax.org/

Integration of Heterogeneous Data Sources. KGEs require integrating data from distributed and diverse sources, including structured, semi-structured, and unstructured formats, while aligning schemas and resolving terminological conflicts. Unlike traditional data integration, KGEs must incorporate semantics using ontologies, mappings, and constraints to harmonize data across components. This challenge is amplified in domains like healthcare and energy, where domain-specific standards (e.g., SNOMED-CT [29], HL7-FHIR [9]) introduce additional semantic heterogeneity. Integration processes must also accommodate the dynamic nature of KGEs, where changes in data sources, ontologies, or mappings ripple through the ecosystem, necessitating re-validation to maintain coherence. Addressing this challenge requires ontology-based frameworks, semantic alignment techniques, and automated tools for entity resolution, all of which ensure seamless semantic data integration without compromising domain-particular characteristics.
Supporting Evolving KGs. KGEs operate in dynamic environments where data, ontologies, and user requirements frequently evolve. This creates the need for continuous updates to maintain semantic and logical consistency across the ecosystem. Unlike static KGs, KGEs must accommodate incremental changes while ensuring the ecosystem remains reliable and trustworthy. Solutions to this challenge include pipelines for ontology versioning, consistency checking, and automated updates to mappings and constraints. These tools must enable the propagation of changes across interconnected components while preserving provenance and traceability.
Enabling Interoperability Across Ecosystem Components. The components of a KGE (e.g., data sources, mappings, ontologies, and constraints) must function cohesively to enable ecosystem-wide reasoning and analysis. Achieving interoperability requires harmonizing syntactic formats and semantic meaning, particularly when integrating diverse standards and vocabularies. Unlike traditional systems, which focus primarily on syntactic alignment, KGEs must resolve semantic inconsistencies across domains. Research should focus on developing shared vocabularies, ontology alignment methods, and standardized APIs to enable seamless interaction among components. They would ensure that data and knowledge can be exchanged and utilized within the ecosystem.
Scalability of Ecosystem Operations. As KGEs grow in complexity and data volume, scalability becomes a critical concern. Managing large-scale KGs, performing computationally expensive reasoning tasks, and ensuring efficient query execution are all challenges that increase as the ecosystem expands. Unlike traditional data integration systems, KGEs involve semantic-centric operations (e.g., reasoning and traversal) that require significant computational resources. Distributed graph storage systems, parallelized reasoning engines, and optimized querying techniques are needed to address these scalability challenges without compromising performance.
Ensuring Data and Knowledge Quality. The reliability of KGEs depends on maintaining high-quality data, mappings, and inferred knowledge. Errors or inconsistencies in these elements can undermine trust in the ecosystem, particularly in critical domains like healthcare. Ensuring data and knowledge quality involves validating the ecosystem's components against domain-specific constraints, cleaning and normalizing data, and continuously monitoring for anomalies. Automated validation frameworks, such as those based on SHACL [33] or ShEx [52], and human oversight for complex scenarios are necessary to uphold quality standards.
Defining and Managing KGE Life Cycles. KGEs require clearly defined life cycles to ensure systematic creation, maintenance, and evolution of their components. Unlike standalone KGs, KGEs encompass dynamic workflows where different lifecycle steps—such as data ingestion, validation, reasoning, and update propagation—must be coordinated. Research should formalize KGE lifecycle models to specify the relationships and dependencies between these steps, enabling systematic management and evolution of the ecosystem.
Tracing and Validation Across the Ecosystem. The complexity of KGEs necessitates robust tracing mechanisms to track the provenance of data, mappings, and knowledge across all lifecycle steps. Additionally, validation frameworks are essential to ensure consistency, completeness, and adherence to domain constraints. Unlike traditional systems, tracing and validation in KGEs must span interconnected components, requiring automated tools to track changes and assess their impact on the overall ecosystem.
Supporting Role-Specific Interactions. KGEs serve a diverse range of users, each with unique expertise, roles, and requirements. This diversity necessitates tailored tools and workflows to ensure usability and adoption. Unlike traditional ecosystems, which often cater to generic user needs, KGEs must support context-specific interactions, enabling knowledge providers, builders, auditors, and consumers to perform their tasks efficiently. Designing intuitive user interfaces, implementing role-based access controls, and providing training resources are key strategies to meet this challenge.
Tracing, Validation, and Explainability Across the Ecosystem. KGEs necessitate mechanisms for tracing, validation, and explainability to ensure both functional reliability and stakeholder trust. Tracing involves tracking the provenance of data, mappings, and knowledge across all lifecycle steps, enabling transparency and accountability. Validation frameworks are essential to ensure consistency, completeness, and adherence to domain-specific constraints. Additionally, explainability is critical for supporting trust in KGE-derived insights by providing clear reasoning paths and clarifying how decisions are made. Addressing this requires developing provenance models, automated validation tools, and explainable AI techniques to track changes, assess their impact on the ecosystem, and ensure that users can interpret, trust, and effectively utilize KGE-driven insights in their decision-making processes.
Addressing challenges in lifecycle specification, tracing, and validation is essential for developing robust and efficient KGEs. Establishing these foundations ensures scalability, adaptability, and the delivery of reliable insights across domains.

ACKNOWLEDGMENTS

The authors thank the Dagstuhl team for hosting Seminar 24061 in February 2024, where the initial ideas for this article were developed. This work has been funded by: champI4.0ns (grant 01MJ22011B - BMWK), PNRR Project FAIR (grant: PE0000013 - MUR), PERKS (grant: 101120323 - Horizon Europe), REOPEN (PID2023-149549NB-I00 - AEI-Spain), ED431G 2023/04, ED431C 2022/19 - Xunta-CE, PRIN 2022 (grant: NEXTCART 2022YXXZH5 - MUR), Flanders Make, UGent Special Research Fund (grant: BOF20/DOC/132 - BOF), TrustKG (grant P99/2020 Leibniz Association).

REFERENCES

[1] Awais Ahmed, Rui Xi, Mengshu Hou, Syed Attique Shah, and Sufian Hameed. 2023. Harnessing Big Data Analytics for Healthcare: A Comprehensive Review of Frameworks, Implications, Applications, and Impacts. *IEEE Access* 11 (2023), 112891–112928. https://doi.org/10.1109/ACCESS.2023.3323574

[2] Fotis Aisopos, Samaneh Jozashoori, Emetis Niazmand, Disha Purohit, Ariam Rivas, Ahmad Sakor, Enrique Iglesias, Dimitrios Vogiatzis, Ernestina Menasalvas, Alejandro Rodríguez González, Guillermo Vigueras, Daniel Gómez-Bravo, Maria Torrente, Roberto Hernández López, Mariano Provencio Pulla, Athanasios Dalianis, Anna Triantafillou, Georgios Paliouras, and Maria-Esther Vidal. 2023. Knowledge graphs for enhancing transparency in health data ecosystems. *Semantic Web* 14, 5 (2023), 943–976. https://doi.org/10.3233/SW-223294

[3] Waqas Ali, Muhammad Saleem, Bin Yao, Aidan Hogan, and Axel-Cyrille Ngonga Ngomo. 2022. A survey of RDF stores & SPARQL engines for querying knowledge graphs. *The VLDB Journal* 31 (2022), 1–26. https://doi.org/10.1007/s00778-021-00711-3

[4] Sören Auer, Lorenz Bühmann, Christian Dirschl, Orri Erling, Michael Hausenblas, Robert Isele, Jens Lehmann, Michael Martin, Pablo N. Mendes, Bert van Nuffelen, Claus Stadler, Sebastian Tramp, and Hugh Williams. 2012. Managing the Life-Cycle of Linked Data with the LOD2 Stack. In *The Semantic Web–ISWC 2012: 11th International Semantic Web Conference, Boston, MA, USA, November 11-15, 2012, Proceedings, Part II 11*. Springer Berlin Heidelberg, 1–16.

[5] Sören Auer, Theodore Dalamagas, Helen Parkinson, François Bancilhon, Giorgos Flouris, Dimitris Sacharidis, Peter Buneman, Dimitris Kotzinos, Yannis Stavrakas, Vassilis Christophides, George Papastefanatos, and Kostas Thiveos. 2012. Diachronic linked data: towards long-term preservation of structured interrelated information. In *Proceedings of the First International Workshop on Open Data* (Nantes, France) *(WOD '12)*. Association for Computing Machinery, New York, NY, USA, 31–39. https://doi.org/10.1145/2422604.2422610

[6] Sören Auer, Viktor Kovtun, Manuel Prinz, Anna Kasprzik, Markus Stocker, and Maria-Esther Vidal. 2018. Towards a Knowledge Graph for Science. In *Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics, WIMS 2018, Novi Sad, Serbia, June 25-27, 2018*, Rajendra Akerkar, Mirjana Ivanovic, Sang-Wook Kim, Yannis Manolopoulos, Riccardo Rosati, Milos Savic, Costin Badica, and Milos Radovanovic (Eds.). ACM, 1:1–1:6. https://doi.org/10.1145/3227609.3227689

[7] Carlos Badenes-Olmedo, David Chaves-Fraga, María Poveda-Villalón, Ana Iglesias-Molina, Pablo Calleja, Socorro Bernardos, Patricia Martín-Chozas, Alba Fernández-Izquierdo, Elvira Amador-Domínguez, Paola Espinoza-Arias, Luis Pozo-Gilo, Edna Ruckhaus, Esteban González-Guardia, Raquel Cedazo, Beatriz López-Centeno, and Óscar Corcho. 2020. Drugs4Covid: Drug-driven Knowledge Exploitation based on Scientific Publications. https://arxiv.org/abs/2012.01953

[8] Sebastian R. Bader, Irlán Grangel-González, Priyanka Nanjappa, Maria-Esther Vidal, and Maria Maleshkova. 2020. A Knowledge Graph for Industry 4.0. In *The Semantic Web - 17th International Conference, ESWC 2020, Heraklion, Crete, Greece, May 31-June 4, 2020, Proceedings (Lecture Notes in Computer Science)*, Andreas Harth, Sabrina Kirrane, Axel-Cyrille Ngonga Ngomo, Heiko Paulheim, Anisa Rula, Anna Lisa Gentile, Peter Haase, and Michael Cochez (Eds.), Vol. 12123. Springer, 465–480. https://doi.org/10.1007/978-3-030-49461-2_27

[9] Duane Bender and Kamran Sartipi. 2013. HL7 FHIR: An Agile and RESTful approach to healthcare information exchange. In *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*. IEEE, 326–331. https://doi.org/10.1109/cbms.2013.6627810

[10] Thomas Berlage, Carsten Claussen, Sandra Geisler, Carlos A. Velasco, and Stefan Decker. 2022. Medical Data Spaces in Healthcare Data Ecosystems. In *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage*, Boris Otto, Michael ten Hompel, and Stefan Wrobel (Eds.). Springer, 291–311. https://doi.org/10.1007/978-3-030-93975-5_18

[11] Mathias De Brouwer, Pieter Bonte, Dörthe Arndt, Miel Vander Sande, Anastasia Dimou, Ruben Verborgh, Filip De Turck, and Femke Ongenae. 2024. Optimized continuous homecare provisioning through distributed data-driven semantic services and cross-organizational workflows. *J. Biomed. Semant.* 15, 1 (2024), 9. https://doi.org/10.1186/S13326-024-00303-4

[12] Virginia Calvo, Emetis Niazmand, Enric Carcereny, Delvys Rodriguez-Abreu, Manuel Cobo, Rafael López-Castro, María Guirado, Carlos Camps, Ana Laura Ortega, Reyes Bernabé, Bartomeu Massutí, Rosario Garcia-Campelo, Edel del Barco, José Luis González-Larriba, Joaquim Bosch-Barrera, Marta Martínez, María Torrente, María-Esther Vidal, and Mariano Provencio. 2024. Family history of cancer and lung cancer: Utility of big data and artificial intelligence for exploring the role of genetic risk. *Lung Cancer* 195 (Sept. 2024), 107920. https://doi.org/10.1016/j.lungcan.2024.107920

[13] David Chaves-Fraga, Pieter Colpaert, Mersedeh Sadeghi, and Marco Comerio. 2023. Editorial of transport data on the web. *Semantic Web* 14, 4 (2023), 613–616. https://doi.org/10.3233/SW-223278

[14] David Chaves-Fraga, Oscar Corcho, Anastasia Dimou, Maria-Esther Vidal, Ana Iglesias-Molina, and Dylan Van Assche. 2024. Are Knowledge Graphs Ready for the Real World? Challenges and Perspective (Dagstuhl Seminar 24061). *Dagstuhl Reports* 14, 2 (2024), 1–70. https://doi.org/10.4230/DagRep.14.2.1

[15] Andrea Cimmino and Raúl García-Castro. 2024. Helio: A framework for implementing the life cycle of knowledge graphs. *Semantic Web* 15, 1 (2024), 223–249.

[16] Diego Conde-Herreros, María Poveda-Villalón, Romana Pernisch, Lise Stork, Oscar Corcho, and David Chaves-Fraga. 2024. Propagating Ontology Changes to Declarative Mappings in Construction of Knowledge Graphs. In *Fifth International Workshop on Knowledge Graph Construction (KGCW2024)*, Vol. 3718. CEUR-ws.org, 1–16.

[17] Federico Croce, Riccardo Valentini, Marianna Maranghi, Giorgio Grani, Maurizio Lenzerini, and Riccardo Rosati. 2024. Ontology-Based Data Preparation in Healthcare: The Case of the AMD-STITCH Project. *SN Computer Science* 5, 4 (April 2024), 1–12. https://doi.org/10.1007/s42979-024-02757-w

[18] Federico Croce, Riccardo Valentini, Marianna Maranghi, Giorgio Grani, Maurizio Lenzerini, and Riccardo Rosati. 2024. Ontology-Based Data Preparation in Healthcare: The Case of the AMD-STITCH Project. *SN Comput. Sci.* 5, 4 (2024), 437. https://doi.org/10.1007/S42979-024-02757-W

[19] Dieter Fensel, Umutcan Simsek, Kevin Angele, Elwin Huaman, Elias Kärle, Oleksandra Panasiuk, Ioan Toma, Jürgen Umbrich, and Alexander Wahler. 2020. *Knowledge Graphs*. Springer Cham. https://doi.org/10.1007/978-3-030-37439-6

[20] Álvaro García-Barragán, Ahmad Sakor, Maria-Esther Vidal, Ernestina Menasalvas, Juan Cristobal Sanchez Gonzalez, Mariano Provencio, and Víctor Robles. 2024. NSSC: a neuro-symbolic AI system for enhancing accuracy of named entity recognition and linking from oncologic clinical notes. *Medical & Biological Engineering & Computing* (Nov. 2024), 1–24. https://doi.org/10.1007/s11517-024-03227-4

[21] Gleb Gawriljuk, Andreas Harth, Craig A Knoblock, and Pedro Szekely. 2016. A Scalable Approach to Incrementally Building Knowledge Graphs. In *Research and Advanced Technology for Digital Libraries: 20th International Conference on Theory and Practice of Digital Libraries, TPDL 2016, Hannover, Germany, September 5–9*. Springer Cham, 188–199. https://doi.org/10.1007/978-3-319-43997-6_15

[22] Sandra Geisler, Maria-Esther Vidal, Cinzia Cappiello, Bernadette Farias Lóscio, Avigdor Gal, Matthias Jarke, Maurizio Lenzerini, Paolo Missier, Boris Otto, Elda Paja, Barbara Pernici, and Jakob Rehof. 2022. Knowledge-Driven Data Ecosystems Toward Data Transparency. *ACM Journal of Data and Information Quality (JDIQ)* 14, 1 (2022), 3:1–3:12. https://doi.org/10.1145/3467022

[23] Irlán Grangel-González, Lavdim Halilaj, Maria-Esther Vidal, Omar Rana, Steffen Lohmann, Sören Auer, and Andreas W. Müller. 2018. Knowledge Graphs for Semantically Integrating Cyber-Physical Systems. In *Database and Expert Systems Applications - 29th International Conference, DEXA 2018, Regensburg, Germany, September 3-6, 2018, Proceedings, Part I (Lecture Notes in Computer Science)*, Sven Hartmann, Hui Ma, Abdelkader Hameurlain, Günther Pernul, and Roland R. Wagner (Eds.), Vol. 11029. Springer, 184–199. https://doi.org/10.1007/978-3-319-98809-2_12

[24] Marco Grassi, Mario Scrocca, Alessio Carenini, Marco Comerio, and Irene Celino. 2023. Composable Semantic Data Transformation Pipelines with Chimera. In *Proceedings of the 4th International Workshop on Knowledge Graph Construction (KGCW 2023)*, Vol. 3471. CEUR-WS.org, 1–12.

[25] Michael Hartung, Anika Groß, and Erhard Rahm. 2013. COnto–Diff: generation of complex evolution mappings for life science ontologies. *Journal of Biomedical Informatics* 46, 1 (2013), 15–32. https://doi.org/10.1016/j.jbi.2012.04.009

[26] Melanie Herschel, Ralf Diestelkämper, and Houssem Ben Lahmar. 2017. A survey on provenance: What for? What form? What from? *The VLDB Journal* 26 (2017), 881–906. https://doi.org/10.1007/s00778-017-0486-1

[27] Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia D'amato, Gerard De Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, Axel-Cyrille Ngonga Ngomo, Axel Polleres, Sabbir M. Rashid, Anisa Rula, Lukas Schmelzeisen, Juan Sequeda, Steffen Staab, and Antoine Zimmermann. 2021. Knowledge Graphs. *ACM Comput. Surv.* 54, 4, Article 71 (jul 2021), 37 pages. https://doi.org/10.1145/3447772

[28] Bernadette Hyland, Ghislain Atemezing, and Boris Villazón-Terrazas. 2014. *Best Practices for Publishing Linked Data*. W3C Group Note. World Wide Web Consortium. https://www.w3.org/TR/ld-bp/ (last accessed date: 2025/03/05).

[29] International Health Terminology Standards Development Organisation - IHTSDO. 2014. SNOMED CT. http://www.snomed.org/ (last accessed date: 2025/03/05).

[30] Valentina Janev, Maria-Esther Vidal, Kemele M. Endris, and Dea Pujic. 2021. Managing Knowledge in Energy Data Spaces. In *Companion of The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19-23, 2021*, Jure Leskovec, Marko Grobelnik, Marc Najork, Jie Tang, and Leila Zia (Eds.). ACM / IW3C2, 7–15. https://doi.org/10.1145/3442442.3453541

[31] Valentina Janev, Maria-Esther Vidal, Dea Pujić, Dušan Popadić, Enrique Iglesias, Ahmad Sakor, and Andrej Čampa. 2022. Responsible Knowledge Management in Energy Data Ecosystems. *Energies* 15, 11 (May 2022), 3973. https://doi.org/10.3390/en15113973

[32] Mayank Kejriwal. 2019. *Domain-Specific Knowledge Graph Construction*. Springer Cham. https://doi.org/10.1007/978-3-030-12375-8

[33] Holger Knublauch and Dimitris Kontokostas. 2017. *Shapes constraint language (SHACL)*. Technical Report. W3C. https://www.w3.org/TR/shacl/ (last accessed date: 2025/03/05).

,

[34] Kalle Koivisto and Toni Taipalus. 2023. Pitfalls in Effective Knowledge Management: Insights from an International Information Technology Organization. https://doi.org/10.48550/ARXIV.2304.07737

[35] Nikolaos Konstantinou, Dimitrios-Emmanuel Spanos, Dimitris Kouis, and Nikolas Mitrou. 2015. An Approach for the Incremental Export of Relational Databases into RDF Graphs. *International Journal on Artificial Intelligence Tools* 24, 02 (April 2015), 1540013. https://doi.org/10.1142/s0218213015400138

[36] Konstantinos I. Kotis, George A. Vouros, and Dimitris Spiliotopoulos. 2020. Ontology engineering methodologies for the evolution of living and reused ontologies: status, trends, findings and recommendations. *The Knowledge Engineering Review* 35 (2020), e4. https://doi.org/10.1017/S0269888920000065

[37] Harry Li, Gabriel Appleby, Camelia Daniela Brumar, Remco Chang, and Ashley Suh. 2024. Knowledge Graphs in Practice: Characterizing their Users, Challenges, and Visualization Opportunities. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2024), 584–594. https://doi.org/10.1109/TVCG.2023.3326904

[38] Jixiong Liu, Yoan Chabot, Raphaël Troncy, Viet-Phi Huynh, Thomas Labbé, and Pierre Monnin. 2023. From tabular data to knowledge graphs: A survey of semantic table interpretation tasks and methods. *Journal of Web Semantics* 76 (2023), 100761. https://doi.org/10.1016/j.websem.2022.100761

[39] Boris Otto. 2022. *The Evolution of Data Spaces.* Springer International Publishing, 3–15. https://doi.org/10.1007/978-3-030-93975-5_1

[40] Raúl Palma, Fouad Zablith, Peter Haase, and Oscar Corcho. 2012. *Ontology Evolution.* Springer, 235–255. https://doi.org/10.1007/978-3-642-24794-1_11

[41] Heiko Paulheim. 2017. Knowledge Graph Refinement: A Survey of Approaches and Evaluation Methods. *Semantic Web Journal* 8, 3 (2017), 489–508. https://doi.org/10.3233/SW-160218

[42] Ciyuan Peng, Feng Xia, Mehdi Naseriparsa, and Francesco Osborne. 2023. Knowledge Graphs: Opportunities and Challenges. *Artif. Intell. Rev.* 56, 11 (apr 2023), 13071–13102. https://doi.org/10.1007/s10462-023-10465-9

[43] Jan Pennekamp, Anastasiia Belova, Thomas Bergs, Matthias Bodenbenner, Andreas Bührig-Polaczek, Markus Dahlmanns, Ike Kunze, Moritz Kröger, Sandra Geisler, Martin Henze, Daniel Lütticke, Benjamin Montavon, Philipp Niemietz, Lucia Ortjohann, Maximilian Rudack, Robert H. Schmitt, Uwe Vroomen, Klaus Wehrle, and Michael Zeng. 2023. *Evolving the Digital Industrial Infrastructure for Production: Steps Taken and the Road Ahead.* Springer International Publishing, 35–60. https://doi.org/10.1007/978-3-031-44497-5_2

[44] Axel Polleres, Romana Pernisch, Angela Bonifati, Daniele Dell'Aglio, Daniil Dobriy, Stefania Dumbrava, Lorena Etcheverry, Nicolas Ferranti, Katja Hose, Ernesto Jiménez-Ruiz, Matteo Lissandrini, Ansgar Scherp, Riccardo Tommasini, and Johannes Wachs. 2023. How Does Knowledge Evolve in Open Knowledge Graphs? *Transactions on Graph Data and Knowledge* 1, 1 (2023), 11:1–11:59. https://doi.org/10.4230/TGDK.1.1.11

[45] Filip Radulovic, María Poveda-Villalón, Daniel Vila-Suero, Víctor Rodríguez-Doncel, Raúl García-Castro, and Asunción Gómez-Pérez. 2015. Guidelines for Linked Data generation and publication: An example in building energy consumption. *Automation in Construction* 57 (2015), 178–187. https://doi.org/10.1016/j.autcon.2015.04.002

[46] Ariam Rivas, Diego Collarana, Maria Torrente, and Maria-Esther Vidal. 2024. A neuro-symbolic system over knowledge graphs for link prediction. *Semantic Web* 15, 4 (Oct. 2024), 1307–1331. https://doi.org/10.3233/sw-233324

[47] Sherif Sakr, Angela Bonifati, Hannes Voigt, Alexandru Iosup, Khaled Ammar, Renzo Angles, Walid Aref, Marcelo Arenas, Maciej Besta, Peter A. Boncz, Khuzaima Daudjee, Emanuele Della Valle, Stefania Dumbrava, Olaf Hartig, Bernhard Haslhofer, Tim Hegeman, Jan Hidders, Katja Hose, Adriana Iamnitchi, Vasiliki Kalavri, Hugo Kapp, Wim Martens, M. Tamer Özsu, Eric Peukert, Stefan Plantikow, Mohamed Ragab, Matei R. Ripeanu, Semih Salihoglu, Christian Schulz, Petra Selmer, Juan F. Sequeda, Joshua Shinavier, Gábor Szárnyas, Riccardo Tommasini, Antonino Tumeo, Alexandru Uta, Ana Lucia Varbanescu, Hsiang-Yun Wu, Nikolay Yakovets, Da Yan, and Eiko Yoneki. 2021. The Future Is Big Graphs: A Community View on Graph Processing Systems. *Commun. ACM* 64, 9 (2021), 62–71. https://doi.org/10.1145/3434642

[48] Tong Shen, Fu Zhang, and Jingwei Cheng. 2022. A comprehensive overview of knowledge graph completion. *Knowledge-Based Systems* 255 (2022), 109597. https://doi.org/10.1016/j.knosys.2022.109597

[49] Umutcan Simsek, Kevin Angele, Elias Kärle, Juliette Opdenplatz, Dennis Sommer, Jürgen Umbrich, and Dieter Fensel. 2021. Knowledge Graph Lifecycle: Building and Maintaining Knowledge Graphs. In *Second International Workshop on Knowledge Graph Construction (KGCW 2021)*, Vol. 2873. CEUR-ws.org, 1–16.

[50] Ljiljana Stojanovic. 2004. *Methods and tools for ontology evolution.* Ph.D. Dissertation. Karlsruhe Institute of Technology, Germany. https://www.deutsche-digitale-bibliothek.de/item/XL2N7EZMAFFJP52KECFFYGBX4JB3627E (last accessed date: 2025/03/05).

[51] Gytė Tamašauskaitė and Paul Groth. 2023. Defining a Knowledge Graph Development Process Through a Systematic Review. *ACM Transactions on Software Engineering and Methodology* 32, 1 (2023), 1–40. https://doi.org/10.1145/3522586

[52] Katherine Thornton, Harold Solbrig, Gregory S. Stupp, José Emilio Labra Gayo, Daniel Mietchen, Eric Prud'hommeaux, and Andra Waagmeester. 2019. Using Shape Expressions (ShEx) to Share RDF Data Models and to Guide Curation with Rigorous Validation. In *The Semantic Web - 16th International Conference, ESWC 2019, Portorož, Slovenia, June 2-6, 2019, Proceedings (Lecture Notes in Computer Science)*, Pascal Hitzler, Miriam Fernández, Krzysztof Janowicz, Amrapali Zaveri, Alasdair J. G. Gray, Vanessa López, Armin Haller, and Karl Hammar (Eds.), Vol. 11503. Springer, 606–620. https://doi.org/10.1007/978-3-030-21348-0_39

[53] Riccardo Valentini, Eugenio Carrani, Marina Torre, and Maurizio Lenzerini. 2023. *Ontology-Based Data Management in Healthcare: The Case of the Italian Arthroplasty Registry.* Springer Nature Switzerland, 88–101. https://doi.org/10.1007/978-3-031-47546-7_7

[54] Riccardo Valentini, Eugenio Carrani, Marina Torre, and Maurizio Lenzerini. 2023. Ontology-Based Data Management in Healthcare: The Case of the Italian Arthroplasty Registry. In *AIxIA 2023 - Advances in Artificial Intelligence - XXIInd International Conference of the Italian Association for Artificial Intelligence, AIxIA 2023, Rome, Italy, November 6-9, 2023, Proceedings (Lecture Notes in Computer Science)*, Roberto Basili, Domenico Lembo, Carla Limongelli, and Andrea Orlandini (Eds.), Vol. 14318. Springer, 88–101. https://doi.org/10.1007/978-3-031-47546-7_7

[55] Dylan Van Assche, Thomas Delva, Gerald Haesendonck, Pieter Heyvaert, Ben De Meester, and Anastasia Dimou. 2023. Declarative RDF graph generation from heterogeneous (semi-)structured data: A systematic literature review. *Web Semant.* 75, C (jan 2023), 24. https://doi.org/10.1016/j.websem.2022.100753

[56] Dylan Van Assche, Sitt Min Oo, Julian Andres Rojas Melendez, and Pieter Colpaert. 2022. Continuous generation of versioned collections' members with RML and LDES. In *Third International Workshop on Knowledge Graph Construction (KGCW 2022)*, Vol. 3141. CEUR-ws.org, Hersonissos, Greece, 1–8.

[57] Denny Vrandecic and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledgebase. *Commun. ACM* 57, 10 (2014), 78–85. https://doi.org/10.1145/2629489

[58] Xiangyu Wang, Lyuzhou Chen, Taiyu Ban, Muhammad Usman, Yifeng Guan, Shikang Liu, Tianhao Wu, and Huanhuan Chen. 2021. Knowledge graph quality control: A survey. *Fundamental Research* 1, 5 (2021), 607–626. https://doi.org/10.1016/j.fmre.2021.09.003

[59] Sascha Welten, Marius de Arruda Botelho Herr, Lars Hempel, David Hieber, Peter Placzek, Michael Graf, Sven Weber, Laurenz Neumann, Maximilian Jugl, Liam Tirpitz, Karl Kindermann, Sandra Geisler, Luiz Olavo Bonino da Silva Santos, Stefan Decker, Nico Pfeifer, Oliver Kohlbacher, and Toralf Kirsten. 2024. A study on interoperability between two Personal Health Train infrastructures in leukodystrophy data analysis. *Scientific Data* 11, 1 (June 2024), 1–20. https://doi.org/10.1038/s41597-024-03450-6

[60] Guohui Xiao, Linfang Ding, Benjamin Cogrel, and Diego Calvanese. 2019. Virtual Knowledge Graphs: An Overview of Systems and Use Cases. *Data Intelligence* 1, 3 (June 2019), 201–223. https://doi.org/10.1162/dint_a_00011

[61] Bingcong Xue and Lei Zou. 2023. Knowledge Graph Quality Management: A Comprehensive Survey. *IEEE Transactions on Knowledge and Data Engineering* 35, 5 (2023), 4969–4988. https://doi.org/10.1109/TKDE.2022.3150080

[62] Hong Yung Yip and Amit Sheth. 2024. The EMPWR Platform: Data and Knowledge-Driven Processes for the Knowledge Graph Lifecycle. *IEEE Internet Computing* 28, 1 (2024), 61–69. https://doi.org/10.1109/MIC.2023.3339858

[63] Amrapali Zaveri, Anisa Rula, Andrea Maurino, Ricardo Pietrobon, Jens Lehmann, and Sören Auer. 2016. Quality assessment for Linked Data: A Survey. *Semantic Web* 7, 1 (2016), 63–93. https://doi.org/10.3233/SW-150175

[64] Lingfeng Zhong, Jia Wu, Qian Li, Hao Peng, and Xindong Wu. 2023. A Comprehensive Survey on Automatic Knowledge Graph Construction. *Comput. Surveys* 56, 4 (2023), 1–62. https://doi.org/10.1145/3618295