Investigating the potential of the semantic web for education: Exploring Wikidata as a learning platform

Shani Evenstein Sigalov[1] · Rafi Nachmias[1]

Abstract

Wikidata is a free, multilingual, open knowledge-base that stores structured, linked data. It has grown rapidly and as of December 2022 contains over 100 million items and millions of statements, making it the largest semantic knowledge-base in existence. Changing the interaction between people and knowledge, Wikidata offers various learning opportunities, leading to new applications in sciences, technology and cultures. These learning opportunities stem in part from the ability to query this data and ask questions that were difficult to answer in the past. They also stem from the ability to visualize query results, for example on a timeline or a map, which, in turn, helps users make sense of the data and draw additional insights from it. Research on the semantic web as learning platform and on Wikidata in the context of education is almost non-existent, and we are just beginning to understand how to utilize it for educational purposes. This research investigates the Semantic Web as a learning platform, focusing on Wikidata as a prime example. To that end, a methodology of multiple case studies was adopted, demonstrating Wikidata uses by early adopters. Seven semi-structured, in-depth interviews were conducted, out of which 10 distinct projects were extracted. A thematic analysis approach was deployed, revealing eight main uses, as well as benefits and challenges to engaging with the platform. The results shed light on Wikidata's potential as a lifelong learning process, enabling opportunities for improved Data Literacy and a worldwide social impact.

✉ Shani Evenstein Sigalov
shanieev@mail.tau.ac.il

Rafi Nachmias
nachmias@post.tau.ac.il

[1] School of Education, Tel Aviv University, 55 Levanon St., 6997801 Tel Aviv, Israel

1 Introduction

We live in an age of information explosion. Knowledge is available in a variety of forms, with the ease of a click-of-a-button or a voice-activated digital agent, almost anytime, anywhere. This abundance of information has driven a search for new modes of meaningful engagement with information, exploring new ways to access, assess and use data. In recent years it has become almost impossible to ignore 'buzz' words, such as 'Big Data', 'Artificial Intelligence (AI)', 'Machine Learning' and 'Data Science'—all manifestations of humanity's efforts to deal with the information overload.

Into this 'sea of knowledge', Wikidata, a "free and open knowledge base that can be read and edited by both humans and machines"(Wikidata.org), was launched in 2012. Wikidata serves as a central, multilingual and free storage for *structured, linked data*, which is drawn from Wikipedia, its sister projects, and from other external sources. Wikidata has been growing exponentially and is maintained by a community of over 25,000 editors (https://w.wiki/PaP). On October 2022, Wikidata celebrated its 10th anniversary, crossing the 100 million items threshold, and to date, it is the largest Semantic knowledge-base in existence. As such, it seems Wikidata holds "many exciting possibilities" (Erxleben et al., 2014), and opens the door for a variety of new research opportunities and "potential applications across all areas of sciences, technology and cultures" (Vrandečić & Krötzsch, 2014).

But what potential does Wikidata holds for its users, especially educators and researchers? How can it be used as a life-long learning platform? And what are the benefits and challenges of engaging with this platform? This exploratory research investigates the phenomenon of Wikidata by analyzing multiple use case studies, focusing specifically on its value in educational contexts. For the initial purpose of investigating Wikidata's value for Education, seven semi-structured, in-depth interviews were conducted with early adopters of the platform, out of which 10 distinct case studies were extracted. A qualitative analysis approach of multiple case studies was used to explore, code and categorize the projects. A thematic analysis of patterns was used to extract main uses, benefits and challenges. The case studies, uses, benefits and challenges are described and discussed herein.

2 Background

2.1 From Web 1.0 to Web 3.0: The emergence of the semantic web

While Web 1.0 used static HTML pages that users could only consume, Web 2.0 allowed users to create and share information, such as social networks and blogs. In recent years we have witnessed another evolution of the web, the emergence of Web 3.0,[1] which among other things includes the Semantic Web, or Linked Data. In 1980, "semantic webbing" was described as organizing information and relationships by visually displaying

---

[1] Web 3.0 is not to be confused with *Web3*, a new term that emerged in 2014 and refers to a "decentralized online ecosystem based on blockchain".

them (Freedman & Reynolds, 1980). Burners-Lee presented the idea of using typed links as a semantics tool, calling it "Semantic Web" (Guns, 2013). After describing its roadmap in 1998, he introduced the modern idea of Semantic Web in 2001 (Berners-Lee et al., 2001). Bizer, Heath and Berners-Lee explain that "Linked Data realizes the vision of evolving the Web into a global data common, allowing applications to operate on top of an unbounded set of data sources, via standardized access mechanisms" (Bizer et al., 2009). According to this vision, "the traditional Web… should be extended to a Web of Data where not only documents and links between documents exist, but [also links among] any entity and any relation" (Färber et al., 2015), in such a way that "machines would be able to participate and help" humans.[2]

This new technology focuses on data rather than on applications and is known as Web 3.0 (Hendler, 2009). In the Social Web (Web 2.0), value is created by aggregates information and knowledge from users and online communities. In the Semantic Web, however, value is created by integrating structured data from many sources (Gruber, 2008) and meaningfully connecting pieces of information. At the heart of the Semantic Web is *structured, linked data*, that "makes substantial reuse of existing ontologies and data" (Shadbolt et al., 2006). This new web theoretically allows both humans and machines to harness the power of a high quality, up-to-date and well-referenced knowledge base of linked data. "Linked Data principles and practices have been adopted by an increasing number of data providers, resulting in the creation of a global data space on the Web containing billions of RDF triples" (Hernández et al., 2016). This brings us closer to fulfilling the prediction that "Linked Data will enable a significant evolutionary step in leading the Web to its full potential" (Bizer et al., 2009).

2.2 Semantic networks as learning platforms

The Semantic Web offers structured linked data that both humans and machines can tap into as a resource. This also means it could be used for learning, not only in the classroom, but as a 'lifelong learning', which has a variety of definitions documented in academic literature (Aspin & Chapman, 2010; Collins, 2004; Dabbagh & Castaneda, 2020; Field, 2000; Laal, 2012; Laal & Salamati, 2012). In this paper, we refer to it as an ongoing, self-motivated pursuit of knowledge, skills, literacies or competencies along one's life, whether for professional or personal reasons. As early as 2003 educators started exploring the Semantic Web relating to education, e-learning and lifelong learning (Anderson & Whitelock, 2004; Koper, 2004; Naeve et al., 2006). That research focused mainly on *using the technology to advance education*, rather than exploring the *types of learning* and *uses* the technology enabled. That is mainly since the Semantic Web vision has not been fully realized on a scale that would allow such exploration. It took a decade longer till Wikidata was launched, marking an important milestone in realizing the Semantic Web dream on a scale, and a few additional years to mature to a point in which educators began experimenting

---

[2] Berners-Lee's road map document could be found here—https://www.w3.org/DesignIssues/Semantic.html

with it in classroom, examining its potential as a learning platform. Meanwhile, Web 2.0 platforms have matured, driving a profusion of academic research related to various applications, among them Wikipedia, as learning platforms and pedagogical frameworks that support learning (Evenstein Sigalov & Nachmias, 2017). These experiences with Web 2.0, and specifically with Wikipedia, would later affect the engagement with Semantic platforms and the use of Wikidata in educational and research contexts, leading to its exploration as a learning platform. Let us therefore explore some relevant theories and frameworks for Web 2.0 platforms that may be relevant to the Semantic Web.

2.3 Web 2.0 theories, frameworks & literacies relevant for the semantic web

There is a wealth of literature exists on Web 2.0 applications as learning platforms, specifically engaging with Wikipedia in educational context. Conversely, almost none deals with Web 3.0 and Semantic Web as learning platforms. While awaiting new research and innovative pedagogical frameworks to emerge, three Web 2.0-related educational theories, frameworks or paradigms (specifically for Wikipedia as a learning platform), have been identified as relevant for the Semantic Web.

The first is **Constructivism.** Rooted in the works of Dewey, Mead and Piaget, its paradigm "describes how learning happens" (Parker & Chao, 2007). Knowledge and meaning are "constructed rather than given" (Parker & Chao, 2007), through a "discussion with peers and teachers, and through reflection" (Higgs & McCarthy, 2005). "The focus on real, authentic problems… force[s] learners to… develop capacity for effective problem-solving behaviours" (Anderson, 2016). Such learning should be "cooperative, collaborative, and conversational, providing students with opportunities to interact…, to clarify and share ideas, to seek assistance, to negotiate problems, and discuss solutions" (Boulos et al., 2006). As Anderson puts it, "multiple perspectives and sustained dialogue lead to effective learning" (Anderson, 2016). To sum up, engaging with a community of learners allows learners to sharpen or gain new skills, and motivates them to attach meaning to what is learned, which results in construction of new knowledge that is better retained.

The second framework is **collaborative learning**. Per Wheeler et al., engaging deeply with "learning objects" and web-based discussion, communities bring forth significant benefits for the "development of professional practice (Boulos et al., 2006; Wheeler et al., 2005). Parker & Chao, as well as Boulos, Maramba and Wheeler, asserted that using a technological collaborative platform encourages a deeper engagement with learning materials. Collaborative learning, then, leads to positive interdependence of group members, individual accountability, and appropriate use of collaborative skills (Parker & Chao, 2007; Schaffert et al., 2006). Collaborative learning also stimulates higher levels of thought and cognitive work, and longer information retention. (Galway et al., 2014; Johnson & Johnson, 1986; Parker & Chao, 2007; Schaffert et al., 2006). To conclude, while engaging with a community, users develop collaborative learning skills and knowledge, while highlighting the importance of the technology as a platform that enables learning.

The third framework, **Self-directed learning**, and more recently, **Heutagogy**, was developed by Hase and Kenyon in 2000, and named after the Greek word for "self". With strong roots in self-directed learning, Heutagogy shifts the focus and control from the teacher to the learner (Anderson, 2016). The educational focus shifts from instructing and testing competencies, towards learning in new and unfamiliar contexts, as a life-long process (Blaschke, 2021; Hase & Blaschke, 2021; Moore, 2020). As Hase and Kenyon put it, "heutagogy looks to the future in which knowing how to learn will be a fundamental skill given the pace of innovation and the changing structure of communities and workplaces" (Hase & Kenyon, 2000). "Heutagogy thus emphasizes self-direction and focuses on the development of efficacy in utilizing the online tools and information available" (Anderson, 2016). While the focus on learners is positive, this shift puts pressure on the learner. As Kop and Hill explain, in order to succeed, the learner needs to be not only capable, but also highly motivated to engage in a self-directed learning (Kop & Hill, 2008). The question of motivation and skills is therefore highlighted as a significant influencer on a successful learning process.

These three educational theories and frameworks have been used by educators to promote acquisition of not only knowledge, but also skills and literacies required for lifelong learning in the digital age, including *Digital Literacy* (Pangrazio et al., 2020; Reddy et al., 1 C.E., n.d.; Spante et al., 2018) and *Data Literacy* (Gummer & Mandinach, 2015; Koltay, 2015; Mandinach & Gummer, 2013; Mandinach et al., 2015; Schield, 2004; Stephenson & Caravello, 2007; Wang et al., 2019). Both terms have multiple definitions in literature, and it is outside the scope of this paper to fully explore them. For **Digital Literacy** we have relied on a systematic review of the term by Spante et al. (2018). While the original definition by Gilster (1997) was "the ability to understand and use information in multiple formats from a wide range of sources when it is presented via computers", the term has evolved over time (Spante et al., 2018). They note that later researchers suggest that the term "originates in a skill-based understanding of the concept and thus relates to the functional use of technology and skills adaptation"; and that, "definitions of digital literacy point towards cognitive skills and competences" (Spante et al., 2018). For Chan et al. (2017), it is "the ability to understand and use information in multiple formats with emphasis on critical thinking rather than information and communication technology skills". They also note that at times, the term is used in the plural, "digital literacies", which "acknowledges new and diverse social practices" and "emphasizes the non-generic and multiply situated nature of the term" (Spante et al., 2018). Finally, they note that some researchers expand the definition to "new textual landscape", including social media and social practices (Spante et al., 2018). Such definitions include Martine's definition (2006), also used by Tang & Chaw (2016), and adopted as well for this research, as "the awareness, attitude and ability of individuals to appropriately use digital tools and facilities to identify, access, manage, integrate, evaluate, analyze and synthesize digital resources, construct new knowledge, create media expressions, and communicate with others, in the context of specific life situations, in order to enable constructive social action; and to reflect upon this process" (Spante et al., 2018).

As for **Data Literacy**, some basic definitions would refer to the ability to "understand, use and manage data" (Qin & D'Ignazio, 2010), or "the ability to understand and use data effectively to inform decisions" (Mandinach & Gummer, 2013). In a world of Information Explosion, Big Data, AI and Machine Learning, it is essential to assist learners in developing critical thinking related not only to digital, online spaces, but more specifically, to data, as the backbone of digital environments. The educational theories and frameworks, as well as the literacies they promote, seem to be relevant not only to Web 2.0, but also to Web 3.0 and Semantic platforms such as Wikidata, as will be demonstrated later. Before delving into Wikidata, we will therefore examine what engaging with Wikipedia has taught us and how it has informed the experimentation with Wikidata later on.

2.4 Wikipedia as learning platforms

In the last decade, a growing number of educators have been using Wikipedia and integrating it into the curricula (Aibar et al., 2015; Dooley, 2010; Evenstein Sigalov & Nachmias, 2017). Initially, Wikipedia was used to teach better information consumption skills and then started to be utilized as a platform for collaborative knowledge construction. But what can research reveal about its benefits as a teaching and learning platform? Wikipedia strives for quality, up-to-date, neutral and well-referenced articles, and offers unique educational opportunities for both teachers and learners (Evenstein Sigalov & Nachmias, 2017; Herbert et al., 2015; Konieczny, 2007, 2016). As a Web 2.0 platform that allows users not only to consume information, but also to create and share knowledge, Wikipedia's pedagogical potential has long been investigated. Educators and researchers have focused on its ability to actively and collaboratively involve learners in the construction of knowledge (Aibar et al., 2013, 2015; Boulos et al., 2006; Evenstein Sigalov & Nachmias, 2017; Konieczny, 2016; LaFrance & Calhoun, 2012; Mareca & Bordel, 2019; Mendes et al., 2021; Minguillón et al., 2018; Naismith et al., 2011; Ramanau & Geng, 2009; Seitzinger, 2006), while aiding them develop skills, such as digital literacy, collaborative skills, critical thinking and academic literacy (Bordel & Mareca, 2019; Di Lauro & Johinke, 2017; Eteokleous et al., 2014; LaFrance & Calhoun, 2012; McKenzie et al., 2018; Selwyn & Gorard, 2016; Soler-Adillon et al., 2018; Staub & Hodel, 2016; Vetter et al., 2019; Zheng et al., 2015).

Most educators have experimented with Wikipedia as an alternative assessment method, substituting traditional assignments such as tests or papers. This type of Open Educational Practice, a form of assignment (and assessment) that contribute to the greater good, has been at times referred to as Renewable Assignment or Non-Disposable Assessment (Wiley & Hilton, 2018). Though used in the classroom for at least 15 years, Wikipedia is still considered relatively new in higher education (Chao, 2007; Evans, 2006; Evenstein Sigalov & Nachmias, 2017; Franklin & Harmelen, 2007; Konieczny, 2007; Schaffert et al., 2006). As Konieczny explains, Wikipedia seems to be gaining acceptance among academics and educators slowly and grudgingly (Konieczny, 2016). Undoubtedly, some progress has been made and a growing number of educators are seeking to incorporate Wikipedia into their

curriculum (Evenstein Sigalov & Nachmias, 2017; Konieczny, 2016). That said, Academia has yet to explore and realize the full potential of Wikipedia as a learning platform, and is yet to formalize the means to promote "deeper learning and integration of learning experiences from both inside the classroom and out" (Chen et al., 2005; p. 96). Educators and instructors are at times still uncertain on how to integrate wikis into the classroom for effective collaboration (Allwardt, 2011; Elgort et al., 2008; Konieczny, 2014, 2016; Naismith et al., 2011; Ramanau & Geng, 2009), and continue to experiment with it, looking for new ways of engagement. Considering educators' endeavors to tap into Wikipedia's potential as a pedagogical tool (Bayliss, 2013; Boulos et al., 2006; Jaroslaw P. Janio, 2014; Kummer, 2013; LaFrance & Calhoun, 2012; Naismith et al., 2011; Ramanau & Geng, 2009; Seitzinger, 2006), one important milestone in expanding the implementation of Wikipedia into the academic curricula has been developing a new elected-course models in 2013 (Evenstein Sigalov & Nachmias, 2017; Mendes et al., 2021). This new model led to for-credit, semester-long, elective courses, in which adding content to Wikipedia has been used as a main assessment model.

2.5 The case of Wikidata

With deep roots in the Semantic Web community, Wikidata came to existence in 2012 when several Wikimedians, including Dr. Denny Vrandečić, tried to answer a question that a Google search failed to accurately address: "What are the 10 largest cities with a female mayor?" (Erxleben et al., 2014; Krötzsch et al., 2007).Vrandečić felt that free and open knowledge should include data that can be searched, analyzed and reused, and as a response developed Wikidata (Vrandečić & Krötzsch, 2014). Wikidata provides a rich, free and multilingual dataset that is constantly improved by users and machines (detailed explanation of Wikidata can be found in Appendix 1). Vrandečić's statement that Wikidata has exceeded his expectations, may be explained in the new learning opportunities it offers its users, as will be explored hereafter. As far as we know, the first experimentations with Wikidata in academic settings began in 2015. Like in the case of Wikipedia, we first encountered it as an exploratory, informative addition to other courses, then as an alternative assessment in various courses; and in 2018, for the first time, Wikidata became a main assessment in an Academic course. Additionally, outside Academia, researchers, industries, cultural and governmental institutions began experimenting with Wikidata, resulting in various types of learning beyond the classroom.

2.6 Types of engagement with Wikidata

While multiple methods of interacting with Wikidata may induce learning and acquisition of knowledge and skills, two main user interactions and their learning opportunities were considered for this research—*data curation*, the process of adding information to Wikidata; and *data extraction*, the process of querying and

extracting information from Wikidata. **Data curation** is performed via four main methods:

1) direct, manual edits directly into Wikidata's interface.
2) the Wikidata Game (https://tools.wmflabs.org/wikidata-game/) and Distributed Games (https://tools.wmflabs.org/wikidata-game/distributed/)—both allow micro-contributions to Wikidata by playing simple games.
3) Quick Statements, a tool that allows users to add multiple statements to multiple items (https://tools.wmflabs.org/wikidata-todo/quick_statements.php).
4) Mass-uploads of metadata donations from external institutions via dedicated tools or bots.

    Data extraction is achieved via three main methods:

1) querying through a service, such as the built-in one (https://query.wikidata.org/). Most services require knowledge of a special coding language called SPARQL (https://w.wiki/N3J), while others do not, such as VizQuery(http://tools.wmflabs. org/hay/vizquery/), and Platypus (https://askplatyp.us/; https://blog.wikimedia. de/2015/02/23/platypus-a-speaking-interface-for-wikidata/).
2) Third party applications that explore data and visualize the results (https://w.wiki/ PaQ) (three examples are available in Appendix 2).
3) Data extraction via Wikidata API (Malyshev et al., 2018), an advanced method that will not be discussed as it is outside the scope of this paper.

2.7 Wikidata users and early adopters

Early adopters of Wikidata include *Wikimedians* (https://w.wiki/PaP); *industries* that use the database to offer various services; *institutions*, that also donate their metadata (Kapsalis, 2019; Klein & Kyrios, 2013; Snyder et al., 2020; Tharani, 2021); and *Researchers* experimenting and conducting various types of research with the platform (Farda-Sarbas & Müller-Birn, 2019; Heftberger et al., 2020; Hernández et al., 2015; Lemus-Rojas & Lee, 2019; Steiner, 2014).

   In the past decade the research community has shared various types of research papers dealing with Wikidata, which could be categorized in a variety of ways. While reviewing the literature, it was found that the majority of existing research focuses on either technological aspects or ontological aspects of the platform. More Specifically, review of the literature illustrated that researchers use Wikidata to conduct **new types of research** (Amaral et al., 2021; Colla et al., 2021; Ferradji & Benchikha, 2021; Good et al., 2016; Kaffee, 2016; Konieczny & Klein, 2018; Lemus-Rojas & Odell, 2018; Li et al., 2022; Meier, 2022; Mietchen et al., 2015; Morshed, 2021; Neelam et al., 2022; Rasberry & Mietchen, 2021; Shenoy et al., 2022; Taveekarn et al., 2019; Waagmeester et al., 2020, 2021; Zhang et al., 2022). Researchers also use Wikidata to conduct **new types of academic analysis** in a

variety of disciplines (Arnaout et al., 2021; Burgstaller-Muehlbacher et al., 2016; Kaffee et al., 2017; Klein et al., 2016; Lemus-Rojas, n.d.; Pfundner et al., 2015; Putman et al., 2017; Rutz et al., 2021; Scharpf et al., 2021a, b; Turki et al., 2019, 2022a, b). Finally, at times researchers use Wikidata to demonstrate **new types of visualizations** (Hernández et al., 2016; Metilli et al., 2019; Nielsen et al., 2017; Nielsen, 2016a, b).

In a meta-review study conducted by Farda-Sarbas & Müller-Birn in 2019, 67 peer-reviewed articles from journal and conference proceedings were classified and categorized (Farda-Sarbas & Müller-Birn, 2019). The researchers divided existing academic research on Wikidata into 5 main categories: **1) Data Oriented Research**, including "data quality issues", and "tools & datasets" (22 articles in total); **2) Knowledge Graph Oriented Research**, including "comparison of knowledge graphs", "common issues of knowledge graphs", and "Wikidata as linked data provider" (15 articles); **3) Community-oriented Research**, including "design decisions", "WD community", and "multilingualism" (14 articles); 4) **Engineering-oriented Research**, including "enhancement features and vandalism detection" 9 articles); and 5) **Application Use Cases,** including "medical & biological data", and "linguistics" (7 articles). In their conclusion, the researchers explain that while Wikipedia has been studied in a variety of disciplines, this is not the case with Wikidata, despite the platform having "the competence to be used in different disciplines" (Farda-Sarbas & Müller-Birn, 2019). They recommend that further investigations must take place "to find out whether Wikidata can be beneficial in the same areas where Wikipedia was used" (Farda-Sarbas & Müller-Birn, 2019). As they note, while their analysis revealed usage of Wikidata in various contexts, the use cases "come from the biomedical domain and linguistics mainly" (Farda-Sarbas & Müller-Birn, 2019). They conclude by suggesting, "It might be valuable to see more use cases from other disciplines, such as social sciences or humanities. It might be valuable, for example, to use Wikidata in educational or museum settings" (Farda-Sarbas & Müller-Birn, 2019). Similar conclusions are to be found in another systematic review of the Wikidata-related literature conducted by Mora-Cantallops et al. (Mora-Cantallops et al., 2019). It appears, then, that the promise that Wikidata holds for education and research is yet to be fully explored and examined. This potential exploration includes what could be learned from existing interactions with the platform, practical uses of the platform, and the benefits and challenges users experience throughout various interactions.

3 The study

3.1 Research goals

It appears that humanity is just beginning to explore the potential of Semantic Web platforms, and more specifically, Wikidata's potential for education and research. As Müller-Birn et al. found, "Peer-production communities addressing the development of structured data have not as yet attracted much attention from the research community" (Müller-Birn et al., 2015). For this reason, questions

relating to different processes of interactions with the platform from a user's learning perspective remained unexplored by academic research. As Wikidata is still relatively young, the results of its continued progress are complex to divine. However, given its close connection to Wikipedia it is apparent that exciting possibilities of both understanding how to contribute to it and how to utilize its data "remain to be explored" (Erxleben et al., 2014). As Müller-Birn and his collaborators put it, "Wikidata provides the prototype of a system that allows even non-technical experts to create and manage semantic data", with the potential to be "the nucleus for a completely new type of system" (Müller-Birn et al., 2015). Erxleben and his collaborators conclude that "It remains for the community of researchers and practitioners in semantic technologies and linked data to show the added value Wikidata can bring about" (Erxleben et al., 2014).

Considering Wikidata's potential, the main purpose of this paper is to investigate its value for education and research, a topic yet to be properly covered by academic research -as noted in the systematic review of existing research (Farda-Sarbas & Müller-Birn, 2019; Mora-Cantallops et al., 2019). More specifically, this paper aims to inform educators and researchers about new learning opportunities enabled via this semantic platform by examining early adopter projects in educational, research and cultural institutions, shed light on the main aspects that make Wikidata valuable to all disciplines, and demonstrate its power as a potential learning platform in diverse contexts.

3.2 Research Questions

Bearing in mind the research goals, the main research questions are:

1) What are some of the distinct **projects** using Wikidata?
2) Considering these projects, what are the **main uses** of Wikidata that induce learning in the context of education or research?
3) Based on the projects and uses, what are the **main benefits and challenges** when using Wikidata in the context of learning and engaging with data?

4 Methodology

4.1 Research design & strategy

We investigate Wikidata's value for its users via multiple projects' case studies. This methodology requires an in-depth examination that draws on multiple sources for information (Creswell, 1998). However, the Semantic Web, and specifically Wikidata, is a relatively new phenomenon, which has not yet been explored in the context of learning (Farda-Sarbas & Müller-Birn, 2019; Mora-Cantallops et al., 2019). Sources of information are still hard to find, and literature on the topic is almost non-existent. To better understand the relatively new phenomenon of

utilizing Wikidata for learning, the study, approved by the university's Ethics Committee, engaged the international community of early adopters and for this article included seven semi-structured, in-depth user interviews that had four goals: 1) Document the different **projects** and **interactions** with Wikidata; 2) Gain a deeper understanding of the different **uses** of Wikidata based on these projects; 3) explore the **benefits and challenges** using the platform; and 4) document workflows, with an emphasis of identifying specific **features or characteristics** of Wikidata that promote, induce or result in learning.

4.2 Participants

When reaching out to the global Wikidata community, we sought participants that could share "success stories", with new, unique or groundbreaking projects involving Wikidata. We strove for diversity, particularly in four main aspects of projects: 1) *geographic location and languages* used – attempting to go beyond English-centric examples; 2) *discipline / type of institution* – striving to include examples from a variety of institution type (educational / cultural / governmental / research / industry); 3) *types of interactions*, attempting to describe different types of interactions, whether data curation, data extraction, or both; and 4) *types of uses,* looking for a cohort in which different projects reveal different aspects and possible uses of Wikidata.

Six participants were affiliated with educational and research institutions. Four were either affiliated with or worked with cultural institutions (GLAMs). The Participants came from England, Scotland, USA, Israel, Brazil, Germany and Australia. Native languages included English (4), Portuguese/French (1), German (1), and Hebrew (1). Only one participant was female, reflecting a known gender gap in Wikimedia projects (Ford & Wajcman, 2017; Hargittai & Shaw, 2015; Klein et al., 2016; Wagner et al., 2015). While most participants' data will remain anonymous, some information is shared—either because it was already public or by explicit consent.

4.3 Data collection

As three types of institutions were targeted (education, research and culture), three interview protocols were developed, which share key questions with appropriate adaptations.

Interviews were conducted online between January 2019 and June 2021, via platforms such as "Hangout on Air" and "StreamYard". Interviews lasted 60–180 min, with most averaging 90–120 min.

4.4 Data analysis

Interviews were transcribed and thematically coded through the "Dedoose" software, and then analyzed. The coding & analysis included an iterative process – enabling the researchers to reflect on the themes, categories and data collected. Next,

similar codes were converged and categorized to reach a final category tree. Since one interview focused on a project directed by one of the authors, to avoid a conflict of interest and strive for neutrality, this project was only included in the descriptive response to the first research question and excluded thereafter. All other interviews were mapped using a bottom-up thematic analysis, followed by quantitative comparisons. First, each statement was coded (coding was not exclusive so statements could be attributed to several categories). Then, an iterative process was used to group similar codes and refine the category tree. To ensure inter-rater reliability of the coding, 30% of the statements were additionally analyzed by a second coder. Agreement level was high, Cohen's Kappa = 0.94. The data collected was classified into categories, sub-categories, and at times, sub-sub-categories. By analyzing these statistically, insights were derived regarding the uses, benefits and challenges of Wikidata as a learning platform. Specific characteristics of interaction with Wikidata that induce learning were highlighted and discussed, as well as implications for education from a wider perspective of life-long learning.

5 Findings

The seven interviews depict **ten distinct case-studies or projects,** that demonstrate learning opportunities in a variety of disciplines, contexts, locations and institutions. The projects are first presented in context of their value for educators, researchers and learners, and then main uses, benefits and challenges of interacting with Wikidata are presented.

5.1 The projects

5.1.1 Bodleian libraries, university of Oxford, UK: The astrolabe explorer

In 2015, the Bodleian Libraries hired a "Wikimedian-in-Residence, longtime Wikimedian, Dr. Martin Poulter. Poulter was to "undertake academic and public outreach work to encourage understanding and development of Wikimedia projects and improve access to the libraries' collections".[3] For over 4 years Poulter focused on making the libraries' collections more visible in Wikimedia projects; exposing academics, students and the public to the benefits of working with Wikimedia projects; and proactively assisting in closing the gender gap. As Poulter explained, in addition to Wikipedia he increasingly focused on Wikidata, and on how to "tell compelling stories with data". One example is a collection of antique Astrolabes, historic devices used for navigation. To make this collection available to the public in an efficient, engaging, and interactive way, Poulter imported all astrolabes data into Wikidata, then over a lunch break, he created a website, as shown in Figure 1. Each tab showcased automatically generated content from Wikidata, focusing on different aspects

---

[3] https://www.bodleian.ox.ac.uk/bodley/news/2015/mar-26

Fig. 1 The Astrolabe Explorer website

of the astrolabes in a variety of languages (https://tinyurl.com/yc97qmqq). This example was a "proof of concept"—simply replace astrolabes with any collection, upload to Wikidata, and easily generate a similar website telling the story of that collection. When new items are added to Wikidata, the website is automatically updated.

5.1.2 University of Edinburgh, Scotland: The witch-hunts project

In 2014, Ewan McAndrew was hired to serve as the University of Edinburgh "Wikimedian-in-Residence". While widespread in cultural, historical, medical and governmental institutions, this position has been a first for a university. One project that showcases contributions to students' learning is uploading a scholarly database about the 16th–17th
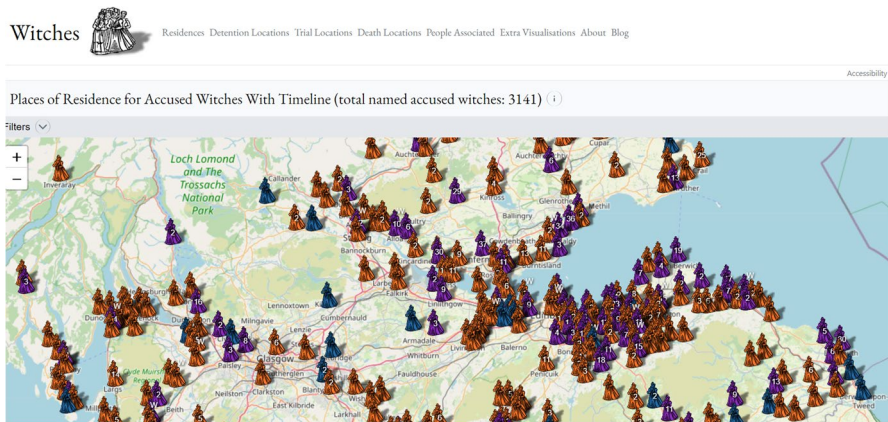
Fig. 2 The Scottish Witch Hunts website

centuries Scottish Witch Hunts to Wikidata. The database was inaccessible to the public and no longer maintained, but held high-quality data from reliable sources curated by researchers. After data was transformed into Wikidata, old locations names were matched with current names and coordinates were added. Once completed, queries were used to display results on an interactive map. A new website was created to tell the story of the witch trials in an engaging, visual and interactive way, as shown in Figure 2 (https://witches.is.ed.ac.uk/timeline/). McAndrew explain that using Wikidata allowed to "breathe new life into it. From a forgotten and unused database… (to) new possibilities for faculty, students and the general public." The project gained much media attention, including this Smithsonian article https://tinyurl.com/y7dvjsdh. Both faculty and students appreciated working with a real dataset, with actual impact. The project encouraged others to add information into Wikidata, and has inspired similar related projects. The university now recognizes Wikidata as a platform that enhances skills, capacities and literacies, and is exploring new ways, additional courses and more databases that could be enhanced by Wikidata.

5.1.3 The Metropolitan Museum of Art, USA: The Portrait of Madam X

The Metropolitan Museum of Art is a leading "encyclopedic" museum, aspiring to showcase the breadth of all human art in a universally accessible way. In 2017, under a new open access policy, the MET released 375,000 images from its collection of over 2 million works under a free license (CC-0) and a Wikimedian-in-Residence supported adding the Met's metadata into Wikidata. In 2018, a new Wikimedian Data Strategist, assisted with an upload of 600,000 artifacts into Wikidata. The goal was to explore how semantically representing the Met's collection in Wikidata can help the Met's physical and virtual visitors explore and learn from a collection of such scale. It was also an investigation of how new technologies can assist in making sense of large data sets. As part of the collaboration, three noteworthy efforts, relevant to education and learning, were undertaken. The first effort, discussed here, was unraveling links, connections and relationships that were not known before. Figure 3 depicts the graphic results of a Wikidata query that demonstrates a connection between the painting "Portrait of Madam X", which inspired the creation of a dress worn by Rita Hayworth in the film "Gilda". Unknown previously to the curators at the Met, only once the painting's metadata was expressed in a structured, linked way on Wikidata, this connection between painting and dress was revealed.

5.1.4 The Metropolitan Museum of Art, USA: The Met's Dashboard

The second notable effort is the creation of the museum's Dashboard, a.k.a. the Met Open Access Portal on Wikidata. This portal (https://w.wiki/Q9J) allows users to explore the Met's collection both statistically and visually. The portal uses a tool called InteGraality (https://tools.wmflabs.org/integraality/) to track the *Completeness* of the collection. The tool automatically generates statistical reports per specific criteria based on queries. Such tools help users explore large-scale data collections, with visual representation of both included and missing items. A potential academic assignment could see students adding missing data to collections in any field (Fig. 4).
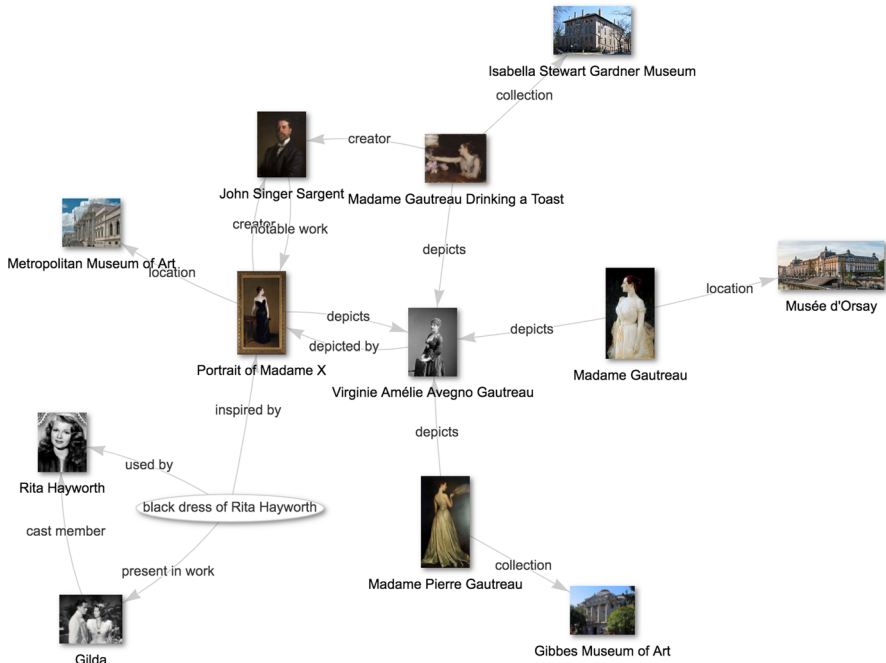
**Fig. 3** Visualizing the connections for the painting "Portrait of Madam X"

Fig. 4 The Metropolitan Museum Dashboard on Wikidata

5.1.5 The Metropolitan Museum of Art, USA: The Depiction Game`

Launched in 2019, the "Depiction" Wikidata Game allows users to make micro-contributions to Wikidata. For this project, the Met's image collection was

Fig. 5 Playing the "Depicts" Wikidata Game at the Met. Credit: Fuzheado, CC-BY-SA 4.0
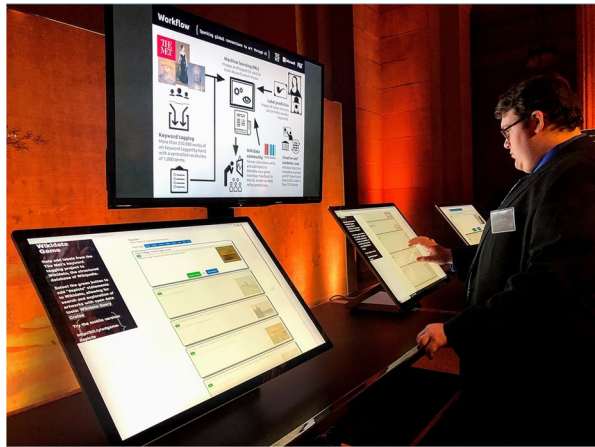
ingested, in collaboration with Microsoft Research, by an Artificial Intelligence (AI) system trained with Met images (https://tinyurl.com/y7aaurpj). The AI algorithm suggests what is depicted in a picture, for instance, a horse. A human playing the "Depiction" game confirms the AI suggestion, as shown in Figures 5 and 6, and if a horse is indeed present, a statement to this effect is automatically added to Wikidata. The game allows a depiction of a variety of objects, such as musical instruments, animals, flowers, vases, etc. This, in turn, allows users to

Fig. 6 The Depiction Wikidata game

explore Met paintings that portray such objects. In the broader perspective, this allows users to get accurate answers to new types of questions, therefore allowing new types of research, not possible before Wikidata.

5.1.6 Tel Aviv University, Israel: An academic course featuring Wikidata

In 2018 a new course opened at Tel Aviv University (TAU): "From Web 2.0 to Web 3.0, from Wikipedia to Wikidata". This for-credit, elective course, the first of its kind worldwide, is available to all undergraduate students at TAU. It focuses on Wikipedia and Wikidata and encourages active learning, while promoting digital literacy, data literacy and academic skills, as well as raising awareness to knowledge gaps, online bias and battling fake news. One of the course's two main assignments is a Wikidata project, which involves curating and extracting data from Wikidata, while presenting it in a visual way. This exposes students to issues such as ontologies, data modeling and basic querying skills, as well as data visualization, gaps, bias, sourcing and completeness, thus strengthening the students' data literacy. A Wikidata project could have students exploring the gender equality among faculty members by creating a query checking how many female faculty members are included in Wikidata, finding gaps, adding missing data based on reliable sources; then re-running the query, watching data being added visually. For its second 2020 iteration, the course was revised based on students' feedback, faculty insights, and the impact of COVID-19. The course is now virtual and more focused on Wikidata (Fig. 7).
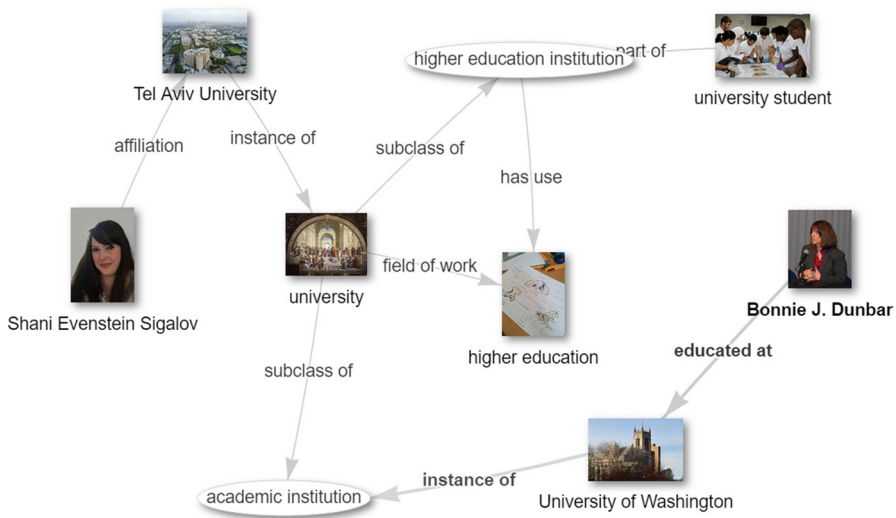
Fig. 7 Example from a TAU class exercise: mapping relationships of 9 items connected to the course in a graph format using a Wikidata query

5.1.7 The School of Journalism, Faculdade Cásper Líbero, Brazil: The Municipal elections case

João Alexandre Peschanski, a professor at the School of Journalism, Faculdade Cásper Líbero, São Paulo, and a researcher at the Center for Neuromathematics at the University of São Paulo, worked with his students to answer the question: "**How can we efficiently and effectively improve content on municipal elections in Brazil on Wikipedia?** While creating election-related Wikipedia articles is important, editing these articles can be tedious, boring and therefore susceptible to human errors. However, bots can do this work easily enough, and the result was a tool that automatically generated Wikipedia articles based on structured data in Wikidata. These articles include not only tables but textual paragraphs that were automatically generated by a template. The final article included 2 empty sections, ready for humans to add details to. Peschanski stated that these articles have already been viewed over 50 million times, and proved to be quite needed and impactful (Fig. 8).

A simpler version of this technique was previously used in order to auto-generate simpler Wikipedia articles about Works of Art, Museums, Libraries, Archives, Theaters, Books, Movies, Earthquakes and Newspapers (https://w.wiki/Peg). The technological tool that enables this is called MBabel (https://w.wiki/NAq). It was adapted from a project originally done at the Metropolitan Museum, and improved upon by the Brazilian community.

A more advanced version of this technology now allows Peschanski and his students to automatically generate a semantic WikiBook – an open textbook (open educational resource, or in short, OER), about a collection in one of Brazil's museums. The WikiBook is created mainly by contributing data about

Fig. 8 Wikipedia article auto-generated by Wikidata-based templates
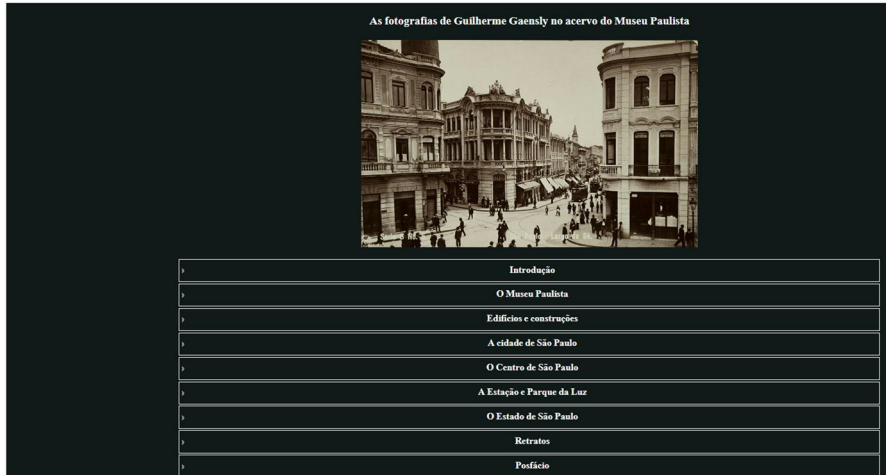
Fig. 9 A Wikibook generated via Wikidata

pictures in the museums catalogue, or by playing a simple Wikidata game, which in turn contributes to the data curated in this open educational resource. The data added to Wikidata is then extracted using queries and added into templates that generate the WikiBook (https://w.wiki/J$R) (Fig. 9).

Using Wikidata as a means to contribute to other Wikimedia projects, such as Wikipedia and WikiBooks, substantiates that Wikidata could be used to generate a much more structured Wikipedia. In this sense, every single article about a notable personality in all 300 language versions, should have the same datum of "date of birth". As far as Peschanski is concerned, everything that could be automatically generated by a bot, should be done that way, so volunteers or students who write articles can focus on more exhilarating work and on details that are not yet structured. This is especially important for smaller language communities that do not have enough volunteers to generate Wikipedia articles and other Open Educational Resources in their own language.

5.1.8 School of Journalism, Faculdade Cásper Líbero, Brazil: Reconciling data from heterogeneous databases case

Peschanski and his students also used Wikidata to reconcile data from heterogeneous databases. Since databases external to Wikimedia can have disagreeing data, humans examined the work of bots making informed decisions. For example, a highly ranked page on Portuguese Wikidata that automatically curates all the people who were killed or missing during the military dictatorship (https://w.wiki/JvF). The page is auto-generated via Listeria bot, based on a query from Wikidata, and the table hosts references for each personality. In Figs. 10 and 11, two different references provide conflicting information. Wikidata curates and displays the different pieces of information, alleviating potential misinformation.

Fig. 10 Wikidata-generated list of killed or missing individuals, demonstrating data based on 4 sources with 2 different death dates

Fig. 11 Wikidata statement about conflicting dates, with their sources

The aggregates data in Wikidata enables research about the reliability of the sources, statistically examining the quality and accuracy of different databases. Bots flag disagreeing sources and students intervene and determine the right answer. The skills gained in such projects are an important part of information and data literacy, which equip students to be better consumers of information in a reality of "fake news" and "post-truth".

5.1.9 Brazil: Digitally recreating lost Museum artifacts

In 2018 Brazil's national museum was completely destroyed in a fire. The museum's collection was not entirely digitized, and the files were also consumed in the fire. The loss for Brazilian culture was so immense that a group of Wikimedia volunteers

Fig. 12 Digitally reconstructed the Afro-Brazilian room

started a process, referred to as "Data Archaeology", to recreate the museum digitally. First, a crowd-sourcing technique was used, asking the public to upload pictures taken at the museum. Then, Wikidata was used to curate information on lost objects via a tool called "tabernacle" (https://tools.wmflabs.org/tabernacle/), which curates structured data in a tabular, multi-lingual, and visual way (https://tinyurl.com/4msa67pc). This digital recreation reveals another use of Wikidata that can be critical for educators and researchers of cultural heritage (Fig. 12).

5.1.10 Germany, Australia, Brazil: Tracking the COVID-19 pandemic with Wikidata

A COVID-19 portal on English Wikipedia (https://w.wiki/QNe) showcases a table tracking the disease progress in different countries. The portal exists in many languages, and in Portuguese is automatically generated and updated according to data added into Wikidata. This process pulls information from diverse databases and sources, the data is being curated in Wikidata and then utilized by Wikipedia in all languages. A Google search also directs users to Wikipedia, which sources its data from Wikidata. This is probably one of the most accurate and reliable online sources on COVID-19, and leads the way toward an online ecosystem that produces query-based digital items (Fig. 13).

Another example of creating a new digital object based on data curated in Wikidata is a collection of Covid-19-related queries created by an Australian academic. The queries allow users to explore notable cases by occupation, age distributions, and birthplace maps, in a visual and engaging way (https://tinyurl.com/y7xbk7ul) (Fig. 14).

COVID-19 cases, deaths, and rates by location[21]

	Country	Deaths / million	Deaths	Cases
⊕	World[a]	860	6,862,848	673,941,526
	Peru	6,441	219,332	4,484,651
	Bulgaria	5,633	38,204	1,296,200
	Bosnia and Herzegovina	5,030	16,267	401,527
	Hungary	4,889	48,738	2,194,102
	North Macedonia	4,610	9,653	346,733
	Georgia	4,526	16,950	1,822,045
	Montenegro	4,465	2,800	287,421
	Croatia	4,452	17,945	1,268,595
	Czech Republic	4,038	42,375	4,601,070
	Slovakia	3,720	20,996	2,663,604
	Moldova	3,655	11,964	600,916
	San Marino	3,621	122	23,521
	Lithuania	3,479	9,570	1,301,860
	Romania	3,441	67,650	3,331,005
	Gibraltar	3,396	111	20,414
	Latvia	3,379	6,255	975,767
	Greece	3,348	34,779	5,548,487
	Slovenia	3,333	7,067	1,325,505
	United States	3,303	1,117,497	103,124,043
	Chile	3,269	64,085	5,149,301
	Brazil	3,242	698,050	36,987,682

Fig. 13 Tracking the pandemic's progress

5.2 The uses

Our second research question aimed to extract and map different **Wikidata uses** in educational, research and cultural contexts by early adopters. The goal was to unravel patterns of current uses and examine Wikidata's potential as a learning platform for education and research in all disciplines. A thematic analysis was performed on 6 interviews, covering 9 out of the 10 projects (excluding the

Fig. 14 A query of notable cases of COVID-19 by occupation visualized in a bubble chart

project in 5.1.6). Coding and analysis of these projects revealed eight main uses of Wikidata that induce learning (n = 435, 41% of the total statements):

1. **Connecting, modeling and cataloguing data from separate sources –** this is the most prevalent use and core ability of Wikidata, allowing users to describe items using rich data, based on a variety of sources. In doing so, Wikidata is serving as a hub of information, aggregating information from a variety of never-before-connected sources of information.
2. **Using Wikidata to make knowledge & culture freely accessible –** this use, almost as frequent, touches on the basic ability of the platform to make knowledge freely accessible to everyone. Due to the use of a Creative Commons license, CC-0, Wikidata is one of the largest Open Educational Resources in existence, sort of a "big data" reservoir that freely available to the public.

3. **As educational platform for teaching & learning –** this use showcases Wikidata utilization in the classroom, enhancing students' skills and generating meaningful work with social impact. In this case, Wikidata is used as a learning platform, similarly to Wikipedia, a platform through which learners gain subject-matter-relevant knowledge, as well as skills, literacies and capacities.

4. **Creating new digital objects that did not exist before –** once structured linked data exists, it can be queried and visualized, thus creating new digital objects that did not exist before. Examples include the Wikibook and the list of killed and disappeared.

5. **Answering new questions & surfacing unknown connections –** with structured, linked data, we can answer hard, or even impossible, questions and reveal new connections between pieces of information – previously unknown. Each datum is described separately, but because it is linked, queries can help reveal unknown relationships between pieces of information, much like in the case of "Portrait of Madam X".

6. **Salvaging data that otherwise would be lost –** "data archeology" can be used to reconstruct lost physical items, or databases on the verge of extinction, presenting a sustainable, central solution to salvaging data and giving it new life, as was the case with the burnt Brazilian museum, or the witch hunts in Scotland.

7. **Using Wikidata to improve external Databases –** Wikidata games use the 'wisdom of the crowd' to map items and improve, or correct, institutional metadata. As many institutions do not have the resources to properly map and update the metadata related to their collections, Wikidata games that allow users' "micro-contributions" through gameplay, is an important step in engaging the public in helping institutions improve their metadata. Players do not need to know anything about Wikidata, but their contributions not only help improve the databases but also allow new types of research through Wikidata.

8. **Auto-generating new content –** least frequent, but important nonetheless, is Wikidata's ability to assist constructing Wikipedia articles or WikiBooks. The system creates an article outline, with volunteers adding aspects that cannot be addressed by machines (Table 1).

A chi-square goodness for fit test, which compared the observed sample distribution with the expected probability distribution based on the proportion of statements in each sub-category, was statistically significant, X^2 (7) = 141.75, $p = 0.000$. The discrepancy between the observed and expected frequencies is used to determine which cells within the contingency table generate residual scores that are larger in magnitude than might be expected by chance (Hadad et al., 2021; Sharpe, 2015). The standardized residual presented in Table 2, shows the degree to which an observed chi-square cell frequency differs from the value expected in the interviews based on their data.

As indicated in Table 2, significant differences were found between the proportion of statements in each sub-category. While the combined top three sub-categories cover over 63% of the coded statements depicting uses, the rest of the sub-categories were much less frequent. Considering the small sample size

Table 1 Mapping uses per project

	Projects Uses	5.1.1 Astrolabe	5.1.2 Witches	5.1.3 Madam X	5.1.4 MET Dashboard	5.1.5 MET Game	5.1.6 TAU Course	5.1.7 Brazil Elections	5.1.8 Reconcile data	5.1.9 Burnt Museum	5.1.10 COVID 19
1	Connecting separate sources	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆
2	Making knowledge accessible	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆
3	Educational platform	◆	◆	◆	◆		◆	◆	◆		
4	New digital objects	◆	◆	◆	◆		◆	◆	◆	◆	◆
5	Answering questions unraveling connections	◆	◆	◆			◆	◆			◆
6	Salvaging data		◆					◆	◆	◆	
7	Improve external databases	◆	◆	◆	◆	◆					
8	Auto-generating content							◆	◆		

Table 2 Uses in numbers

Category	Sub-category	Number of statements ($N=435$)	Standardized residuals
Uses	1. Connecting, modeling and cataloguing data from separate sources in a structured way	101 (23.22%)	+6.32
	2. Using WD to make knowledge & culture freely accessible	92 (21.15%)	+5.1
	3. Educational platform for teaching & learning	82 (18.85%)	+3.75
	4. Creating new digital objects that did not exist before	51 (11.72%)	-0.046
	5. Answering new questions & surfacing unknown connections	39 (8.97%)	-2.09
	6. Salvaging data that otherwise would be lost	35 (8.05%)	-2.63
	7. Using WD to improve external Databases	20 (4.59%)	-4.66
	8. Auto-generating content	15 (3.45%)	-5.34

Uses ($N=435$, 41%); X^2 (7) $=141.75$, $p=0.000$

1. $N=$ total number of statements in the main category.

2. Percentage is out of all statements.

3. The rule of thumb for interpreting Standardized Residuals is a standard score beyond the range of ± 2, then that cell can be considered to be a major contributor (Sharpe, 2015)

and the innovative nature of several projects, it is expected that some uses will be less frequent. Frequency, then, does not imply significance or importance of use, as will be discussed below.

5.3 Benefits and challenges

Analyzing the interviews and projects revealed benefits of using Wikidata that encourage and support learning, and challenges that should be addressed or considered. Benefits included 485 statements (46% of total). Challenges were less frequent (as expected) and included 132 statements (13%). A chi-square goodness for fit test examining the benefits and challenges, was statistically significant, X^2 (1) $=200.82$, $p=0.000$. Three main categories emerged and repeated for both the benefits and challenges: outreach, education and platform related statements, though with alternating order of frequency.

5.3.1 Benefits

Examining the benefits of interacting with Wikidata (N = 485, 46%)**,** the order was: **outreach-related** (n = 230, 47.42% of statements), **Education-related** (n = 165, 34.02%) and **platform-related** (n = 90, 18.56%). A chi-square goodness for fit test was statistically significant, X^2 (3) = 65.86, p = 0.000. Additional chi-square goodness for fit tests were performed on each sub-category and sub-sub category, and were statistically significant.

Outreach-related benefits appeared most frequently, attesting to the need for demonstrating the advantages of Wikidata to institutions and stakeholders. Sub-categories included: highlighting the benefits of making collections accessible to increase the completeness, quality and reliability of data, allow a positive social impact, and offer a multitude of uses and applications; allowing the discovery of information that otherwise would be hidden or inaccessible; using an external plat-form as more cost-effective solution, compared to developing one in house; and the ability to extract specific details and tell compelling stories with data.

Education-related benefits showcased the significant educational potential this platform has for educators and learners. Sub categories revealed that being able to engage with data motivated users to contribute, especially knowing their work will last and benefit others. It also revealed that engagement with Wikidata helped improve different skills and highlighted new opportunities for faster collaboration in Academia.

Platform-related benefits were the least frequent, focusing on data visualization and the ability to easily explore information. Additional benefits were the multilin-gual nature of the platform, overcoming language barriers and even engaging with machines; the power of a cross-disciplinary, global community to work with; the ability to flexibly model items and reconcile different sources of information; and the tools built around Wikidata, which allow users to work more efficiently and scale their efforts.

The category tree for all benefits of interacting with Wikidata is presented in Table 3. The standardized residual score shows the degree to which an observed chi-square cell frequency differs from the value expected in the interviews, based on their data for the categories, sub-categories and sub-sub-categories.

5.3.2 Challenges

Examining the challenges of interacting with Wikidata (N = 132, 13%), the order of the main categories was: **platform-related** (n = 77, 58.33% of statements), **outreach-related** (n = 45, 34.09%), and **education-related** (n = 10, 7.58%). A chi-square goodness for fit test, was statistically significant, X^2 (2) = 51.05, p = 0.000. Additional chi-square goodness for fit tests were performed on each sub-category and were statistically significant for the platform-related and outreach-related sub-categories. A Chi-squared test was not performed on the education-related sub-category, as their values did not meet the assumption of the test.

The high frequency of **platform-related** challenges**,** especially considering it was the lowest scored benefits category, attest to the complexity of this relatively new

Table 3 Benefits in numbers

Category	Standardized residuals	Sub-category 1	Standardized residuals	Sub-sub-cat 2	Standardized residuals
Benefits (N=485, 46%)					
Outreach-related benefits – N=230 (47.42%)	+5.37	Advocacy for Open: sharing as a default – N=75 (15.46%)	+2.31	Better completeness, quality & reliability as more is shared – N=29 (5.98%)	+0.8
				Positive social impact – N=25 (5.15%)	0
				Diverse uses & applications for different stakeholders – N=21 (4.33%)	-0.8
				$X^2(2)=1.28, p=0.527$	
		Discoverability of info – N=70 (14.43%)	+1.65		
		Sustainability and cost-effective over time – N=43 (8.87%)	-1.91		
		Storytelling: telling a story with data – N=42 (8.66%)	-2.04		
		$X^2(3)=15.88, p=0.0012$			
Education-related benefits – N=165 (34.02%)	+0.26	Allows engagement with data – N=77 (15.88%)	+5.57	Allows innovation and exploring new frontiers – N=24 (4.95%)	+3.05
		Motivations for participation –N=52 (10.72%)	+1.67	More fun / interesting to engage with and has real impact on the world – N=18 (3.71%)	+1.39
				Allows quality assessment, research & data-driven solutions – N=5 (1.03%)	-2.22

Table 3 (continued)

Category	Stand-ardized residuals	Sub-category 1	Stand-ardized residuals	Sub-sub-cat 2	Stand-ardized residuals
				Sense of pride for producing something tangible, durable – N = 5 (1.03%) $X^2 (3) = 21.08, p = 0.000$	-2.22
		Improved Data Literacy and other skills – N = 29 (5.98%)	-1.91		
		Allows faster, collaborative, academic work – N = 7 (1.44%) $X^2 (3) = 65.86, p = 0.000$	-5.33		
Platform-related benefits – N = 90 (18.56%)	-5.64	Data visualized and easily explored – N = 45 (9.28%)	+6.36		
		Overcoming language barriers, going beyond human languages – N = 22 (4.54%)	+0.94		
		Allows working with a cross-disciplinary, global community to engage and learn from – N = 10 (2.06%)	-1.89		
		Allows flexible modeling and reconciling different sources of information – N = 8 (1.65%)	-2.36		
		Allows Humans-Machines collaboration and use of tools to scale work – N = 5 (1.03%) $X^2 (4) = 59.89, p = 0.000$	-3.06		

Table 3 (continued)

Category	Stand-ardized residuals	Sub-category 1	Stand-ardized residuals	Sub-sub-cat 2	Stand-ardized residuals
Benefits (N=485, 46%); X²(2)=60.72, p=0.000					

1. N = total number of statements in the main category, including percentage out of all statements

2. The rule of thumb for interpreting Standardized Residuals is a standard score beyond the range of ± 2, then that cell can be considered to be a major contributor (Sharpe, 2015)

Table 4 Challenges in numbers

Category	Standardized residuals	Sub-category 1	Standardized residuals
Challenges – N=132 (13%) Platform-related – N=77 (58.33%)	+4.97	Unknown, complex platform that requires skills for engagement – N=58 (43.94%)	+6.38
		Modeling challenges & biases – N=12 (9.09%)	-2.7
		Platform limitations – N=7 (5.3%)	-3.68
		$X^2(2)=61.58, p=0.000$	
Outreach-related – N=45 (34.09%)	+0.15	Need to persuade about benefits – N=24 (18.18%)	+5
		Tracking Incompleteness – N=7 (5.3%)	-0.67
		Fear of losing control – N=6 (4.55%)	-1
		The mental burden of working in less resourced countries – N=5 (3.79%)	-1.33
		Not being paid for digital labor – N=3 (2.27%)	-2
		$X^2(4)=32.22, p=0.000$	
Education-related – N=10 (7.58%)	-5.13	Hard to motivate students to invest in it – N=5 (3.79%)	--*
		Pace of course design changes is slow – N=3 (2.27%)	--*
		Address needs from various stakeholders when used as a learning platform – N=2 (1.52%)	--*
		--*	
Challenges (N=132, 13%); X²(2)=51.05, p=0.000			

1. N = total number of statements in the main category, including percentage out of all statements

2. The rule thumb for interpreting Standardized Residuals is a standard score beyond the range of ±2, then that cell can be considered to be a major contributor (Sharpe, 2015). An asterix (*) was used to indicate when a Chi-squared test was not performed on the table cells that did not meet the assumption of the test

platform. It seems that there is still a high threshold for engagement, requiring specific skills to use the platform to its full potential. Modeling issues, specifically trying to model challenging items or addressing biases, as well as the platform's own limitations, emerged as additional challenges.

The majority of **outreach-related** challenges related to the need to persuade others of the benefit of Wikidata in order to implement its use. Additional challenges included the need to track incompleteness of datasets to clarify which portion of topic mapping has been achieved; fears expressed by experts from cultural institutions or Academia of "losing control" of their contributions; and the mental burden of volunteers in less resourceful countries who felt they had to do all the work themselves or it will never be done.

Education-related challenges were the least frequent. It seems that users found more benefits than challenges for incorporating Wikidata as a learning platform. Sub-categories highlighted challenges with: students' motivation to invest in Wikidata; the complex and time-consuming task of implementing Wikidata into academic curriculum; the slow pace of changes to course design; and the need to address a variety of, sometimes conflicting, needs of different stakeholders (students, faculty, institutions, Wikidata community).

The category tree for all challenges of interacting with Wikidata is presented in Table 4.

6 Discussion

For years, the captivating idea of a Semantic Web inspired various attempts to realize this dream of a `web of data`, one that both humans and machines can access and make use of. But the Semantic Web is no longer a dream. Wikidata, Wikibase (the open-source platform Wikidata is based on, similarly to Mediawiki, Wikipedia's platform), and similar Semantic or Linked Data projects, have forever changed the interactions between humans and knowledge, creating new learning opportunities for their users, both in and outside of classroom. Considering the plea for action from the research community to explore the potential of the semantic web, specifically, the lack of imperative research about semantic networks and Linked Data platforms as learning platforms, this study aimed to investigate Wikidata's value for education and research in its broad sense – not only in the classroom, but rather as a lifelong learning platform for diverse disciplines, contexts and narratives.

We examined noteworthy projects from around the world that showcase how early adopters are interacting with Wikidata and using it as a learning platform. Thematic analysis exposed different uses of the platform, as well as benefit and challenges that emerged from two main interactions: **Data Curation**, adding data into Wikidata, in order to curate, salvage, and enrich datasets; and **Data Extraction**, querying Wikidata to answer difficult (or impossible till now) questions, visually examining data in an engaging way, and exploring relationships and expose connections previously unknown. The analysis has also revealed two additional interactions. The first was **Data Creation or Auto-generation,** using Wikidata to create new digital objects based on scattered external data, as well as auto-generating content on

other Wiki projects, thus freeing humans to work on less technical tasks. The other interaction was **Teaching with Wikidata:** using Wikidata as a teaching and learning platform, sharpening not only learners' digital literacy, but also promoting Data Literacy, which included touching on skills like data modeling, ontologies, critical thinking, and data analysis. This aspect of Wikidata helps fight misinformation, disinformation and fake news, for example, by reconciling contradicting sources; and finally, Wikidata assists in teaching related topics such as "Semantic Web", "Linked Open Data" and "Digital Humanities".

Analysis revealed that the four **interactions** described **led to eight different uses** explored above, out of which various **benefits and challenges** of engaging with the platform were mapped, as described in the findings. Considering the uses and benefits in light of the three pedagogical frameworks presented above, constructivism, collaborative learning and self-directed learning / heutagogy, is appears that Wikidata is an ideal platform to induce learning. First, Wikidata is built through a collaborative effort of a global community. As suggested by **Constructivists**, knowledge and meaning are "constructed rather than given" (Parker & Chao, 2007), through a "discussion with peers… and through reflection" (Higgs & McCarthy, 2005). The focus on solving real-life problems helps users "develop capacity for effective problem-solving behaviors (Anderson, 2016)". Adding data to Wikidata requires users to engage in constant dialogue and negotiation of how to correctly describe the world, while taking into account the multitude of global perspectives. In order to be equitable and inclusive in describing our diverse and complex world, users in the community constantly rethinks the ontology, the modeling schemes for certain items, and how to better represent complex knowledge.

As suggested by the **Collaborative Learning** framework, it is specifically the engagement with a technological collaborative platform, that encourages a deeper engagement with information, stimulates a higher-level thinking, and a longer information retention (Boulos et al., 2006; Galway et al., 2014; Johnson & Johnson, 1986; Parker & Chao, 2007; Schaffert et al., 2006; Wheeler et al., 2005). The process of extracting information from Wikidata is reliant not only on the technology, but also on the community, and specifically on collaborating and learning from others, as new users to the platform seldom know SPARQL, a Semantic Web programing language that allows querying Wikidata. Users of the platform often use existing query examples, as well as community experts, to learn how to write required queries and gain insights from this vast knowledge-base.

The **Self-directed learning** and Heutagogy frameworks also suggest that Wikidata is a platform that promotes learning. As noted in the literature, there seems to be a shift in focus from instructing and testing competences, toward equipping learners with skills and literacies that teach them how to learn (Anderson, 2016). In a world where the structure of communities and workplace is constantly changing and new knowledge is rapidly emerging, more efforts are invested in learners gaining skills, competencies and literacies that allows them to engage with information and data in unfamiliar contexts as a lifelong process (Hase & Kenyon, 2000). Anderson stresses that self-direction and the focus on developing skills is highly connected with "utilizing the online tools and information available" (Anderson, 2016). And indeed, no matter the type of interaction, engaging with Wikidata drives

users to engage with its ecosystem of technological tools, which improve various workflows. Researchers also highlight self-motivation as key prerequisite for a successful engagement. While a detailed exploration of Wikidata users' motivation is outside our scope and will be explored in future research, an examination of the uses and benefits suggest high levels of self-motivation and engagement in self-directed learning. Specifically, it seems that the ecosystem of additional tools is important to users in improving and enhancing various workflows. To sum up, the collaborative, technology-based and tools-reliant, self-motivated effort to engage with the platform makes Wikidata an effective and propitious learning platform that allows users to gain both knowledge and skills on an ongoing basis.

Further examining the different uses and their benefits, it seems that Wikidata has some **key features or characteristics** that encourage and enable learning, as well as the improvement of digital and data literacies. The first notable feature is **Data Visualization.** The different projects suggest that some of Wikidata's relevance and value for education stems first from the ability to get accurate answers to questions previously difficult or impossible to address. More specifically, findings suggest that it is the ability to *visualize* the results that seems to be one of the most important features of Wikidata for education, research and learning. While "Data visualized and easily explored" was coded in only 45 statements (9.28%), many of the benefits described rely heavily on data visualization, including "Advocacy for Open" (75, 15.46%), "discoverability" (70, 14.43%), "storytelling" (42, 8.66%), "engagements" 77 (15.88%), "motivations" ("fun to engage with", 18, 3.71%), and "improved Data Literacy and other skills" (29, 5.98%). Combined, these benefits add up to 356 statements, or 73% of all benefits (34% of all statements). Data Visualization, then, appears to be a key element in Wikidata's power as a teaching and learning tool – it allows us to explore not only what is there, but also what is missing, as well as learning through context. Visualizations of structured, linked data allows us to tell stories in new and engaging ways, making sense of the abundance of data, and in turn, of our world.

Another characteristic that emerged from the findings is that using Wikidata **promotes higher-order & critical thinking.** While only 29 statements (5.98%) were directly coded as "Improved Data Literacy and other skills", interviews and thematic analysis revealed that other benefits either rely on, or result in, higher-order & critical thinking, such as contemplating on or dealing with "completeness, quality & reliability" (29, 5.98%), "social impact" (25, 5.15%), "diverse uses and applications for different stakeholders" (21, 4.33%), "discoverability of info" (70, 14.43%), "storytelling" (42, 8.66%), "engagements" 77 (15.88%), "motivations" (52, 10.72%), "overcoming language barriers" (22, 4.54%), "flexible modeling and reconciling sources of info" (8, 1.65%) and "human–machine collaborations and use of tools to scale" (5, 1.03%). Combined, these add up to 380 statements, or 78% of all benefits (36% of all statements). Thus, interacting with Wikidata (whether via curation, extraction, creation or teaching), drive learners to deal with higher-order thinking and questioning of a given topic.

It seems that various **data-related issues,** such as data modeling, data verification, systematic bias, data manipulation, data access, and data completeness, **become clearer to learners, as they see data visualized**, for example, a map with

missing areas, or a timeline with missing information. Simply put, the abilities to answer questions and visualize data seem to encourage users to apply critical thinking regarding the results they encounter. It is worth noting that while many of the projects highlight learning benefits that result from extracting data and visualizing it, some benefits emerged from curating information in Wikidata. These benefits include addressing ontological issues, such as how to best model items, how to make sure the hierarchies of information make sense, and how similar objects can be consistently represented in the database. A deeper understanding of data modeling was also reported to assist in critically analyzing query results. Dealing with modeling of items made users acutely aware of querying limitations, realizing that the exact way one models an item will affect the way the data would, or would not, be discoverable in a query or autogenerate content.

Both of the characteristics discussed above, Data Visualization and Higher-order & critical thinking, seem to promote **improved digital and data literacies**. This could include, among other things, issues relating to data modeling, data analysis, data verifiability, date completion, and systematic data bias. Some examples include: 1) dealing with mass-uploads of data involving other capacities such as 'data wrangling', (transforming and mapping data from one "raw" data form into another), the need to "clean" datasets, and prepare them for upload in a structured, linked way; and 2) modeling items on Wikidata and working on ontological issues drive users to find the right hierarchies for information, deconstruct and analyze the world—just to reconstruct it in a structured, yet flexible way. This, in turn, makes users wiser, more informed, consumers of knowledge, leading to better digital citizens. Interacting with Wikidata, therefore, plays an important role in stimulating structural and organized thinking as well as critical thinking—one of the key skills for survival in the digital age. Educators must consider that modern learners do not have to work hard to get answers; they simply ask Siri, Alexa and other AI-based digital agents. But do users ever stop to evaluate the answers they get? In a world of `post-truth`, in which dealing with `fake news` and even `deep fake` is part of being digital citizens, it is essential to equip users with skills to evaluate and analyze data. It appears that Wikidata can assist learners develop these necessary skills.

It is important to note that despite the many benefits, interacting with Wikidata and other Semantic platforms is far from perfect and can hold challenges that may hinder learning. As the findings suggest, some criticism of the platform include: a high threshold for newcomers—both in modeling, tools and knowledge of programming; problematic and inconsistent modeling; missing data; missing references; inability to rate good sources of information; data bias; poor documentation of tools; and lack of platform interactivity. That said, despite these growing pains, interacting with Wikidata, via all uses, benefits and key features presented in this research, suggests that Wikidata holds a variety of learning opportunities. The platform appears to drive users toward improving critical thinking and acquiring a higher level of Data Literacy.

7 Conclusion

An anonymous famous proverb, inspired by an Antoine de Saint Exupéry's text says, "If you want to build a ship, don't drum up the men and women to gather wood, divide the work, and give orders. Instead, teach them to yearn for the vast and endless sea." It seems that Tim Burners-Lee's vision of the Semantic Web has inspired such yearning for a world in which humans and machines can make use of the vastness of data available in new, more informed ways that advance humanity. Wikidata cannot fulfil the dream of a Semantic Web by itself, as it requires an ecosystem of structured-data-driven websites that are connected to each other. That said, it stands as an important step forward in a reality where humans and machines can have easier and more meaningful access to data. Additionally, Wikidata has its limitations, in terms of what could be structured in it. Wikidata cannot contain everything possible, and not everything could be structured or modeled in it. Nevertheless, it is still an important milestone for humanity, one that keeps inspiring further technological developments, including recent AI advancements such as ChatGPT, despite various challenges.[4]

One important aspect of Wikidata that has not been fully explored in this paper is its open license, meaning that data modeled in it is considered an Open Educational Resource (OER). The term OERs was defined by UNESCO back in 2002 as "teaching, learning, or research materials that are in the public domain or released with an intellectual property license that allows free use, adaption, and distribution" (UNESCO, 2002). Thus, an OER is defined primarily (though not exclusively) by its license, with Creative Commons licenses being the most widespread. For some educators, the main incentive for using OERs is minimizing textbooks' cost, still a financial burden in many countries (Hegarty, 2015; Lin, 2019). For some, it is the desire to create a ubiquitous, mobile learning experience by accessing materials anywhere, anytime (Hegarty, 2015; Lin, 2019). For others, the preference for OERs is part of a wider pedagogical, if not ideological, perception that values OERs not only as a means of knowledge equity, but also as means to acquire relevant skills, competencies, capacities and literacies in a world where learners are also digital citizens (Cronin & MacLaren, 2018; Evenstein Sigalov & Nachmias, 2017; Hegarty, 2015; Lin, 2019; Wiley & Hilton, 2018). Using emerging open technologies, for both knowledge acquisition and knowledge creation, entails gaining relevant skills for 21st century learners.

Using Wikidata to create OERs and improve skills also connects to UNESCO's framework introduced in 2015, called the "Sustainable Development Goals" (SDGs).[5] The SDGs are a collection of 17 global goals that were designed to be a "blueprint to achieve a better and more sustainable future for all", and were approved by the UN's General Assembly in 2017[6.] Out of the 17, goal number 4 is

---

[4] For further exploration of Semantic Web and bias in AI, it is recommended to refer to the following article: Reyero Lobo, P., Daga, E., Alani, H., & Fernandez, M. (2022). Semantic Web technologies and bias in artificial intelligence: A systematic literature review. *Semantic Web*, (Preprint), 1-26.

[5] https://sdgs.un.org/goals

[6] https://en.wikipedia.org/wiki/File:A_RES_71_313_E.pdf

focused on "Quality Education", with the full title being "Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all" (UNESCO, SDG 4). Goal number 4, then, highlights the importance of open and equal access to education and educational resources (UNESCO, SDG 4). More specifically, it highlighted the role of **Open Education** (OE), sometimes referred to as **Open Education Practices** (OEP), **Open Pedagogy** (OP), or simply OERs, in achieving this SDG (Jha et al., 2019; Lane, 2017; Ossiannilsson, 2019; Tlili et al., 2020; Urbančič et al., 2019). Fulfilling the 4th SDG, Quality Education, through OERs, ties strongly to the final characteristic or key feature highlighted while reviewing the uses and benefits—Wikidata seems to support knowledge equity, by **empowering less established communities.** This is done by offering opportunities to undertake projects with positive social impact, and to **semi-automate creation of content** in local languages, especially those with less volunteers to support them, thus overcoming knowledge gaps and language barriers.

The latter is especially important considering the emergence of **Abstract Wikipedia** and **Wikifunctions**. Abstract Wikipedia, approved in July 2020, is "a strategic effort and a new Wikimedia project".[7] It is "an extension to Wikidata that aims to create a language-independent version of Wikipedia using its structured data" (from Wikipedia). It aims to expand the range of what could be expressed *with Wikidata*. This allows overcoming language barriers by structuring bigger portions of Wikipedia articles, thus enabling auto-generation of content translation into languages with smaller communities. Wikifunctions, announced in December 2020, is " a collaboratively edited catalog of computer functions that aims to allow the creation, modification, and reuse of source code, closely related to Abstract Wikipedia" (from Wikipedia). It is meant to help express in a structured way, facts that currently are impossible to express via Wikidata, due to current structure limitations. An example of such structural limitation was given by Vrandečić, who is leading the development of Abstract Wikipedia and Wikifunctions, during an interview on the future of Wikidata and Abstract Wikipedia. He notes that in the case of Marie Curie, in most Wikipedia articles, the intro narrative would usually mention that she is the only person to receive 2 Nobel Prizes, in Physics and Chemistry. But it is not as simple to structure this fact on Wikidata. While the platform allows to structure the fact that she received a Nobel in both these disciplines, it is currently impossible to demonstrate in a structured way the uniqueness and importance of her double-win. This is something Wikifunctions will allow users to do in the future. Wikidata, then, is at the heart of these, and many other, future advancements. It is an important part of the data ecosystem and a catalyst for the Semantic Web and Linked Data initiatives, especially with Wikibase being increasingly adopted by other institutions.

To conclude, in its ten years of existence, Wikidata has shown great potential that we are merely beginning to explore. The implications, from Education, through Research, to actual applications for industries, is at burgeoning phase, and though appear propitious, additional research is required to fully explore them. It is hoped that despite its limitations, this research will be a stepping stone in investigating

---

learning with semantic networks. It is also hoped that this research will inspire educators to experiment with Semantic Web and Linked Data platforms and applications as a learning tool, implementing them into the academic curriculum. Finally, it is hoped that the findings of this research will encourage further investigation by researchers, institutions and industries, contributing to future semantic applications, leading towards a more sophisticated future, where the existing data is better utilized for the benefit of learners and the general public.

8 Research limitations

This research aims to shed light on the potential and value Wikidata has for educators and learners around the world. Only 7 users and 10 use cases or projects were discussed in this specific paper. When a small sample is concerned, there is always a chance that some descriptive elements may be used to induce from the specific to the general in an inaccurate way. Moreover, even though an emphasis was put on finding diverse example, some users or existing cases were not discussed and fell out of scope for this research. It is especially challenging to determine whether the diversity reached in the sample represents the larger population of Wikidata users, mainly since most users are unknown or might be unreachable. Therefore, while there is value in describing this phenomenon, larger-scale research is needed, that might analyze the topic from quantitative lenses. Finally, this field of research is rapidly evolving, constantly changes and is influenced by other technological advancements. It is not unlikely that at some point there will be technological breakthroughs that may change the relevance of accuracy of some of this research findings.

Appendix 1: How Wikidata is constructed

Wikidata can be read by both humans and machines and as customary for Semantic Web platforms, utilizes information "triples": **Item − > Property − > Value.** *An item* is any topic (a person, place, thing, etc.)**;** *a property* is a specific single kind of data relevant to this item (e.g. the height of a mountains; a capital of a country; the gender of a human); *a value* is either a reference to another item (capital of Germany: Berlin) or a literal value (e.g. 8848 m). The system of triples can be read by both humans and machines since each item has a unique identifying number starting with the letter **Q**, and each property has a unique identifying number starting with the letter **P**.

Every item in Wikidata is first described by **labels** in a variety of languages, a short **description** and **aliases**. Then follows a series of **statements** describing the items, each asserting a single datum or fact about the item—using a property and its value. Each statement could be expanded on, annotated, or contextualized beyond what can be expressed using simple property-value pairs by using **qualifiers.** Qualifiers are also built by pairing a property with a value, to give details on a specific statement. In addition, to ensure statements' verifiability and accuracy each statement could be backed up by **reference**s. Figure 15 details these components in the Wikidata item for Douglas Adams.

Fig. 15 Components of a Wikidata item via the example of item Q42, Douglas Adams

Considering that one of the characteristics of linked open data is interconnectedness of datasets (Bizer et al., 2009), items in Wikidata can also have a special set of statements called **identifiers**—statements that store reference numbers for the item in external databases, allowing ease of use and flexibility in finding details in various datasets. Figure 16 illustrates how identifiers are stated in the Marie Curie Wikidata item.

Fig. 16 Identifiers for the Marie Curie Wikidata item

Fig. 17 The lexeme for the word "חי" in Hebrew

Lately, a new type of entity was introduced into the Wikidata: Lexemes, which are marked with the letter **L**. Each Lexeme has statements that describe the *senses* of a given word, as well as its different *form*s. Lexemes will be able to provide accurate and meaningful online translations that both humans and machines can understand. Figure 17 illustrates the lexeme for the word "חי" in Hebrew, which could have 2 meanings: alive and "live", i.e. in real-time.

Appendix 2: Existing tools for data extraction & visualization

Histropedia, is a third-party application that strives to "bring history to life" and allows users to "see all of history on an interactive timeline" (http://histropedia.com/). Users can explore different timelines, merge timelines, create new ones, and explore items in Wikipedia and other sources. Figure 18 displays a timeline of Italian painters.
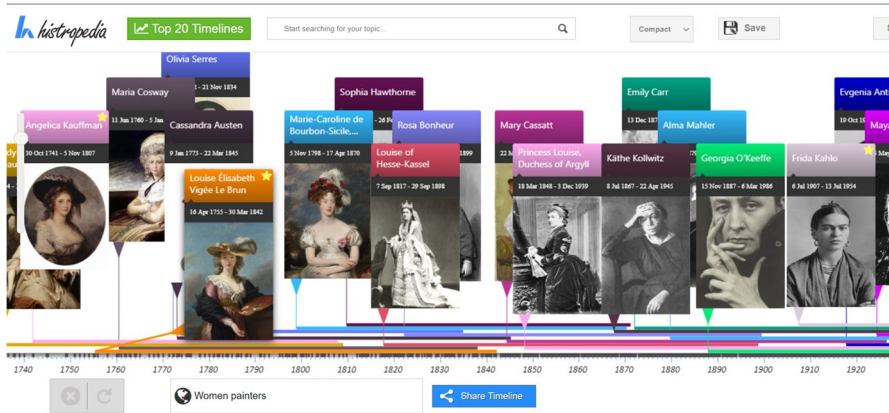
Fig. 18 Histropedia's interactive timeline of women painters

Listeria allows users to curate tabular lists on specific topics. Once the customized table is generated by a Wikipedia query, it gets automatically updated whenever new information is added to Wikipedia—thus assisting in monitoring topics' completeness. In the example below columns showcase existing or missing data—making it easier for contributors to add missing pieces to the "Women-in-Red" WikiProject, a part of a more extensive effort to close the Gender Gap in Wikipedia (Ford & Wajcman, 2017; Hargittai & Shaw, 2015; Konieczny et al., 2018; Wagner et al., 2015). In that sense, Listeria helps generate better transparency on coverage of specific topics, and highlights missing information. In the example below, the list shows articles that do not exist on English Wikipedia, but do have a Wikidata item with some details in it. The different columns showcase basic data that either exists or not (picture, date of birth, place of birth, etc.), making it easier for contributors to add the missing data and then write the article (Fig. 19).

Fig. 19 A Listeria list for women scientists without a Wikipedia article, Women-in-Red project

Scholia (https://tools.wmflabs.org/scholia/), allows users to explore information on research papers, researchers and institutions (Nielsen et al., 2017). While many academics use search engines like Google Scholar, Scholia automatically generates information in a visualized, interactive way, making is superior to non-semantic browsing. In Fig. 20, Sir Tim Berners-Lee's academic contributions are explored, through a series of Wikidata queries, including a list of publications, publications per year, and co-author graph. The query results get updated when new information is added (Figs. 21 and 22).

Fig. 20 Scholia results for Tim Berners-Lee: https://tools.wmflabs.org/scholia/author/Q80
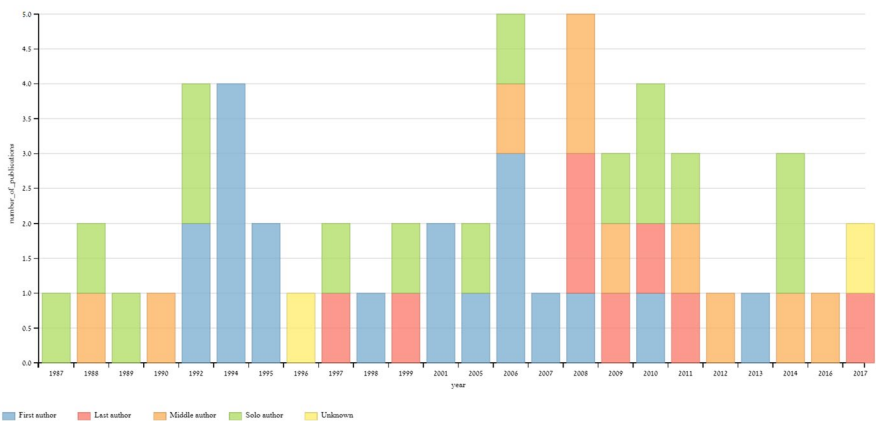
Fig. 21 Scholia results for Tim Berners-Lee
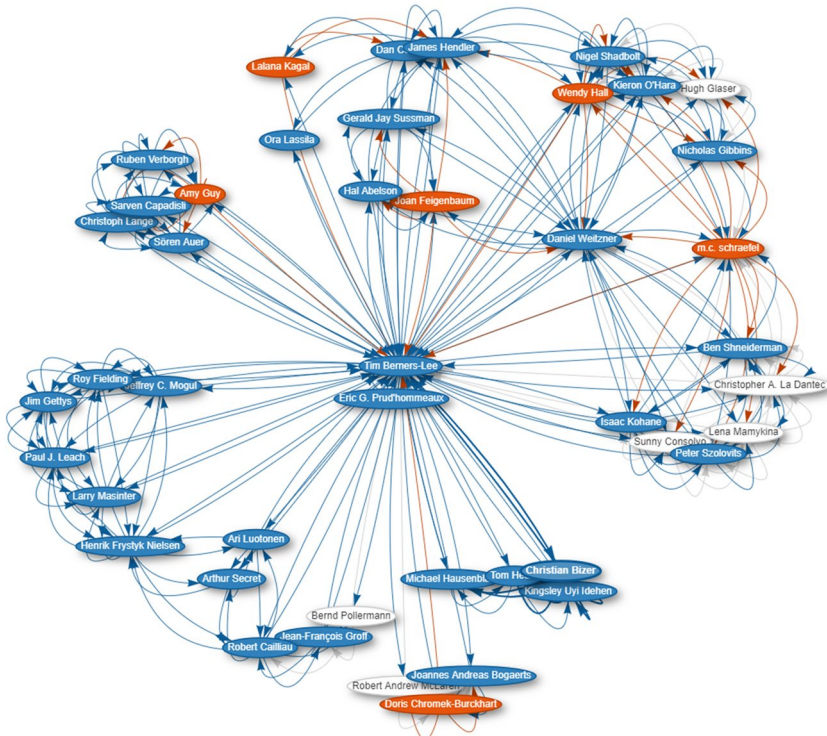
Fig. 22 Scholia results for Tim Berners-Lee

Data availability The datasets generated during and/or analyzed during the current study are not publicly available due, as it is part of a larger-scale research project that has not been published yet, but are available from the corresponding author on reasonable request.

Declarations

Conflict of interest / disclaimer In her capacity as a volunteer, the first author of this paper is also an active contributor to Open Knowledge initiatives, including Wikimedia projects, and serves as the Vice Chair of the Board of Trustees of the Wikimedia Foundation. The research described in this paper was conducted in her capacity as a PhD candidate, and there have been no financial or ethical conflict of interests related to the subject matter of the article.

References

Aibar, E., Lladós-Masllorens, J., Meseguer-Artola, A., Minguillón, J., & Lerga, M. (2015). Wikipedia at university: What faculty think and do about it. *Electronic Library, 33*(4), 668–683. https://doi.org/10.1108/EL-12-2013-0217

Aibar, E., Lerga, M., Llados, J., Meseguer, A., & Minguillon, J. (2013). Wikipedia in higher education: An empirical study on faculty perceptions and practices. In *EDULEARN13 Proceedings: 5th International Conference on Education and New Learning Technologies*, 4269–4275.

Allwardt, D. E. (2011). Writing with wikis: A cautionary tale of technology in the classroom. In *Journal of Social Work Education* (Vol. 47, Issue 3, pp. 597–605). https://doi.org/10.5175/JSWE.2011.200900126

Amaral, G., Kaffee, L.-A., Rodrigues, O., Simperl, E., & Piscopo, A. (2021). Assessing the quality of sources in Wikidata across languages: A hybrid approach. *ACM Journal of Data and Information Quality, 13*(23), 1–35. https://doi.org/10.1145/3484828

Anderson, T., & Whitelock, D. M. (2004). The educational semantic web: Visioning and practicing the future of education. *Journal of Interactive Media in Education*, *2004*(1). https://doi.org/10.5334/2004-1

Anderson, T. (2016). Emergence and innovation in digital learning: Foundations and applications. In *Emergence and Innovation in Digital Learning: Foundations and Applications*. https://doi.org/10.15215/aupress/9781771991490.01

Arnaout, H., Razniewski, S., Weikum, G., & Pan, J. Z. (2021). Negative knowledge for Open-world Wikidata. *The Web Conference 2021 - Companion of the World Wide Web Conference, WWW, 2021*, 544–551. https://doi.org/10.1145/3442442.3452339

Aspin, D. N., & Chapman, J. D. (2010). Lifelong learning: concepts and conceptions. *19*(1), 2–19. https://doi.org/10.1080/026013700293421

Bayliss, G. (2013). Exploring the cautionary attitude toward Wikipedia in Higher Education: Implications for higher education institutions. *New Review of Academic Librarianship, 19*(1), 36–57. https://doi.org/10.1080/13614533.2012.740439

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American, 284*(5), 1–5.

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data - The story so far. *International Journal on Semantic Web and Information Systems, 5*(3), 1–22. https://doi.org/10.4018/jswis.2009081901

Blaschke, L. M. (2021). The dynamic mix of heutagogy and technology: Preparing learners for lifelong learning. *British Journal of Educational Technology*, *52*(4). https://doi.org/10.1111/bjet.13105

Bordel, B., & Mareca, P. (2019). New teaching and learning methodologies in the smart higher education era, a study case, Wikipedia. *International Journal of Technology and Human Interaction, 15*(2), 70–83. https://doi.org/10.4018/IJTHI.2019040106

Boulos, M. N., Maramba, I., & Wheeler, S. (2006). Wikis, blogs and podcasts: A new generation of Web-based tools for virtual collaborative clinical practice and education. *BMC Medical Education, 6*, 41. https://doi.org/10.1186/1472-6920-6-41

Burgstaller-Muehlbacher, S., Waagmeester, A., Mitraka, E., Turner, J., Putman, T., Leong, J., Naik, C., Pavlidis, P., Schriml, L., Good, B. M., & Su, A. I. (2016). Wikidata as a semantic framework for the gene Wiki initiative. *Database, 2016*. https://doi.org/10.1093/database/baw015

Chan, B. S., Churchill, D., & Chiu, T. K. (2017). Digital literacy learning in higher education through digital storytelling approach. *Journal of International Education Research (JIER), 13*(1), 1–16. https://doi.org/10.19030/jier.v13i1.9907

Chao, J. (2007). Student project collaboration using wikis. *Of the 20th Conference on Software Engineering Education and Training*. https://doi.org/10.1109/CSEET.2007.49

Chen, H. L., Cannon, D., Gabrio, J., Leifer, L., Toye, G., & Bailey, T. (2005). Using wikis and weblogs to support reflective learning in an introductory engineering design course. *American Society for Engineering Education Annual Conference & Exposition*. https://doi.org/10.18260/1-2-14895

Colla, D., Goy, A., Leontino, M., & Magro, D. (2021). Wikidata support in the creation of rich semantic metadata for historical archives. *Applied Sciences (Switzerland)*, *11*(10). https://doi.org/10.3390/app11104378

Collins, J. (2004). *Education Techniques for Lifelong Learning Principles of Adult Learning*. https://doi.org/10.1148/rg.245045020

Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. SAGE Publications.

Cronin, C., & MacLaren, I. (2018). Conceptualising OEP: A review of theoretical and empirical literature in Open Educational Practices. *Open Praxis*, *10*(2). https://doi.org/10.5944/openpraxis.10.2.825

Dabbagh, N., & Castaneda, L. (2020). The PLE as a framework for developing agency in lifelong learning. *Educational Technology Research and Development, 68*(6), 3041–3055. https://doi.org/10.1007/s11423-020-09831-z

Di Lauro, F., & Johinke, R. (2017). Employing Wikipedia for good not evil: innovative approaches to collaborative writing assessment. In *Assessment and Evaluation in Higher Education* (Vol. 42, Issue 3, pp. 478–491). https://doi.org/10.1080/02602938.2015.1127322

Dooley, P. L. (2010). Wikipedia and the two-faced professoriate. *Proceedings of WikiSym 2010 - The 6th International Symposium on Wikis and Open Collaboration*. https://doi.org/10.1145/1832772.1832803

Elgort, I., Smith, A. G., & Toland, J. (2008). Is wiki an effective platform for group course work? *Australasian Journal of Educational Technology*, *24*(2). https://doi.org/10.14742/ajet.1222

Erxleben, F., Günther, M., Krötzsch, M., Mendez, J., & Vrandečić, D. (2014). Introducing wikidata to the linked data web. *In International Semantic Web Conference*, 50–65. https://doi.org/10.1007/978-3-319-11964-9

Eteokleous, N., Ktoridou, D., & Orphanou, M. (2014). Integrating Wikis as educational tools for the development of a community of inquiry. *American Journal of Distance Education, 28*(2), 103–116. https://doi.org/10.1080/08923647.2014.896572

Evans, P. (2006). The Wiki Factor. *BizEd*, *5(2)*, 28–32. Accessed 1 Dec 2022

Evenstein Sigalov, S., & Nachmias, R. (2017). Wikipedia as a platform for impactful learning: A new course model in higher education. *Education and Information Technologies, 22*(6), 2959–2979. https://doi.org/10.1007/s10639-016-9564-z

Färber, M., Ell, B., Menne, C., & Rettinger, A. (2015). A comparative survey of dbpedia, freebase, opencyc, wikidata, and yago. *Semantic Web Journal, 1*(1), 1–5. http://semantic-web-journal.org/system/files/swj1141.pdf

Farda-Sarbas, M., & Müller-Birn, C. (2019). *Wikidata from a Research Perspective -- A Systematic Mapping Study of Wikidata*. https://arxiv.org/abs/1908.11153v2

Ferradji, M. A., & Benchikha, F. (2021). Enhanced metrics for temporal dimensions toward assessing Linked Data: A case study of Wikidata. *Journal of King Saud University - Computer and Information Sciences*. https://doi.org/10.1016/J.JKSUCI.2021.05.010

Field, J. (2000). *Lifelong learning and the new educational order*. Trentham Books. Accessed 1 Dec 2022

Ford, H., & Wajcman, J. (2017). 'Anyone can edit', not everyone does: Wikipedia's infrastructure and the gender gap. *Social Studies of Science, 47*(4), 511–527. https://doi.org/10.1177/0306312717692172

Franklin, T., & Harmelen, M. Van. (2007). *Web 2.0 for content for learning and teaching in higher education*. https://staff.blog.ui.ac.id/harrybs/files/2008/10/web-2-for-content-for-learning-and-teaching-in-higher-education.pdf

Freedman, G., & Reynolds, E. G. (1980). Enriching basal reader lessons with semantic webbing. *The Reading Teacher, 33*(6), 677–684. https://www.jstor.org/stable/20195100

Galway, L. P., Corbett, K. K., Takaro, T. K., Tairyan, K., & Frank, E. (2014). A novel integration of online and flipped classroom instructional models in public health higher education. *BMC Medical Education, 14*(1), 181. https://doi.org/10.1186/1472-6920-14-181

Gilster, P. (1997). Digital literacy. New York: John Wiley.

Good, B. M., Burgstaller-Muehlbacher, S., Putman, T., Su, A., Waagmeester, A., & Mitraka, E. (2016). Opportunities and challenges presented by Wikidata in the context of biocuration. *CEUR Workshop Proceedings*, *1747*. https://ceur-ws.org/Vol-1747/BT105_ICBO2016.pdf

Gruber, T. (2008). Collective knowledge systems: Where the Social Web meets the Semantic Web. *Web Semantics: Science, Services and Agents on the World Wide Web, 6*(1), 4–13. https://doi.org/10.1016/j.websem.2007.11.011

Gummer, E. S., & Mandinach, E. B. (2015). Building a conceptual framework for data literacy. *Teachers College Record*, *117*(4). https://doi.org/10.1177/016146811511700401

Guns, R. (2013). Tracing the origins of the semantic web. *Journal of the American Society for Information Science and Technology, 64*(10), 2173–2181. https://doi.org/10.1002/asi.22907

Hadad, S., Shamir-Inbal, T., Blau, I., & Leykin, E. (2021). Professional Development of Code and Robotics Teachers Through Small Private Online Course (SPOC): Teacher centrality and pedagogical

strategies for developing computational thinking of students. *Journal of Educational Computing Research, 59*(4), 763–791. https://doi.org/10.1177/0735633120973432

Hargittai, E., & Shaw, A. (2015). Mind the skills gap: The role of Internet know-how and gender in differentiated contributions to Wikipedia. *Information Communication and Society, 18*(4), 424–442. https://doi.org/10.1080/1369118X.2014.957711

Hase, S., & Blaschke, L. M. (2021). Heutagogy, Work and Lifelong Learning. In *The SAGE Handbook of Learning and Work*. https://doi.org/10.4135/9781529757217.n6

Hase, S., & Kenyon, C. (2000). From andragogy to heutagogy. In *ultiBASE* (Vol. 28). https://www.webarchive.nla.gov.au/awa/20010220130000/http://ultibase.rmit.edu.au/Articles/dec00/hase2.htm

Heftberger, A., Höper, J., Müller-Birn, C., & Walkowski, N.-O. (2020). Opening up research data in film studies by using the structured knowledge base Wikidata. In *Digital Cultural Heritage* (pp. 401–410). https://doi.org/10.1007/978-3-030-15200-0_27

Hegarty, B. (2015). Attributes of open pedagogy: A model for using open educational resources. *Educational Technology, 55*(4), 3–13. https://tinyurl.com/y2tpd3ho.

Hendler, J. (2009). Web 3.0 emerging. *Computer, 42*(1), 111–113. https://doi.org/10.1109/MC.2009.30

Herbert, V. G., Frings, A., Rehatschek, H., Richard, G., & Leithner, A. (2015). Wikipedia - Challenges and new horizons in enhancing medical education. *BMC Medical Education, 15*(1), 32. https://doi.org/10.1186/s12909-015-0309-2

Hernández, D., Hogan, A., & Krötzsch, M. (2015). Reifying RDF: What works well with wikidata? *CEUR Workshop Proceedings, 1457*, 32–47.

Hernández, D., Hogan, A., Riveros, C., Rojas, C., & Zerega, E. (2016). Querying Wikidata: Comparing SPARQL, relational and graph databases. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9982 LNCS*, 88–103. https://doi.org/10.1007/978-3-319-46547-0_10

Higgs, B., & McCarthy, M. (2005). Active learning — from lecture theatre to field-work. In *Emerging Issues in the Practice of University Learning and Teaching* (pp. 37–44). https://www.edin.ie/wp-content/uploads/2021/11/HiggsMcCarthy.pdf

Janio, J. P. (2014). Observations on Wikipedia and its uses in higher education. In Czerepaniak-Walczak & Perzycka (Eds.), *Media and Trust:Theoretical, research and practical contexts* (pp. 297–303). https://www.u-pad.unimc.it/bitstream/11393/201898/1/Trust%20As%20A%20Systemic%20Problem%20-%20Media%20and%20Trust%20(libro%20finale).pdf#page=298

Jha, R. K., Ganguly, S., & Mishra, S. (2019). Alignment of OER platforms with SDGs: An exploratory study. In *Handbook of Research on Emerging Trends and Technologies in Library and Information Science*. https://doi.org/10.4018/978-1-5225-9825-1.ch006

Johnson, R. T., & Johnson, D. W. (1986). Action research: Cooperative learning in the science classroom. *Science and Children, 24*(2), 31–32.

Kaffee, L.-A., Piscopo, A., Vougiouklis, P., Simperl, E., Carr, L., & Pintscher, L. (2017). *A Glimpse into Babel: An Analysis of Multilinguality in Wikidata*. https://doi.org/10.1145/3125433.3125465

Kaffee, L. A. (2016). *Generating article placeholders from Wikidata for Wikipedia: increasing access to free and open knowledge* (Doctoral dissertation, Hochschule für Technik und Wirtschaft Berlin). https://www.upload.wikimedia.org/wikipedia/commons/9/99/Generating_Article_Placeholders_from_Wikidata_for_Wikipedia_-_Increasing_Access_to_Free_and_Open_Knowledge.pdf

Kapsalis, E. (2019). Wikidata: Recruiting the crowd to power access to digital archives. *Journal of Radio and Audio Media, 26*(1), 134–142. https://doi.org/10.1080/19376529.2019.1559520

Klein, M., Gupta, H., Rai, V., Konieczny, P., & Zhu, H. (2016). Monitoring the gender gap with Wikidata human gender indicators. *Proceedings of the 12th International Symposium on Open Collaboration, OpenSym 2016*, 1–9. https://doi.org/10.1145/2957792.2957798

Klein, M., & Kyrios, A. (2013). VIAFbot and the Integration of Library Data on Wikipedia. *Code4Lib Journal, 22*, 85–107.

Koltay, T. (2015). Data literacy: In search of a name and identity. *Journal of Documentation, 71*(2), 401–415. https://doi.org/10.1108/JD-02-2014-0026

Konieczny, P. (2007). Wikis and Wikipedia as a teaching tool. *International Journal of Instructional Technology and Distance Learning, 4*(1), 15–34.

Konieczny, P. (2014). Rethinking Wikipedia for the classroom. *Contexts, 13*(1), 80–83. https://doi.org/10.1177/1536504214522017

Konieczny, P. (2016). Teaching with Wikipedia in a 21st-century classroom: Perceptions of Wikipedia and its educational benefits. *Journal of the Association for Information Science and Technology, 67*(7), 1523–1534. https://doi.org/10.1002/asi.23616

Konieczny, P., & Klein, M. (2018). Gender gap through time and space: A journey through Wikipedia biographies via the Wikidata Human Gender Indicator. *New Media and Society, 20*(12), 4608–4633. https://doi.org/10.1177/1461444818779080

Kop, R., & Hill, A. (2008). Connectivism: Learning theory of the future or vestige of the past? In *International Review of Research in Open and Distance Learning* (Vol. 9, Issue 3). https://doi.org/10.19173/irrodl.v9i3.523

Koper, R. (2004). Use of the semantic web to solve some basic problems in education: Increase flexible, distributed lifelong learning; decrease teacher's workload. *Journal of Interactive Media in Education, 2004*(1). https://doi.org/10.5334/2004-6-koper

Krötzsch, M., Vrandečić, D., Völkel, M., Haller, H., & Studer, R. (2007). Semantic Wikipedia. *Web Semantics, 5*(4), 251–261. https://doi.org/10.1016/j.websem.2007.09.001

Kummer, C. (2013). Factors influencing Wiki collaboration in higher education. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2208522

Laal, M. (2012). Benefits of lifelong learning. *Procedia - Social and Behavioral Sciences, 46*, 4268–4272. https://doi.org/10.1016/J.SBSPRO.2012.06.239

Laal, M., & Salamati, P. (2012). Lifelong learning; why do we need it? *Procedia - Social and Behavioral Sciences, 31*, 399–403. https://doi.org/10.1016/J.SBSPRO.2011.12.073

LaFrance, J., & Calhoun, D. W. (2012). Student perceptions of Wikipedia as a learning tool for educational leaders. *International Journal of Educational Leadership Preparation, 7*(2), 2.

Lane, A. (2017). Open education and the sustainable development goals: Making change happen. *Journal of Learning for Development, 4*(3), 275–286. Retrieved February 19, 2023 from https://www.learntechlib.org/p/189219/

Lemus-Rojas, M., & Odell, J. D. (2018). Creating structured linked data to generate scholarly profiles: A pilot project using wikidata and Scholia. *Journal of Librarianship and Scholarly Communication, 6*(1). https://doi.org/10.7710/2162-3309.2272

Lemus-Rojas, M., & Lee, Y. Y. (2019). Using wikidata to provide visibility to women in STEM. *Proceedings of the International Conference on Dublin Core and Metadata Applications*. https://doi.org/10.1016/j.acalib.2021.102326

Lemus-Rojas, M., & Ramirez Rojas, M. (2021). Wikidata Projects in Times of COVID-19: IUPUI Libraries' Engagement in Open Knowledge. *InULA Notes, 33*(1). Retrieved June 4, 2022. https://www.scholarworks.iupui.edu/bitstream/handle/1805/26112/Wikidata%20Projects%20in%20Times%20of%20COVID-19.pdf?sequence=1

Li, M., Reddy, R. G., Wang, Z., Chiang, Y. S., Lai, T., Yu, P., & Ji, H. (2022). Covid-19 claim radar: A structured claim extraction and tracking system. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 135–144)

Lin, H. (2019). Teaching and learning without a textbook: Undergraduate student perceptions of open educational resources. *International Review of Research in Open and Distance Learning, 20*(3). https://doi.org/10.19173/irrodl.v20i4.4224

Malyshev, S., Krötzsch, M., González, L., Gonsior, J., & Bielefeldt, A. (2018). Getting the most out of wikidata: Semantic technology usage in wikipedia's knowledge graph. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11137 LNCS*, 376–394. https://doi.org/10.1007/978-3-030-00668-6_23

Mandinach, E. B., Parton, B. M., Gummer, E. S., & Anderson, R. (2015). Ethical and appropriate data use requires data literacy. *Phi Delta Kappan, 96*(5). https://doi.org/10.1177/0031721715569465

Mandinach, E. B., & Gummer, E. S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher, 42*(1), 30–37. https://doi.org/10.3102/0013189X12459803

Mareca, M. P., & Bordel, B. (2019). The educative model is changing: toward a student participative learning framework 3.0 - editing Wikipedia in the higher education. *Universal Access in the Information Society, 18*(3), 689–701. https://doi.org/10.1007/s10209-019-00687-6

McKenzie, B., Brown, J., Casey, D., Cooney, A., Darcy, E., Giblin, S., & NíMhórdha, M. (2018). From poetry to Palmerstown: using Wikipedia to teach critical skills and information literacy in a first-year seminar. *College Teaching, 66*(3), 140–147. https://doi.org/10.1080/87567555.2018.1463504

Meier, F. (2022). *TWikiL-The Twitter Wikipedia Link Dataset*. https://doi.org/10.5281/zenodo.5845374

Mendes, T. B., Dawson, J., Evenstein Sigalov, S., Kleiman, N., Hird, K., Terenius, O., Das, D., Geres, N., & Azzam, A. (2021). Wikipedia in health professional schools: From an opponent to an Ally. *Medical Science Educator, 31*(6). https://doi.org/10.1007/s40670-021-01408-6

Metilli, D., Bartalesi, V., & Meghini, C. (2019). A Wikidata-based tool for building and visualising narratives. *International Journal on Digital Libraries, 20*(4), 417–432. https://doi.org/10.1007/s00799-019-00266-3

Mietchen, D., Hagedorn, G., Willighagen, E., Rico, M., Gómez-Pérez, A., Aibar, E., Rafes, K., Germain, C., Dunning, A., Pintscher, L., & Kinzler, D. (2015). Enabling Open Science: Wikidata for Research (Wiki4R). *Research Ideas and Outcomes*, *1*. https://doi.org/10.3897/rio.1.e7573

Minguillón, J., Aibar, E., Lerga, M., Lladós, J., & Meseguer-Artola, A. (2018). Wikipedia in academia as a teaching tool: From averse to proactive faculty profiles. arXiv preprint arXiv:1801.07138. https://www.arxiv.org/abs/1801.07138

Moore, R. L. (2020). Developing lifelong learning with heutagogy: contexts, critiques, and challenges. *Distance Education*, *41*(3). https://doi.org/10.1080/01587919.2020.1766949

Mora-Cantallops, M., Sánchez-Alonso, S., & García-Barriocanal, E. (2019). A systematic literature review on Wikidata. In *Data Technologies and Applications* (Vol. 53, Issue 3). https://doi.org/10.1108/DTA-12-2018-0110

Morshed, M. (2021). *Modeling syntactic dependency relationships in Wikidata lexicographical data*. In Wikidata@ ISWC. https://www.ceur-ws.org/Vol-2982/paper-7.pdf

Müller-Birn, C., Karran, B., Lehmann, J., & Luczak-Rösch, M. (2015). Peer-production system or collaborative ontology engineering effort: What is Wikidata? *Proceedings of the 11th International Symposium on Open Collaboration, OPENSYM 2015*, 20. https://doi.org/10.1145/2788993.2789836

Naeve, A., Lytras, M., Nejdl, W., Balacheff, N., & Hardin, J. (2006). Editorial - Advances of the Semantic Web for e-learning: Expanding learning frontiers. In *British Journal of Educational Technology* (Vol. 37, Issue 3, pp. 321–330). https://doi.org/10.1111/j.1467-8535.2006.00608.x

Naismith, L., Lee, B. H., & Pilkington, R. M. (2011). Collaborative learning with a wiki: Differences in perceived usefulness in two contexts of use. *Journal of Computer Assisted Learning, 27*(3), 228–242. https://doi.org/10.1111/j.1365-2729.2010.00393.x

Neelam, S., Sharma, U., Karanam, H., Ikbal, S., Kapanipathi, P., Abdelaziz, I., Mihindukulasooriya, N., Lee, Y.-S., Srivastava, S., Pendus, C., Dana, S., Garg, D., Fokoue, A., Bhargav, S., Khandelwal, D., Ravishankar, S., Gurajada, S., Chang, M., Uceda-Sosa, R., Subramaniam, V. (2022). *A benchmark for generalizable and interpretable temporal question answering over knowledge bases*. https://www.arxiv.org/abs/2201.05793

Nielsen, F. Å., Mietchen, D., & Willighagen, E. (2017). Scholia, Scientometrics and Wikidata. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10577 LNCS*. https://doi.org/10.1007/978-3-319-70407-4_36

Nielsen, F. Å. (2016a). *Fundamentals — from data to visualisation Big Data Business Academy*. http://www.2.imm.dtu.dk/pubdb/edoc/imm6950.pdf

Nielsen, F. (2016). *Literature, Geolocation and Wikidata*. In *Proceedings of the International AAAI Conference on Web and Social Media 10*(2), 61–64. https://www.ojs.aaai.org/index.php/ICWSM/article/view/14833/14683

Ossiannilsson, E. (2019). OER and OEP for access, equity, equality, quality, inclusiveness, and empowering lifelong learning. *International Journal of Open Educational Resources*, *1*(2). https://doi.org/10.18278/ijoer.1.2.9

Pangrazio, L., Godhe, A. L., & Ledesma, A. G. L. (2020). What is digital literacy? A comparative review of publications across three language contexts. *17*(6), 442–459. https://doi.org/10.1177/2042753020946291

Parker, K., & Chao, J. (2007). Wiki as a teaching tool. *Interdisciplinary Journal of E-Learning and Learning Objects, 3*(1), 57–72. https://doi.org/10.28945/386

Pfundner, A., Schönberg, T., Horn, J., Boyce, R. D., & Samwald, M. (2015). Utilizing the Wikidata system to improve the quality of medical content in Wikipedia in diverse Languages:a pilot study. *Journal of Medical Internet Research*, *17*(5). https://doi.org/10.2196/jmir.4163

Putman, T. E., Lelong, S., Burgstaller-Muehlbacher, S., Waagmeester, A., DIesh, C., Dunn, N., Munoz-Torres, M., Stupp, G. S., Wu, C., Su, A. I., & Good, B. M. (2017). WikiGenomes: An open web application for community consumption and curation of gene annotation data in Wikidata. *Database*, *2017*(1). https://doi.org/10.1093/database/bax025

Qin, J., & D'Ignazio, J. (2010). Lessons learned from a two-year experience in science data literacy education. *International Association of Scientific and Technological University Libraries, 31st Annual Conference*.

Ramanau, R., & Geng, F. (2009). Researching the use of Wiki's to facilitate group work. *Procedia - Social and Behavioral Sciences, 1*(1), 2620–2626. https://doi.org/10.1016/j.sbspro.2009.01.463

Rasberry, L., & Mietchen, D. (2021). FAIR and open multilingual clinical trials in Wikidata and Wikipedia. *Research Ideas and Outcomes, 7*. https://doi.org/10.3897/rio.7.e66490

Reddy, P., Sharma, B., & Chaudhary, K. (2020). Digital literacy: A review of literature. *International Journal of Technoethics (IJT), 11*(2), 65–94. https://doi.org/10.4018/IJT.20200701.oa1

Rutz, A., Sorokina, M., Galgonek, J., Mietchen, D., Willighagen, E., Graham, J., & Stephan, R. (2021). *Open natural products research: curation and dissemination of biological occurrences of chemical structures through Wikidata*. https://doi.org/10.1101/2021.02.28.43326

Schaffert, S., Bischof, D., Bürger, T., Gruber, A., Hilzensauer, W., & Schaffert, S. (2006). Learning with semantic wikis. *CEUR Workshop Proceedings, 206*, 109–123.

Scharpf, P., Schubotz, M., & Gipp, B. (2021a). Fast linking of mathematical Wikidata entities in Wikipedia articles using annotation recommendation; fast linking of mathematical Wikidata entities in wikipedia articles using annotation recommendation. *Companion Proceedings of the Web Conference 2021*. https://doi.org/10.1145/3442442

Scharpf, P., Schubotz, M., & Gipp, B. (2021). Mathematics in wikidata. In *2nd Wikidata Workshop (Wikidata 2021) co-located with the 20th International Semantic Web Conference (ISWC 2021), Virtual Conference, October 24, 2021* (Vol. 2982, p. 1). Aachen, Germany: RWTH Aachen. https://www.oa.tib.eu/renate/bitstream/123456789/8963/1/paper-1-1-Mathematics.pdf

Shields, M. (2005). Information literacy, statistical literacy, data literacy. *IASSIST quarterly, 28*(2-3), 6–6.

Seitzinger, J. (2006). Be constructive: Blogs, podcasts, and wikis as constructivist learning tools. *Learning solutions e-magazine, 31*, 1–12. https://www.innovationlabs.com/newhighschool/2006/reading%20materials/constructivist.pdf

Selwyn, N., & Gorard, S. (2016). Students' use of Wikipedia as an academic resource - Patterns of use and perceptions of usefulness. *Internet and Higher Education, 28*, 28–34. https://doi.org/10.1016/j.iheduc.2015.08.004

Shadbolt, N., Hall, W., & Berners-Lee, T. (2006). The semantic web revisited. In *IEEE Intelligent Systems* (Vol. 21, Issue 3). https://doi.org/10.1109/MIS.2006.62

Sharpe, D. (2015). Chi-square test is statistically significant: Now what? *Practical Assessment, Research, and Evaluation, 20*(1), 8. https://doi.org/10.7275/tbfa-x148

Shenoy, K., Ilievski, F., Garijo, D., Schwabe, D., & Szekely, P. (2022). A study of the quality of Wikidata. *Journal of Web Semantics, 72*, 100679. https://doi.org/10.1016/J.WEBSEM.2021.100679

Snyder, E., Lorenzo, L., & Mak, L. (2019). Linked open data for subject discovery: Assessing the alignment between library of congress vocabularies and Wikidata. In *International Conference on Dublin Core and Metadata Applications* (pp. 12–20). https://www.dcpapers.dublincore.org/pubs/article/view/4225

Soler-Adillon, J., Pavlovic, D., & Freixa, P. (2018). Wikipedia in higher education: Changes in perceived value through content contribution. *Comunicar, 26*(54). https://doi.org/10.3916/C54-2018-04

Spante, M., Hashemi, S. S., Lundin, M., & Algers, A. (2018). Digital competence and digital literacy in higher education research: Systematic review of concept use. *Cogent Education, 5*(1), 1–21. https://doi.org/10.1080/2331186X.2018.1519143/SUPPL_FILE/OAED_A_1519143_SM9855.DOCX

Staub, T., & Hodel, T. (2016). Wikipedia vs. Academia: An investigation into the role of the internet in education, with a special focus on Wikipedia. *Universal Journal of Educational Research, 4*(2), 349–354. https://doi.org/10.13189/ujer.2016.040205

Steiner, T. (2014). Bots vs. Wikipedians, Anons vs. Logged-ins (Redux): A global study of edit activity on Wikipedia and Wikidata. *Proceedings of the 10th International Symposium on Open Collaboration, OpenSym 2014*, 25. https://doi.org/10.1145/2641580.2641613

Stephenson, E., & Caravello, P. S. (2007). Incorporating data literacy into undergraduate information literacy programs in the social sciences A pilot project. *Reference Services Review, 35*(4). https://doi.org/10.1108/00907320710838354

Tang, C. M., & Chaw, L. Y. (2016). Digital literacy: A prerequisite for effective learning in a blended learning environment? *Electronic Journal of e-Learning, 14*(1), 54–65.

Taveekarn, W., Yimudom, C., Sukkanta, S., Lynden, S., Sawangphol, W., & Tuarob, S. (2019). Data++: An automated tool for intelligent data augmentation using wikidata. In *2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE)* (pp. 91–96). IEEE.

Tharani, K. (2021). Much more than a mere technology: A systematic review of Wikidata in libraries. *Journal of Academic Librarianship, 47*(2). https://doi.org/10.1016/j.acalib.2021.102326

Tlili, A., Nascimbeni, F., Burgos, D., Zhang, X., Huang, R., & Chang, T. W. (2020). The evolution of sustainability models for Open Educational Resources: Insights from the literature and experts. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2020.1839507

Turki, H., Shafee, T., Hadj Taieb, M. A., Ben Aouicha, M., Vrandečić, D., Das, D., & Hamdi, H. (2019). Wikidata: A large-scale collaborative ontological medical database. In *Journal of Biomedical Informatics* (Vol. 99). https://doi.org/10.1016/j.jbi.2019.103292

Turki, H., Ali Hadj Taieb, M., Shafee, T., Lubiana, T., Jemielniak, D., Ben Aouicha, M., Emilio Labra Gayo, J., Youngstrom, E. A., Banat, ab, Das, D., Mietchen, D., behalf of WikiProject COVID-, on, Ainali, J., Änäs, S., Azzellini, E., Boccone, A., Darnell, J., Denis, L., Farmbrough, R., Willighagen, E. (2022a). Representing COVID-19 information in collaborative knowledge graphs: The case of Wikidata. *Semantic Web*, *13*(2), 233–264. https://doi.org/10.3233/SW-210444

Turki, H., Priskorn, D., Ali Hadj Taieb, M., Ben Aouicha, M., & Piad-Morffis, A. (2022b). *Enhancing multilingual and biomedical named entity recognition using Wikidata semantic relations ACM Reference Format*. https://pypi.org/project/langdetect/

UNESCO. (2002). *Forum on the Impact of Open Courseware for Higher Education in Developing Countries, UNESCO, Paris, 1–3 July 2002: final report*. https://unesdoc.unesco.org/notice?id=p::usmarcdef_0000128515

Urbančič, T., Polajnar, A., & Jermol, M. (2019). Open education for a better world: A mentoring programme fostering design and reuse of open educational resources for sustainable development goals. *Open Praxis*, *11*(4). https://doi.org/10.5944/openpraxis.11.4.1026

Vetter, M. A., McDowell, Z. J., & Stewart, M. (2019). From opportunities to outcomes: The Wikipedia-based writing assignment. *Computers and Composition, 52*, 53–64. https://doi.org/10.1016/j.compcom.2019.01.008

Vrandečić, D., & Krötzsch, M. (2014). Wikidata: A free collaborative knowledgebase. *Communications of the ACM*. https://doi.org/10.1145/2629489

Waagmeester, A., Willighagen, E. L., Su, A. I., Kutmon, M., Gayo, J. E. L., Fernández-Álvarez, D., Groom, Q., Schaap, P. J., Verhagen, L. M., & Koehorst, J. J. (2021). A protocol for adding knowledge to Wikidata: Aligning resources on human coronaviruses. *BMC Biology, 19*(1), 1–14. https://doi.org/10.1186/S12915-020-00940-Y/FIGURES/10

Waagmeester, A., Stupp, G., Burgstaller-Muehlbacher, S., Good, B. M., Griffith, M., Griffith, O. L., Hanspers, K., Hermjakob, H., Hudson, T. S., Hybiske, K., Keating, S. M., Manske, M., Mayers, M., Mietchen, D., Mitraka, E., Pico, A. R., Putman, T., Riutta, A., Queralt-Rosinach, N., … Su, A. I. (2020). Wikidata as a knowledge graph for the life sciences. *ELife*, *9*. https://doi.org/10.7554/ELIFE.52614

Wagner, C., Garcia, D., Jadidi, M., & Strohmaier, M. (2015). It's a Man's Wikipedia? Assessing gender inequality in an online encyclopedia. *Proceedings of the 9th International Conference on Web and Social Media, ICWSM 2015*.

Wang, B., Wu, C., & Huang, L. (2019). Data literacy for safety professionals in safety management: A theoretical perspective on basic questions and answers. *Safety Science*, *117*. https://doi.org/10.1016/j.ssci.2019.04.002

Wheeler, S., Kelly, P., & Gale, K. (2005). The influence of online problem-based learning on teachers' professional practice and identity. *Research in Learning Technology*, *13*(2). https://doi.org/10.3402/rlt.v13i2.10986

Wiley, D., & Hilton, J. (2018). Defining OER-enabled pedagogy. *International Review of Research in Open and Distance Learning*, *19*(4). https://doi.org/10.19173/irrodl.v19i4.3601

Zhang, C. C., Houtti, M., Smith, C. E., Kong, R., & Terveen, L. (2022). Working for the invisible machines or pumping information into an empty void? An exploration of Wikidata contributors' motivations. *Proceedings of the ACM on Human-Computer Interaction, 6*(CSCW1), 1–21. https://doi.org/10.1145/3512982

Zheng, B., Niiya, M., & Warschauer, M. (2015). Wikis and collaborative learning in higher education. *Technology, Pedagogy and Education, 24*(3), 357–374. https://doi.org/10.1080/1475939X.2014.948041