Empowering Data Literacy Among High School Learners: Insights from a Linked Open Data Workshop

Publisher:

Springer Science and Business Media Deutschland GmbH

Published version:

DOI:10.1007/978-3-031-85561-0_17

Empowering Data Literacy Among High School Learners*

Insights from a Linked Open Data Workshop

Alessia Antelmi[1] [0000−0002−6366−0546] and
Pasquale Esposito[2] [0009−0006−3464−5861]

[1] Università degli Studi di Torino
alessia.antelmi@unito.it
[2] Università degli Studi di Salerno
pasesposito@unisa.it

Abstract. The growing accessibility of (linked) open data offers a rich opportunity for stakeholders to harness their potential value. Nonetheless, significant obstacles like inadequate technical skills and a lack of awareness hinder their effective utilization. As a result, there is a pressing need for learners to cultivate competencies in data and information literacy, which are integral components of 21st-century skill sets.

To foster discussion on initiatives to raise awareness of linked open data and enhance users' capabilities to work with them, we proposed a workshop to introduce Italian high school learners to retrieve information from encyclopedic sources modeled as knowledge graphs. Specifically, the workshop guided the learners through the process of querying the Wikidata endpoint via `KGSnap!`, a block-based programming interface that allows lay users to build and run queries over a SPARQL endpoint.

This article describes the activities of the workshop, delineating its protocol, participants, and outcomes. The results indicate that participants were highly engaged and achieved accurate results in the assigned tasks within a short timeframe, suggesting promising expectations for further initiatives in this domain.

Keywords: Linked Open Data · Data literacy · K-12 learners.

1 Introduction

We are living in a data-driven society that is rapidly evolving in Society 5.0, where the massive availability of data is strictly intertwined with our daily activities [7]. In this landscape, data is converted into information and knowledge, influencing the world through a dual mechanism [3]. On one hand, data indirectly affects the world via humans as it informs and guides human decision-making.

* This work has been partially supported by the spoke "FutureHPC & BigData" of the ICSC – Centro Nazionale di Ricerca in High-Performance Computing, Big Data and Quantum Computing funded by European Union – NextGenerationEU.

On the other hand, data directly drives the world through automated processes. This transition is not merely technological, as it fundamentally reshapes how decisions are made, businesses operate, and public policies are formulated [11]. In this context, Open Data (OD) plays a critical role and is considered one of the crucial drivers for the sustainable economy as well as innovation and creativity in the industry ecosystem, online public services, and public good [12]. OD is data that can be freely used, shared, and built on by anyone, anywhere, for any purpose - subject only to the requirement to attribute and share-alike [15]. OD comes in several formats and quality. Tim Berners-Lee, the inventor of the World Wide Web, proposed a 5-star scheme for grading the quality of OD, where Linked OD (LOD) represents the best format to realize OD, requiring data to be open, structured as a Knowledge Graph (KG), and interlinked. Nonetheless, the exploitation of KGs is hindered by challenges such as the complexity of query languages like SPARQL, which are too difficult for lay users, and conceptualization issues related to understanding how the data are modeled [26].

The success of a data-driven society relies not only on the availability of data but also on individuals' capacity to access, interpret, and utilize it effectively. This is where the need for data literacy becomes pivotal. Data literacy is the ability to comprehend, analyze, create, and communicate data as information effectively, thereby fostering knowledge acquisition [25]. Data literacy is increasingly considered to be a 21st-century life skill, as our interactions with data become increasingly prevalent in everyday life, and individuals frequently rely on data to make judgments and decisions regarding personal information [28, 29]. Despite current literature recognizes the need to develop more effective data literacy education at University level [5], there currently are only a few documented and reproducible efforts to teach OD literacy to the younger generations [18]. When it comes at initiatives in aiming at teaching LOD, these are even fewer [18].

To foster discussion on initiatives aimed at raising awareness of LOD and enhancing users' capabilities to work with them, we proposed a workshop to enhance the awareness and understanding of KGs among high-school students using `KGSnap!`, an extension of Snap!, designed to query KGs with exposed SPARQL endpoints. The didactic approach of the workshop stems from learning-by-doing mechanism, which integrates learning from experiences directly originating from one's own actions [19]. This article describes the activities of the workshop, delineating its protocol, participants, and outcomes. The results indicate that participants were highly engaged and achieved accurate results in the assigned tasks within a short timeframe, suggesting promising expectations for further initiatives in this domain.

The remainder of the paper is organized as follows. Section 2 describes the terms used throughout the text and `KGSnap!`, the tool used in the workshop. Section 3 reviews related work about block-based programming environments and initiatives to teach LOD concepts via such tools. Section 4 details the protocol followed in the workshop activities, participants, and the data collection mechanism. Section 5 describes and discusses quantitative and qualitative observations

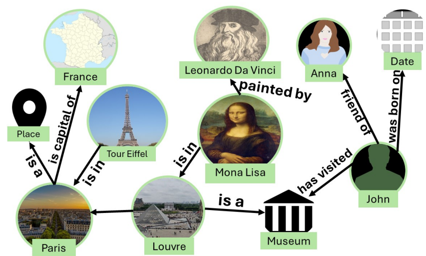


Fig. 1. An example of a KG.

collected during the workshops concerning participants' engagement. Section 6 concludes with final remarks and suggestions for future LOD workshops.

2 Background

2.1 Terminology

Linked (Open) Data (LOD or LD according to the presence of the openness requirements) refers to a set of principles for publishing and interlinking structured data on the web, following standards such as RDF (Resource Description Framework) and SPARQL (SPARQL Protocol and RDF Query Language).

Knowledge Graphs (KGs) are structured representations of knowledge, consisting of entities (such as people, places, and things) and the relationships between them [8]. Data stored in a KG are represented via RDF triples as (subject, predicate, object). For example, we can use the RDF triple `(Da Vinci, painted, Mona Lisa)` to model that Leonardo Da Vinci authored Mona Lisa. By graphically representing subjects and objects as nodes and relationships as edges, we obtain a KG recalling the example visible in Fig. 1. KGs leverage LD principles for data integration and interoperability. In fact, by visually mapping out these connections, KGs enable easier navigation, discovery, and analysis. However, in RDF, all entities and relationships must have a URI (Uniform Resource Identifier). The URI is similar to the primary key for a row in a relational database table, but it is intended to be globally unique. Transitioning from labels to URIs mandates an entity resolution stage. For instance, referencing the Mona Lisa unambiguously in Wikidata requires utilizing https://www.wikidata.org/wiki/Q12418 as a URI.

SPARQL is the standard query language for RDF datasets and it is based on the notion of triple pattern, i.e., an RDF triple in which each component can be replaced by a variable[3]. A SPARQL query can be run over KGs exposing a publicly accessible SPARQL endpoint. Among the available query types, we focus on the SELECT queries, which return a set of tuples that can be rendered as a data table.

[3] SPARQL: https://www.w3.org/TR/2013/REC-sparql11-query-20130321

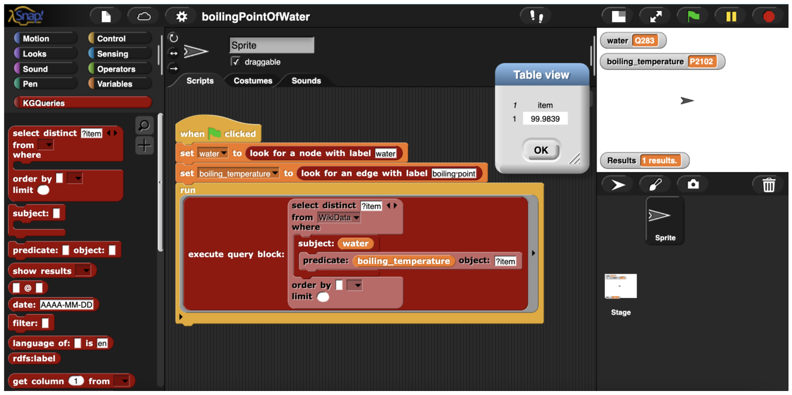


Fig. 2. An illustrative query created using the `KGSnap!`interface, asking for the boiling point of water. The second and third blocks extract the URIs linked to the entities 'Water' and 'Boiling point', respectively, while the query itself is composed within the final block. The result(s) are displayed in a popup table.

2.2 The `KGSnap!` tool

`KGSnap!` is a freely available[4] and open-source extension of Snap! that enables users to easily formulate and execute SELECT SPARQL queries directly on active SPARQL endpoints through a user-friendly block-based programming interface. `KGSnap!` aims to simplify the complexities associated with SPARQL syntax as much as possible without completely hindering the KG modeling process to ensure users maintain a clear understanding of the underlying data structure. This tool supports various fundamental graph operations, including path traversal, filtering, sorting, and manipulating results. Each of its features is encapsulated within a dedicated programming block. By default, `KGSnap!` is configured to query Wikidata, but it can be easily extended to query any operational SPARQL endpoints. To craft a SELECT SPARQL query, users must initially perform an entity resolution step, transitioning from natural language labels to the respective URIs representing nodes or edges in the KG. This process is facilitated by specialized functionalities within the `KGSnap!` framework. Once users have gathered all the relevant URIs, they can proceed with the query construction. Fig. 2 shows an example query.

3 Related work

In block-based programming environments, users interact with a visual interface where they can drag and drop graphical elements known as blocks. These blocks represent code snippets that can be interconnected, similar to assembling pieces of a jigsaw puzzle. This approach enables users to compose code in a syntactically

[4] KGSnap!: https://isislab-unisa.github.io/Snap/snap.html

correct manner, with blocks coming in various shapes to indicate how they can be visually combined. Such a design approach not only showcases the language's capabilities but also aids in preventing syntax errors. Block-based programming environments find applications in various fields, from designing relational algebra expressions [6, 21] to querying databases [10, 27, 22], due to its effectiveness compared with traditional text-based programming [24].

In the Semantic Web community, Sanctorum et al. [20] and Öztürk and Özacar [16] contributed to supporting non-expert users in authoring ontologies and populating KGs through block-based environments. Similarly, RDF Playground [9], SPARQL Playground [1], and Punya [17] are block-based programming environments for performing SPARQL queries and handling query results.

When it comes to evaluating these tools, the process typically involves users in assessing usability, as demonstrated in the evaluations documented by Inostroza et al. [9] and Öztürk and Özacar [16]. KGs and LOD are seldom the primary learning objectives. This observation is supported by a survey conducted by Fettach et al. [4], which noted that KGs in education are primarily utilized by educators for preparing teaching materials, facilitating personalized learning or question-answering, and for assessment purposes, such as authoring final questionnaires. Moreover, tools focusing on learning assessment typically target adults and university students, as illustrated by the workshop documented by Nurmikko-Fuller [13], who employed SPARQL Playground to teach LOD within the cultural heritage community. It is worth emphasizing that, in general, all works on KG block-based programming roughly adhere to the learning-by-doing didactic approach, which integrates learning from experiences directly stemming from one's own actions [19].

While significant efforts have been made to introduce linked data concepts to adults and adult learners, further efforts should be invested in letting K-12 learners familiarize themselves with Semantic Web technologies. To fill this gap, this article proposes a workshop to enhance the awareness and understanding of KGs among high-school students using `KGSnap!`, an extension of Snap!, designed to query KGs with exposed SPARQL endpoints. Regarding performing SPARQL queries, `KGSnap!` offers similar functionalities to related tools, including implemented SPARQL query patterns, results exploration, and export capabilities. However, a notable distinction lies in its extension of the block-based programming environment. Specifically, while Punya extends App Inventor and SPARQL Playground extends Blockly, `KGSnap!` extends Snap!.

4 Material and method

Protocol. The proposed workshop strongly encourages continuous engagement via hands-on and collaborative activities, minimizing reliance on traditional frontal learning settings wherever feasible. Its structure is outlined as follows.

i. *Training stage.* First, the moderator introduced the concepts of LOD and KG through an interactive frontal lecture, prompting participants to formulate hypotheses and answer questions to maintain engagement. In this phase, clarifying the terminology is critical to participants' comprehension of the task. For this reason, a few minutes were dedicated to explaining unfamiliar terms and concepts.

The following phase consisted of a three-task exercise with incremental complexity. The presented tasks are schematically reported in Table 1. For each task, the moderator began by posing the question in natural language, reflecting the query's intent. Participants were then prompted to hypothesize the underlying structure within the KG. The moderator recorded all hypotheses on a whiteboard accessible to everyone, and participants collectively voted on the various options until reaching a consensus. Subsequently, once the underlying structure was correctly derived, the entire class transitioned to formulating the query within the `KGSnap!` interface.

At this point, the moderator displayed a step-by-step guide on the screen, illustrating how to develop and execute the SPARQL query. Afterward, the moderator walked among the desks, addressing unresolved queries and allowing participants to ask questions and resolve uncertainties. Once all participants successfully formulated and executed the query, the class proceeded to the next exercise, following the same procedure. This iterative process continued until all three examples were completed.

ii. *Hands-on stage.* During this phase, participants were tasked with collaboratively addressing the following three questions previously presented in natural language, mirroring the patterns and complexity explored during the training stage.

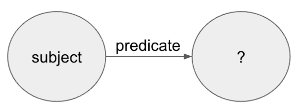
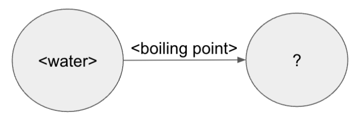
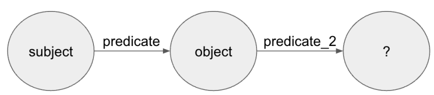
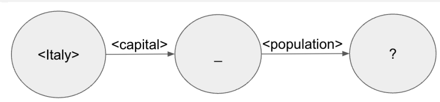
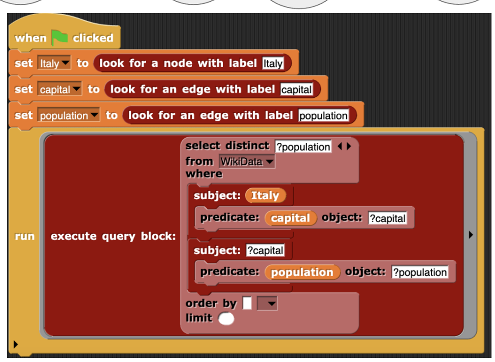
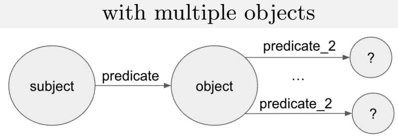
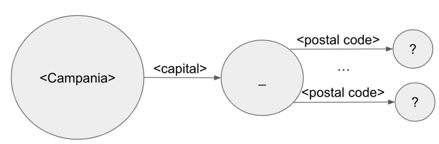
- *What is the building date of the Royal Palace of Caserta?* (1-hop path)
- *What is the birth date of the mayor of Naples?* (2-hop path with a single object as a result)
- *How many brothers does the mayor of Naples have?* (2-hop path with more than one object as a result)

The moderator oversaw participants' progress throughout the tasks, providing technical and terminological clarification while refraining from directly suggesting solutions to the assignments.

Participants and Settings. The experiment took place in February 2024, in presence. 24 learners (58% female), comprising students from the second and third years of a scientific upper-secondary school, participated in the activity. None of the participants were familiar with semantic web technologies such as LD, KGs, and SPARQL queries. Participation was voluntary and free of charge. However, it is worth noting that this workshop was among several activities organized by the school and strongly recommended to the entire class as supplementary to the curriculum. In practice, learners were given the freedom to choose workshops aligned with their interests and curiosities.

The activity lasted two and a half hours and was structured as follows: the first hour was allocated to the training stage, followed by a ten-minute break,

Table 1. Tasks used during the training. First row: general pattern, query in natural language. Second row: general pattern in the KG, corresponding KG in the specific use case. Third row: SPARQL query in `KGSnap!`.

<table>
<tr><th>Pattern</th><th>Use case</th></tr>
<tr><td>1-hop path traversal</td><td>What is the boiling point of water?</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td colspan="2">Query visible in Fig. 2</td></tr>
<tr><td>2-hop path traversal
with a single object</td><td>What is the population of
the capital of Italy?</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td colspan="2">
</td></tr>
<tr><td>2-hop path traversal
with multiple objects</td><td>What are the postal codes of
the capital of the Campania region?</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td colspan="2">The query in <code>KGSnap!</code> has the same format of the previous pattern.</td></tr>
</table>

and the remainder of the time was dedicated to the hands-on stage. At the beginning of the activity, learners organized themselves into groups, forming seven teams ranging in size from individual participants to groups of five. Based on the availability of computers, each group was assigned a single device.

Data collection mechanisms. During the hands-on stage, per each of the three tasks, all the groups were invited to report (i) the identified pattern or the formulated query, (ii) the result returned by `KGSnap!`, and (iii) the time required to complete the task on a format-free blank sheet. These notes have then been used to assess the ease of familiarizing with semantic web technologies and the readiness to learn how to formulate and run SELECT SPARQL queries effectively and accurately in a block-based programming environment.

Additionally, an anonymous questionnaire on Google form was compiled by participants to assess (i) the general engagement, (ii) the perceived complexity, (iii) the level of entertainment achieved via the proposed approach, (iv) the interest in using KGs both before starting the workshop and immediately after the ending of the hands-on stage, and (v) the behavioral intention (BI) to reuse and suggest data literacy workshops to others. All the questions concerning engagement were rated via a 5-point Likert scale.

The moderator acted as an observer, taking notes of the observed behavior and level of engagement throughout the workshop and transcribing notes using the BROMP codes at the end of the activity. Subsequently, those observations were compared at the class level with participants' auto-assessed engagement (via the questionnaire). It is worth highlighting that while the observed engagement and the quality of the performed queries could be compared on a group-by-group basis to identify any potential correspondence between low performance and disengagement, observed and auto-assessed engagement could only be compared at a class level due to the anonymous nature of the questionnaire.

5 Results and Discussion

This section presents and discusses the results, highlighting key promising aspects and various points for reflection.

Accurate results in a short time. Similarly to rapid prototyping experiences, learners are challenged to acquaint themselves with a new modeling approach and query mechanism within a limited timeframe. Despite this constraint, most participants completed the challenges by identifying the correct

Table 2. Accuracy of tasks performed by each group. Groups that returned no feedback are omitted. The symbol ✓represents correct responses, x corresponds to incorrect replies, - reports missing information.

Group ID	Task ID	Correct pattern	Reply	Time in minutes
	1	✓	✓	2
1	2	x	-	15
	3	✓	✓	5
	1	✓	✓	<10
2	2	✓	✓	<5
	3	✓	✓	<5
	1	✓	✓	4
3	2	✓	✓	17
	3	✓	✓	5
	1	✓	✓	4
4	2	$\sim$	✓	16
	3	✓	✓	5
	1	✓	✓	15
6	2	✓	✓	17
	3	-	-	-

conceptualization pattern, formulating queries in `KGSnap!`, and retrieving the desired responses. Interestingly, upon reviewing the time data presented in Table 2, it becomes apparent that completion times do not necessarily correlate with the complexity of the task. While many participants took time to accurately identify the second, more complex pattern, they rapidly demonstrated their ability to apply it in the third and last task. Notably, a group experienced a "cold start", requiring more time to approach the first task than the others.

High engagement. Examining both the observed engagement (see Table 3) and the self-assessed engagement (see Table 4), participants showed consistently high levels of engagement throughout the workshop. Notably, concentration emerged as the most prevalent behavior across all stages, as evidenced by a mean value of 4.4 with a standard deviation of 0.55, as reported by participants. Regarding their inclination towards further participation in similar activities or recommending them to others, opinions appeared somewhat varied (see Table 4, column "BI"), ranging from 3 to 5, with an average score of 4.3. However, the workshop effectively impacted their interest in learning about KGs, as indicated by their self-assessed interest both before and after the activity. Specifically, the scores increased from a minimum of 3 to a minimum of 4, with the mean score rising from 4.2 to 4.6 (see Table 4, columns "Interest in KGs - PRE and POST").

Stronger focus on conceptualization rather than technical challenges. Data literacy workshops should encompass the essential phases of accessing, querying, and leveraging data. Consequently, participants should be encouraged to hypothesize how data can be structured and conceptualize the underlying KGs to formulate queries accordingly. They should also learn how to articulate queries by dealing with SPARQL syntax and explore the results returned. While KGSnap! mitigates the complexity of formulating and executing SPARQL queries through a block-based programming environment, the conceptualization

Table 3. BROMP codes [14] observed at the group level by the moderator, acting as an observer during the training and hands-on stages. BROMP is a protocol for qualitative and quantitative field observations of (on/off-task) behaviors (**B**) and affective (**A**) states in learning with technology, which indicates engagement in tasks.

Group id	Number of members	Training stage				Hands-on	
		Frontal lesson		Step-by-step hands-on			
		B	A	B	A	B	A
1	3	on	engaged	on	concentrate	on	concentrate
2	4	on	concentrate	on on	concentrate engaged	on off	concentrate boredom
3	5	off		on	concentrate	on	delight
4	1	on	concentrate	on	concentrate	on	engaged
5	2	on off	confusion	on on	concentrate confusion	on off	confusion
6	4	on on	concentrate confusion	on	concentrate	on on	concentrate confusion
7	5	on off	concentrate	on	concentrate	off on	 concentrate

Table 4. Perceived engagement, auto-assessed by participants through a questionnaire, with agreement reached within the group

	Perceived easiness	Perceived fun	Overall engagement	Interest in KGs PRE	Interest in KGs POST	BI Re-join	BI Suggest
Range	1-5	1-5	1-5	1-5	1-5	1-5	1-5
	3	3	4	5	4	5	5
	4	3	5	5	5	4	5
	4	4	4	3	5	5	5
	4	4	5	3	5	3	3
	4	4	4	5	4	4	4
Min	3	3	4	3	4	3	3
Mean	3.8	3.6	4.4	4.2	4.6	4.2	4.4
St.Dev.	0.45	0.55	0.55	1.10	0.55	0.84	0.89
Max.	4	4	5	5	5	5	5

stage remains the responsibility of the learners. As indicated in Table 2, the conceptualization phase is non-trivial and likely influenced the perceived "easiness" scores reported in Table 4, as not all participants successfully identified the query pattern on their initial attempt. This necessitated a trial-and-error approach to contemplate alternative modeling strategies, experiment with query formulations, and verify the resulting outcomes.

6 Conclusions and Future Directions

Data literacy, a specialized form of literacy centered around data, is becoming increasingly vital in our data-driven world. It involves learning how to read, interpret, and manage data and its representations to enable critical thinking and informed decision-making. KGs are emerging as valuable educational tools, especially given the growing number of resources available as KGs and the projected increase in their use in the coming years. However, LOD presents challenges, primarily related to querying them via SPARQL and dealing with conceptualization issues. These technical difficulties can become time-consuming obstacles for both teachers and learners [2].

There are several works exploring the convergence of LOD and education [4], rarely focusing on how to engage learners in actively working with them. In this direction, the workshop described in this contribution aims to raise awareness within a broader community about the potential of these technologies to facilitate the widespread utilization of data in various formats. While data representations based on graphs may seem intricate for both data curators and users, they inherently resemble conceptual maps commonly used in education from an early age. Graphs and LD naturally evoke the ways in which humans conceptualize learned concepts. By leveraging the similarities between KGs and conceptual maps, moderators can help young learners become more comfortable with complex data modeling approaches.

Our primary objective is to meet the threshold required for effectively integrating KGs in learning environments via a scaffolding approach [23], utilizing blocks to simplify SPARQL complexity. Nonetheless, we are also considering transitioning to a text-based programming environment in the future.

Acknowledgment

We express our gratitude to BIMED, represented by its president Andrea Iovino, for allowing us to conduct the research upon which this article is based. Further, we sincerely thank Maria Angela Pellegrino, who participated in the activity discussed in this paper and contributed valuable insights and feedback.

References

1. Bottoni, P., Ceriani, M.: SPARQL Playground: A Block Programming Tool to Experiment with SPARQL. In: Proceedings of the International Workshop on Visualizations and User Interfaces for Ontologies and Linked Data co-located with 14th the Int. Semantic Web Conference. vol. 1456, p. 103 – 108. CEUR-WS (2015)
2. Cui, J., Yu, S.: Fostering deeper learning in a flipped classroom: Effects of knowledge graphs versus concept maps. British Journal of Educational Technology **50**(5), 2308–2328 (2019)
3. Deguchi, A., Hirai, C., Matsuoka, H., Nakano, T., Oshima, K., Tai, M., Tani, S.: What Is Society 5.0?, pp. 1–23. Springer Singapore, Singapore (2020)
4. Fettach, Y., Ghogho, M., Benatallah, B.: Knowledge Graphs in Education and Employability: A Survey on Applications and Techniques. IEEE Access **10**, 80174–80183 (2022)
5. Ghodoosi, B., West, T., Li, Q., Torrisi-Steele, G., Dey, S.: A systematic literature review of data literacy education. Journal of Business & Finance Librarianship **28**(2), 112–127 (2023)
6. Gorman, J., Gsell, S., Mayfield, C.: Learning relational algebra by snapping blocks. In: ACM technical symposium on Computer science education. pp. 73–78 (2014)
7. Harayama, Y., Fukuyama, M.: Society 5.0: Aiming for a New Human-centered Society (2017), https://www.hitachi.com/rev/archive/2017/r2017_06/trends/index.html, [Online] Last access April 8, 2024
8. Hogan, A. et al.: Knowledge Graphs. ACM Comp. Sur. **54**(4), 71:1–71:37 (2022)
9. Inostroza, B., Cid, R., Hogan, A.: RDF Playground: An Online Tool for Learning about the Semantic Web. In: Companion Proceedings of the ACM Web Conference. p. 111–114 (2023)
10. Jacobs, S., Jaschke, S.: SQheLper: A block-based syntax support for SQL. In: Global Engineering Education Conference (EDUCON). pp. 478–481. IEEE (2021)
11. Mourtzis, D., Angelopoulos, J., Panopoulos, N.: A Literature Review of the Challenges and Opportunities of the Transition from Industry 4.0 to Society 5.0. Energies **15**(17) (2022)
12. Nikiforova, A., Flores, M.A.A., Lytras, M.D.: The Role of Open Data in Transforming the Society to Society 5.0: A Resource or a Tool for SDG-Compliant Smart Living?, pp. 219–252. Smart Cities and Digital Transformation: Empowering Communities, Limitless Innovation, Sustainable Development and the Next Generation, Emerald Publishing Limited (2023)
13. Nurmikko-Fuller, T.: Teaching Linked Open Data using Bibliographic Metadata. Journal of Open Humanities Data **8** (2022)

14. Ocumpaugh, J.: Baker Rodrigo Ocumpaugh monitoring protocol (BROMP) 2.0 technical and training manual. New York, NY and Manila, Philippines: Teachers College, Columbia University and Ateneo Lab. for the Learning Sciences **60** (2015)
15. Open Knowledge Foundation: Defining Open Data. https://blog.okfn.org/2013/10/03/defining-open-data (2013), [Online, Last access May 2024]
16. Öztürk, Ö., Özacar, T.: A case study for block-based linked data generation: Recipes as jigsaw puzzles. Journal of Information Science **46**(3), 419–433 (2020)
17. Patton, E.W., Van Woensel, W., Seneviratne, O., Loseto, G., Scioscia, F., Kagal, L.: The Punya platform: building mobile research apps with linked data and semantic features. In: The Semantic Web–ISWC. pp. 563–579. Springer (2021)
18. Pellegrino, M., Antelmi, A.: At school of open data: A literature review. In: Proceedings of the 15th International Conference on Computer Supported Education - Volume 2: CSEDU. pp. 172–183. SciTePress (2023)
19. Reese, H.: The learning-by-doing principle **17**, 1–19
20. Sanctorum, A., Riggio, J., Sepehri, S., Arnesdotter, E., Vanhaecke, T., De Troyer, O.: A jigsaw-based end-user tool for the development of ontology-based knowledge bases. In: International Symposium on End User Development. pp. 169–184. Springer (2021)
21. Silva, Y.N., Chon, J.: Dbsnap: Learning database queries by snapping blocks. In: ACM Technical Symposium on Computer Science Education. pp. 179–184 (2015)
22. Silva, Y.N., Nieuwenhuyse, A., Schenk, T.G., Symons, A.: Dbsnap++: creating data-driven programs by snapping blocks. In: the 23rd Annual Conference on Innovation and Technology in Computer Science Education. pp. 170–175. ACM (2018)
23. Stanier, C.: Scaffolding in a higher education context. In: ICERI2015 Proceedings. pp. 7781–7790. Inter. Conf. of Education, Research and Innovation, IATED (2015)
24. Sun, D., Ouyang, F., Li, Y., Zhu, C., Zhou, Y.: Using multimodal learning analytics to understand effects of block-based and text-based modalities on computer programming. Journal of Computer Assisted Learning (2024)
25. Van Audenhove, L., Van den Broeck, W., Mariën, I.: Data literacy and education: Introduction and the challenges for our field. Journal of Media Literacy Education **3**, 1–5 (2020)
26. Vargas, H., Aranda, C.B., Hogan, A., López, C.: RDF explorer: A visual SPARQL query builder. In: The Semantic Web - ISWC 2019 - 18th International Semantic Web Conference. Lecture Notes in Computer Science, vol. 11778, pp. 647–663. Springer (2019)
27. Vinayakumar, R., Soman, K., Menon, P.: DB-learn: studying relational algebra concepts by snapping blocks. In: 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT). pp. 1–6. IEEE (2018)
28. Wolff, A., Gooch, D., Cavero Montaner, J., Rashid, M., Kortuem, G.: Creating an Understanding of Data Literacy for a Data-driven Society. The Journal of Community Informatics **12**(3) (2016)
29. Wolff, A., Wermelinger, M., Petre, M.: Exploring design principles for data literacy activities to support children's inquiries from complex data. International Journal of Human-Computer Studies **129**, 41–54 (2019)