Umanistica Digitale – ISSN: 2532–8816 – n. 22, 2026: Special Issue
G. Renda, M. Daquino – *Do Information Visualisation and Semantic Web Get on Well? A Study on the Usage of MELODY to Improve Data Literacy Skills*
DOI: https://doi.org/10.60923/issn.2532–8816/21755

Do Information Visualisation and Semantic Web Get on Well? A Study on the Usage of MELODY to Improve Data Literacy Skills

Giulia Renda

University of Bologna, Bologna, Italy
giulia.renda3@unibo.it

Marilena Daquino

University of Bologna, Bologna, Italy
marilena.daquino2@unibo.it

Abstract

Data literacy and visualisation literacy are pivotal for Humanities scholars to explore and interpret digital Cultural Heritage collections. In the Digital Humanities realm, such skills facilitate engagement with data–driven information systems, making abstract information tangible and actionable. This study investigates the role of a web–based storytelling editor, i.e. MELODY, in enhancing students' ability to interpret and create narratives from Linked Open Data sources using information visualisation and storytelling techniques. By analysing how students integrated visual elements into their data stories and reflecting on their project feedback, we characterise the way essential competencies in data and visualisation literacy are unlocked by the usage of WYSIWYG tools. Findings reveal a general preference for exploratory data analysis approaches, characterised by linear narratives, and numerous, easy–to–consume, visual aids. Critical thinking is fostered by the need of presenting explanatory narratives, and the immediate feedback from the platform supports iterative learning. Challenges faced with incomplete data sources remain a significant frustration element in the learning experience. Insights underscore the need for improved educational tools that effectively support trial–and–error exploratory learning approaches to the Semantic Web.

Keywords: data visualisation literacy, semantic web, education, data storytelling, exploratory data analysis

La data literacy e la visualisation literacy sono fondamentali per i ricercatori delle discipline umanistiche nell'esplorazione e nell'interpretazione delle collezioni digitali di patrimonio culturale. Nell'ambito delle Digital

Humanities, tali competenze facilitano l'interazione con sistemi informativi basati sui dati, rendendo l'informazione astratta tangibile e fruibile. Questo studio indaga il ruolo di un editor per data storytelling, i.e. MELODY,[1] nella crescita di capacità interpretative a partire da Linked Open Data. Analizzando come un gruppo di studenti ha integrato elementi visivi nelle loro storie e riflettendo sui feedback dati al completamento del loro progetto, vogliamo caratterizzare le modalità attraverso cui competenze essenziali come data literacy e data visualisation literacy vengono incentivate tramite l'uso di strumenti WYSIWYG. I risultati rivelano una preferenza generale per approcci di analisi esplorativa dei dati, caratterizzati da narrazioni lineari e numerosi supporti visivi facili da comprendere. Il pensiero critico è stimolato dalla necessità di presentare narrazioni esplicative e dal feedback immediato della piattaforma, che supporta l'apprendimento iterativo. Le difficoltà legate a fonti di dati incomplete restano un elemento di frustrazione significativo nell'esperienza formativa. I risultati evidenziano la necessità di strumenti educativi più efficaci, in grado di supportare approcci di apprendimento esplorativo e basati sul trial–and–error applicati al Web Semantico.

Parole chiave: alfabetizzazione alla visualizzazione dei dati, web semantico, educazione, narrazione dei dati, analisi esplorativa dei dati

1. Introduction

Data and visualisation literacy are critical competencies in today's information society. In the realm of Digital Humanities, scholars and practitioners increasingly engage with complex cultural heritage data collections in order to answer their research questions [23]. Such an approach requires one to manipulate raw data into meaningful insights, and leverage both their analytical and creative capabilities to present the results of their analyses to a broad audience. In this context, mastering Semantic Web technologies – a set of standards that provide formal and structured descriptions of concepts, terms and relationships that enable machines to process web content –is a desired expected skill, since its paramount role in the dissemination of Cultural Heritage data [20].

It has been argued that information visualisation – particularly when applied to Linked Open Data (LOD), i.e. structured and interlinked datasets published in open formats – can ease the learning curve required by Semantic Web technologies and serve as a meaningful educational tool [12]. Prior studies have demonstrated that combining narrativity with information visualisation techniques fosters critical thinking and interpretative skills, which are fundamental for humanists to navigate and engage with complex datasets [11]; [2]. Specifically, integrating storytelling into visual analytics encourages users to critically assess data representation choices, mitigating the risk of oversimplification. This is particularly relevant in the Humanities, where uncertainty, vagueness, and incompleteness are inherent characteristics of knowledge production. By embedding narrative elements into data exploration, students are prompted to reflect on the epistemological limitations of their sources, fostering a more nuanced and responsible engagement with data [19].

However, data analysis, information visualisation, storytelling, and Semantic Web technologies are often taught in separate learning paths. Learning (and teaching) such competencies poses a number of challenges, particularly for learners with limited technical backgrounds [35]. Therefore, many have advocated for easy–to–use tools [22] to bridge different levels of competences and facilitate a comprehensive learning experience. To the best of our knowledge,

[1] https://projects.dharc.unibo.it/melody/

previous studies on data literacy [4]; [26]; [38] lack a focus on teaching Semantic Web technologies together with information visualisation strategies, and offer limited consideration of data visualisation and storytelling tools to support such a learning experience [30].

MELODY is a what–you–see–is–what–you–get (WYSIWYG) platform – meaning that users can design content via graphical user interfaces (GUI), with the final output closely matching what is displayed during editing – for data storytelling design and publication. It integrates curated texts and data visualisations (charts, maps, text searches, and tables) built on top of users' input SPARQL queries –a query language used to retrieve and manipulate LOD – into natural language narratives enriched [30]. Notably, MELODY has been designed specifically to allow the interrogation and integration of Linked Open Data sources into "data stories", i.e. web documents including texts and chart components in a linear, author–driven narrative. MELODY has recently been used as a teaching tool in a Humanities master course program[2] that introduces Semantic Web technologies to undergraduates with no technical background. The objective of the course was to investigate a music–related topic using Linked Open Datasets and produce data stories.

In this paper we examine the collection of students' data stories realised with MELODY. Our goal is to (1) characterise data and visualisation literacy skills developed by students while using the platform, and (2) identify key aspects of how WYSIWYG tools impact on data and visualisation literacy. In particular, we are interested in how students engage with data collections, which data exploratory solutions they adopt and how they communicate them in their stories, as well as the challenges and opportunities that emerge in the learning journey. Findings reveal how intuitive visual storytelling tools can effectively foster data and visualisation literacy by motivating students to explore and communicate complex datasets with simple and easy–to–consume solutions. Still, challenges related to incomplete data sources and technical usability are frustrating elements that significantly affect the perception of learners. The article extends our prior work [29] by including new insights on students' 1) usage of textual aspects, 2) SPARQL proficiency, and 3) perception of usefulness of visual/textual components. Moreover, we include results of the usability study performed during the course to better appreciate how the tool chosen for the study may have affected results.

The article is organised as follows. In Literature Review we discuss previous works on data literacy and the Semantic Web, highlighting a clear gap in adopting information visualisation as a learning tool. In Materials and Methods we present the corpus of data stories produced by students using MELODY, and in Results we present results of our analysis. In Usability Study and Feedback we share the results of the usability test and students' feedback reported as part of the exam. Finally, in Discussion and Conclusion we discuss results, drawing conclusions and future works.

2. Literature Review

It has been argued that information visualisation can enhance data literacy and critical thinking [8]; [12]. Likewise, storytelling has demonstrated being a creative, project–driven approach to cultivate data literacy [22]; [21]. By allowing students to integrate visualisations into data stories, storytelling bridges the gap between data analysis and communication practices that can foster

[2] https://www.unibo.it/it/studiare/dottorati–master–specializzazioni–e–altra–formazione/insegnamenti/insegnamento/2023/402023

ethical awareness, as students reflect on the implications of how they represent and share data. Such an approach makes data meaningful to students, promoting critical engagement and communication skills [36]. Indeed, while experts prefer technically precise visualisations, less experienced users favour contextually rich and engaging formats, where contextual and narrative elements are present. A balance between the two approaches is essential in educational contexts, where students must learn not only to design accurate visualisations but also to understand their narrative potential. Moreover, data visualisations play a dual role: reducing cognitive load and fostering critical engagement with datasets [36]. To this extent, scholars have argued on the importance of integrating visualisation literacy courses into curricula, so as to make it a critical skill for learners [36]; [32]. The integration of competences such as web development, data analysis, and data preparation is increasingly recognised as necessary for mature engagement with complex datasets across various knowledge domains [24]. This multifaceted skill set becomes especially relevant when students are expected to interpret and manipulate Linked Open Data using interactive visualisation and storytelling platforms.

[38] offer a broad view of the data literacy learning landscape, identifying information visualisation in educational paths as a pivotal competency. Exploratory platforms like the Comédie–Française Registers Project (CFRP) exemplify this potential, allowing users to interact with datasets through intuitive interfaces [39]. However, studies suggest that while such tools simplify initial engagement, they often lack the depth to support advanced exploration, underscoring the need for balanced designs that foster both comprehension and creativity [17].

Studies have shown that the combination of narrativity and information visualisation helps expand critical skills essential for working with complex data, encouraging users to question, contextualise, and interpret visual representations rather than passively consuming it [22]; [11]; [3]; [13]; [19]. To the best of our knowledge, MELODY is the only open–source platform that integrates both aspects–narrative and data visualisation–within a single environment, which would require the use of multiple tools [30].

MELODY is a web–based platform designed for data storytelling with Linked Open Data. It allows users to query any SPARQL endpoint, visualise results dynamically as interactive charts, and therefore integrate structured data into curated narratives. The main requirement to create the data story is the URL of a SPARQL endpoint. In the data story canvas, users can select and arrange various UI components–such as charts, maps, tables, and text blocks–to construct their narrative. These data stories go beyond simple data visualisations by structuring visual insights within a narrative context, which facilitates interpretation and communication, qualities that distinguish data storytelling from mere data presentation [31]. Content creation is simplified through WYSIWYG forms, allowing authors to input plain text or SPARQL queries, with immediate preview of results. The final output is a web document containing an ordered list of components, which can be modified and rearranged at any time.

MELODY is specifically tailored for querying and visualising LOD. Created to address the research needs of humanists–linguists, historians, musicologists–and validated through their input [10]; [9], the platform was designed to facilitate the exploration of cultural heritage datasets, which today are predominantly served as linked data [20]. Despite its intuitive interface, learning Semantic Web technologies remains challenging for humanists, as they must acquire SPARQL querying skills and an understanding of linked data structures, which are often beyond their traditional expertise [6]; [17].

Teaching the Semantic Web often presents unique challenges due to its technical nature and steep learning curves, especially for students from non–technical backgrounds. Interdisciplinary approaches to data science education, as highlighted by [35], emphasise the importance of

integrating qualitative and quantitative reasoning to make complex technical subjects more accessible to diverse learners. While not directly referencing Semantic Web technologies, their advocacy for inclusive and interdisciplinary pedagogies underscores the need for teaching methods that bridge these gaps and cater to students from a variety of academic and professional backgrounds. [12] provide evidence of the educational potential of Semantic Web technologies, focusing on Wikidata as a teaching platform. They highlight how data visualisations make LOD accessible, by enabling learners to explore connections, identify gaps, and critically assess data quality. This approach fosters key skills such as data modelling, ontology design, and critical thinking, establishing visualisation as a vital tool in Semantic Web education. However, their study also identifies significant challenges, including high thresholds for newcomers, inconsistent data modelling practices, and usability issues, which underscore the need for more intuitive educational tools.

In summary, scholars agree on the benefits derived from adopting information visualisation and storytelling strategies to foster data literacy skills. However, the integration of such approaches into Semantic Web education remains underexplored. Studies suggest that easy–to–use tools would effectively help to bridge such different competences, fostering critical skills along the learning journey of new technologies. In this study we investigate a unique example of the intertwining of such methods in a master course, and we analyse novices' behaviours to characterise their data literacy skills and whether these are affected by the usage of a dedicated tool.

3. Materials and Methods

MELODY provides users (i.e. students in our case study) with a flexible environment for creating web–based data stories that combine curated text elements (e.g., titles, paragraphs) with interactive data visualisations. The platform supports various visualisation types, including counters, bar charts, line charts, scatterplots, doughnut charts, tables, and maps with interactive filters.

These components answer the need to select the appropriate visualisation strategies based on the author's intent [19]. Given that MELODY places full control of data access in the hands of authors through custom SPARQL queries, the platform supports multiple visualisation types suited to different goals: a single piece of information (counter), parts of a whole (bar chart, doughnut chart), comparisons (bar chart, table), trends (line chart), distributions (bar chart), relationships (scatterplot), and spatial information (map). This variety enables authors to construct semantically meaningful visual expressions tailored to their storytelling objectives.

Additionally, it allows more sophisticated exploratory journeys, such as interactive text searches, i.e. a follow–your–nose approach to explore a dataset moving from a text search to a result set, and from an item of the result set to another result set. Searches return results as tables, where each result is associated with buttons representing bespoke journeys (designed by the author of the data story). By clicking on buttons in cells, final users can dynamically generate new tables answering the research question previously designed by the author of the story, enabling an iterative exploration of a graph. Figure 1 illustrates an example of a chart configuration and a text search.

MELODY allows authors with a GitHub profile to publish their data stories for free in a dedicated online catalogue.[3] Additionally, authors and final users can export their data stories as HTML, PDF, or JSON files, ensuring flexibility in presentation and sharing. The JSON file includes a readable, actionable, structured view of the text/visual components chosen by authors, including their order, the queries, and the parameters to setup the visualisations.

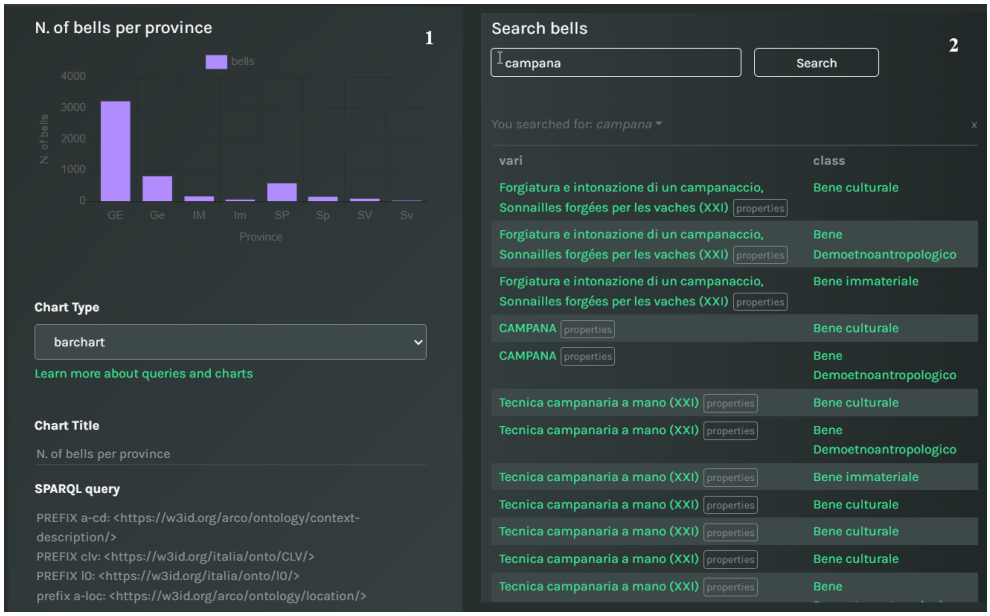

Figure 1 The WYSIWYG interface showing a chart preview while typing the SPARQL query (1) and an example of text search (2).

In this study we analyse the JSON files of a subset of stories collected on the catalogue. In particular, the study is conducted on stories produced by students enrolled in the Language, Society and Communication master's program at the University of Bologna, as part of the course on Semantic Web technologies. Students were briefly introduced to data journalism [15] but were free to use the storytelling technique of their choice. MELODY was introduced through a short tutorial, during which they learned how to use its core functionalities, such as performing SPARQL queries, selecting visualisation types and parameters, designing advanced interactive features like iterative text searches and filters. Students received immediate feedback from the platform, which gives previews of query results and selected charts on–the–fly. After the introductory session, students independently explored the platform to complete their projects. As part of their coursework, students were required to create one or more data–driven stories to explore topics in music history using the online MELODY platform. The analysis is available as a notebook on Zenodo [27].

Stories without visualisations have been pruned from the corpus. The final corpus that we analysed consists of 51 data stories. Students had the possibility to build these projects on top

[3] https://melody-data.github.io/stories/

of any Linked Open Dataset. Although, the majority (82% of stories) chose Wikidata as their primary data source.

Additionally, we collected qualitative feedback from students in two moments. We first performed a usability study during the course [14], asking for feedback on their experience on the platform. Secondly, during the project presentations part of the course examination, we collected qualitative feedback on the whole learning experience and in creating the data stories. Such qualitative reflections were systematically classified for sentiment analysis purposes. The dataset of students' data stories and the structured feedback data are publicly available in a Zenodo repository [28].

To answer our research question –whether data literacy skills are positively affected by easy–to–use tools for data storytelling–, our analysis focuses on (1) the frequency and distribution of visual elements (e.g., charts, tables, maps), as well as their intertwining in structuring narratives, (2) the proficiency in SPARQL, and (3) the perceived satisfaction of the learning experience. Aspect 1 tells us about data visualisation literacy skills achieved by students, 2 reveals whether adequate queries have been performed to support the storytelling, and 3 provides us with reflections on how the storytelling platform affected the former aspects.

4. Results

In the usability study, we gathered insights on the population of our study, which corresponds to the same population of students that later submitted one or more data stories as part of their assignment. A subset of the same population provided feedback during the exam. We asked students to rate their level of expertise in both Data visualisation and Semantic Web technologies (Figure 2). The population of the study is rather homogeneous, clearly identifying a group of novices, with slightly less confidence in data visualisation than in Semantic Web technologies such as SPARQL. All respondents declared they have a background in Languages and Linguistics.

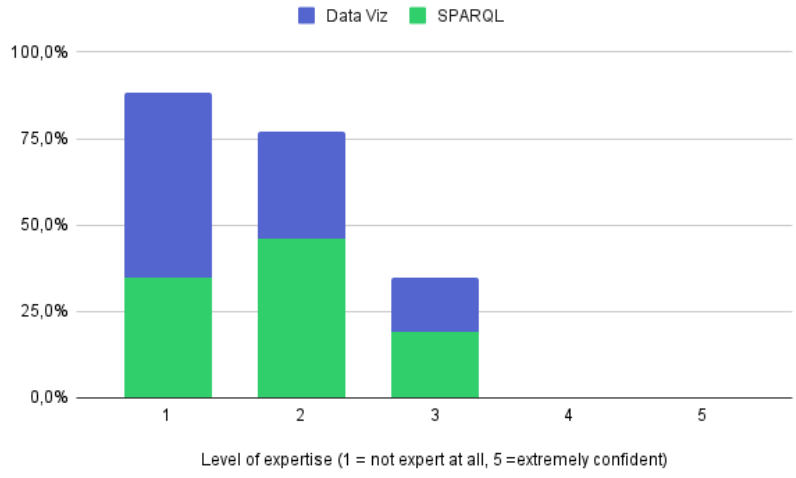

Figure 2 The population of the study is homogeneously formed by novices, with little or no knowledge of both data visualisation and SPARQL.

Data literacy and storytelling skills. The collected 51 data stories account for a total of 708 components, with an average of 13.9 components per story, although they can span from 4 to 31 (Figure 3). Components available in MELODY include: text fields, counters, data visualisations (i.e., bar charts, line charts, doughnut charts, maps, scatterplots), tables, and text searches. Texts, tables, counters, and maps are the elements that appear the most across data stories, which on average include at least 4.5 text elements, 3 tables, 2 counters, and one of the data visualisations.

In Figure 4 we show the distribution of aforementioned types of components across stories. Tables are the most popular visual form to present information clearly and concisely, which appear in copious amounts in 46 stories out 50. Other visual forms, such as doughnut charts, maps, and bar charts, were also widely used, although in a smaller amount per each story. Instead, line charts and scatter plots saw limited application.

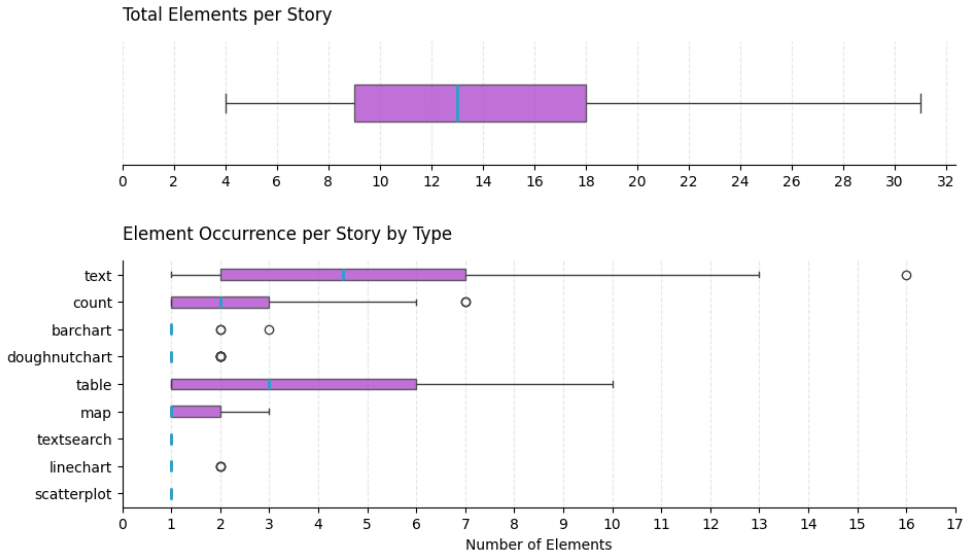

Figure 3 Average number of components (elements) per story.

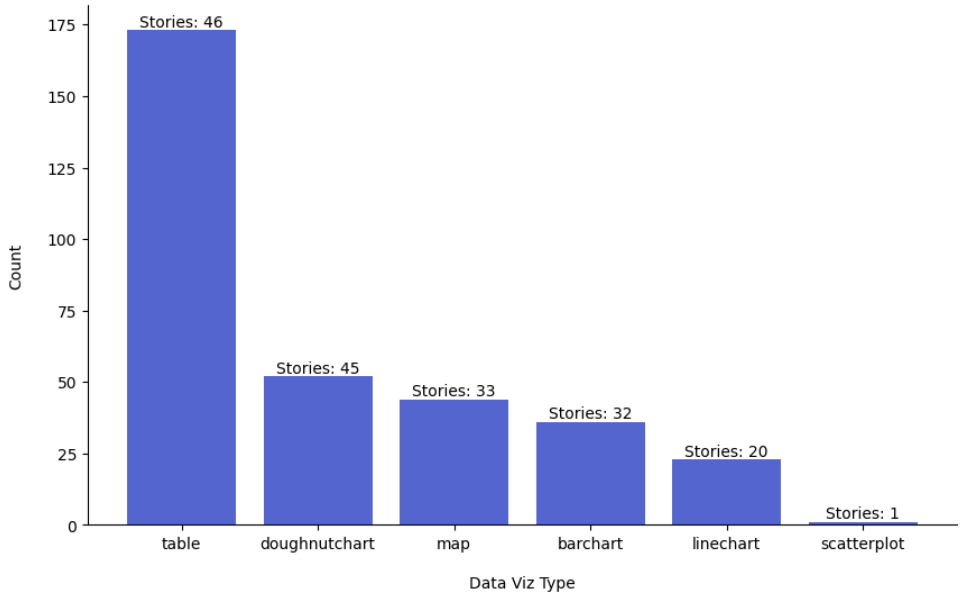

Figure 4 Tables are still the preferred tools of novices. Distribution of the usage of data visualisation types in data stories.

The interplay of visualisations and curated texts gives us precious insights on both communication and data literacy skills. Figure 5 illustrates the distribution and the order of components across narratives, namely: free texts, data visualisations (grouping tables, bars, line, scatter, doughnut charts and maps), counters, and text searches. Counters are distinguished from data visualisations because of their peculiar usage when providing figures in a larger narrative. Indeed, our assumption is that an author may want to juxtapose several counters next to each other to provide an overview of a phenomenon. Instead, data visualisations should be regarded differently, as they should appear in a more balanced amount, comparable to the number of explanatory texts. Since charts potentially condense a lot of information, we expect them to appear in an alternating order, so that explanatory texts accompany the reader through results [33].

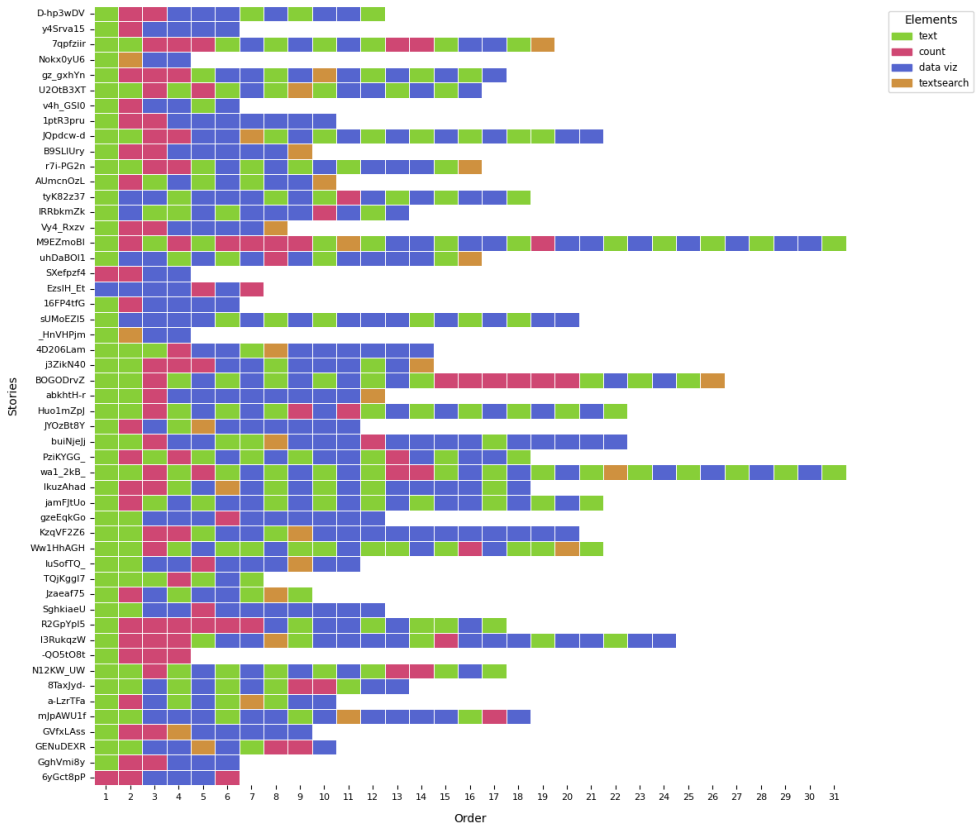

Figure 5 Alternating texts and charts. Distribution and order of appearance visual and text components in students' data stories.

Data visualisations (blue cells, Figure 5) are the most frequently used elements in data stories (329 in 51 stories), which appear scattered across the linear narratives, with an abundance in the initial positions (mostly between 3rd and 15th position). Similarly, text blocks (green cells) appear 244 times in 48 stories, with only a few exceptions not including any text description. Texts appear as introductions, but also along the whole story, showing in many cases a good alternation with charts. Counters (red cells) appear 106 times in 48 stories. Like texts, counters are often placed at the beginning of stories to highlight key statistics and provide a visual introduction to the dataset at hand. Text searches (orange cells) were less used, although they appear 29 times in 29 stories (56% of stories). Notably, they did not constitute a decisive element in the story, e.g. being the only interactive component of the story or being placed at the end of the story. 19 stories out of 29 integrate the text search in the middle of the story, along with charts and texts before and after it.

From Figure 5 emerges how text blocks are frequently used alongside visualisations, possibly to provide descriptions or explanations for charts. Likewise, the placement of text blocks at the end of the story suggests that students use them to summarise results highlighted by previous visualisations. However, only a limited number of stories ends with a conclusive text (23%), showing a limited attention to develop summarisation skills.

Moreover, we measured a negative correlation between the length of data stories (in terms of number of components) and the length of the last textual element (Figure 6). That is, long stories are often concluded with relatively concise summaries, while shorter stories tend to allocate extensive descriptions at the end. The highlighted pattern reflects the existence of two main approaches in combining visuals and text, possibly influenced by the complexity of the story at hand.

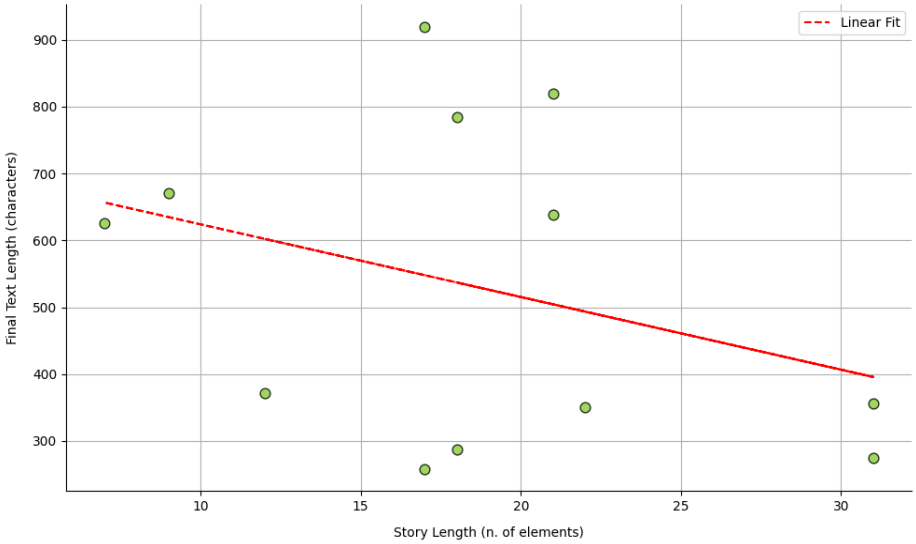

Figure 6 Long stories have short conclusions. Negative correlation between the length of a story and the length of the last text component.

It is worth noting that long stories refer to those having many components, regardless of these being text or visual ones, and regardless of the length of each text component. In Figure 7 we show the average length of text components (in characters). The visualisation reveals that on average natural language descriptions are rather short.

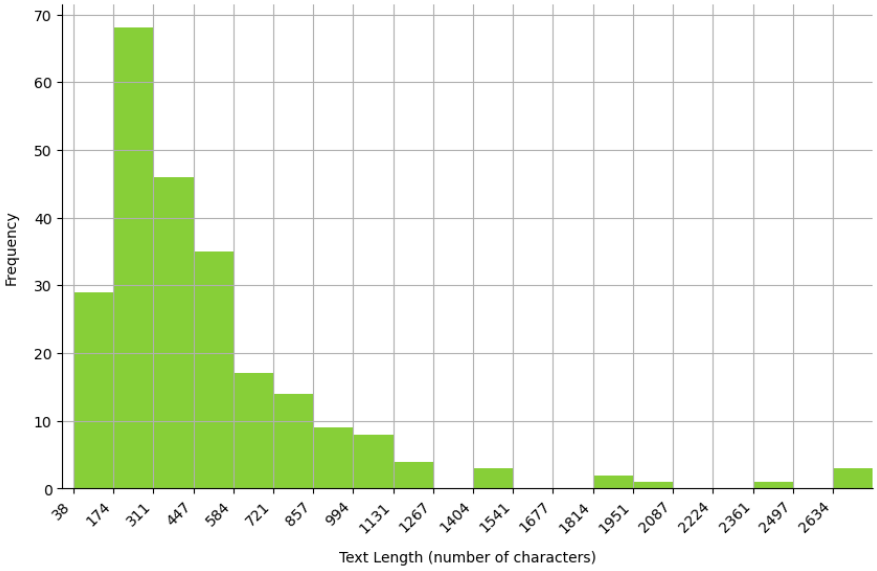

Figure 7 Explanations are short. Distribution of characters length of text components across data stories.

SPARQL proficiency. We analysed queries performed by students to produce the data visualisations above. Notice that we have not stored logs showing how students reached the latest queries later used to publish their data story, therefore, we can speculate on the latter only (i.e. the final queries), rather than on their iterative design process. We collected 464 SPARQL queries. Out of these, 26 are federated queries (5.6%), showing a small advanced group of students capable of appreciating the potential of the Semantic Web in answering complex questions via data integration.

Following the approach of [37], we analyse number of triples per query, types of triple patterns, and SPARQL features. We analysed SPARQL queries extracting subjects, predicates, objects, which together form triple patterns (e.g. "?s ?p ?o") and SPARQL operators (e.g. "SERVICE", which allows one to perform federated queries across several datasets at the same time).

Figure 8 shows that most queries tend to include less than 6 triple patterns, the majority with 2 patterns only, and a long tail of queries that include from 7 to 16 triple patterns.

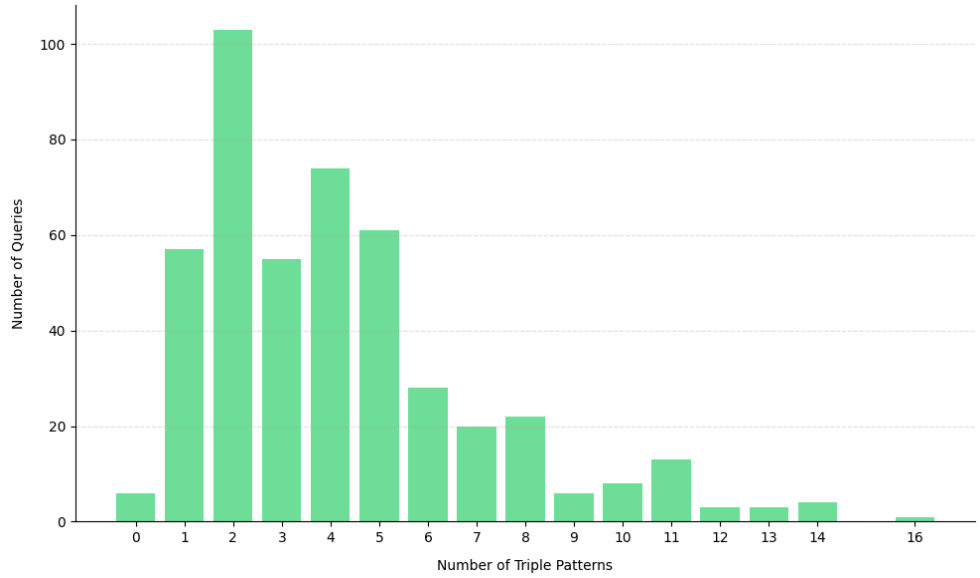

Figure 8 SPARQL queries tend to have often less than 6 triple patterns.

In Table 1 we show triple patterns including constants (C), i.e., URIs, and variables (V) in each of the subject, predicate and object positions. Students' SPARQL queries mostly rely on patterns such as VCC and VCV, which appear in over 60% of queries.

#	Pattern	Percentage
0	VVC	8.12%
1	VCC	**23.37%**
2	VCV	**40.28%**
3	CCC	15.81%
4	CCV	10.09%
5	CVC	1.48%
6	VVV	0.43%
7	CVV	0.43%

Table 1 SPARQL queries use triple patterns for exploratory purposes.

Lastly, we examined the usage of operators, shown in Figure 9. Notice that the operator "SERVICE" is used almost 300 times mostly to call Wikidata Labels service rather than to integrate external data sources. The graph shows the majority of queries using the DISTINCT and SERVICE, followed by COUNT, LIMIT and FILTER. COUNT is often a requirement to build a chart (e.g. counters, bar charts), while the usage of filtering methods and optional triple patterns are to be related to "post–processing" operations that are often required to retrieve clean or incomplete results.

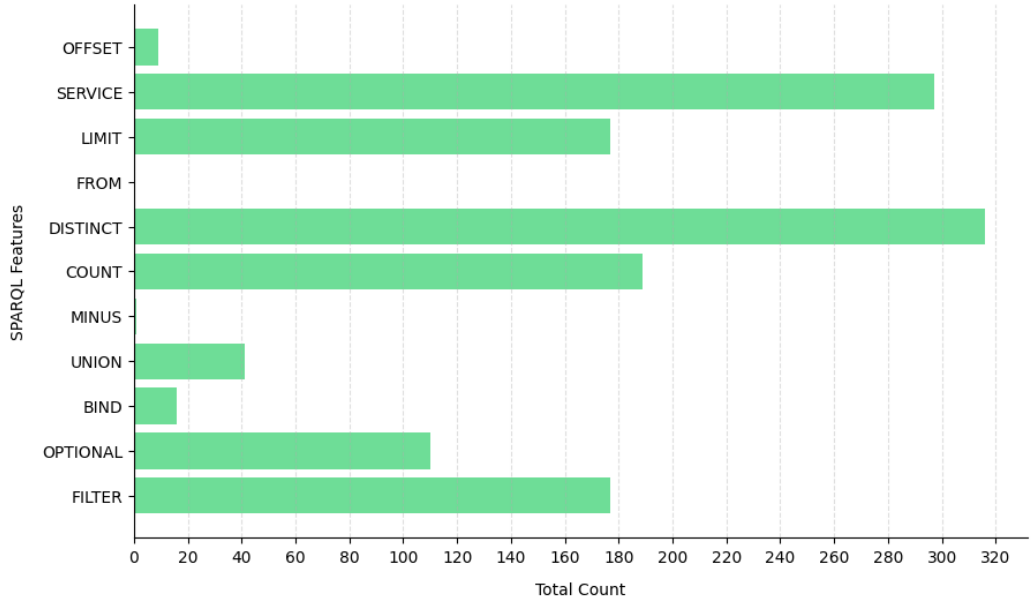

Figure 9 SPARQL operators are used to cope with data source issues, e.g. filter, optional, and for presentational aspects, e.g. limit.

5. Usability Study and Feedback

The usability test and student feedback provided us with valuable insights into their experiences using MELODY, highlighting both strengths and areas for improvement for WYSIWYG tools [28].

Usability test. After the tutorial and hands–on session, we asked students some follow–up questions. First, we asked to elaborate on their perception of the usefulness and complexity of text and visual components in MELODY (Figure 10). 26 students answered the questionnaire. The majority of participants deem counters, tables, and text boxes the most useful components to build a data story. The survey confirms the actual distribution of components usage (see Figure 3, Figure 4 and Figure 5). Notably, the fourth most useful component is the text search, which is instead less used. Usefulness and perception of complexity seem to go along when describing textboxes, counters and tables. Notably, all types of charts are similarly rated by more than half of students, which do not find them difficult to implement. Text searches are instead the least simple components.

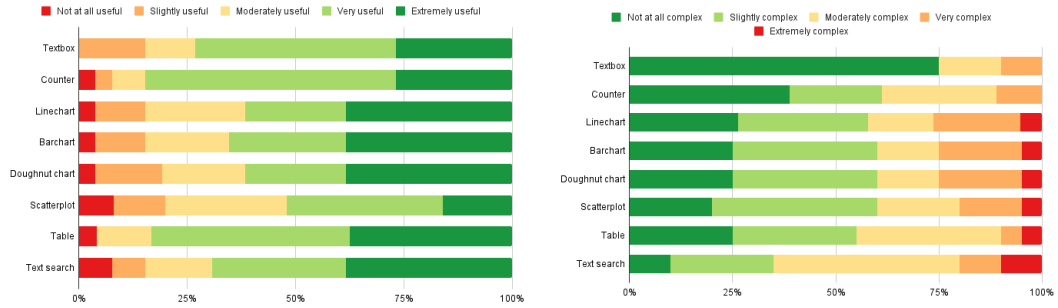

Figure 10 Novices find easy–to–consume visual components more useful than charts.

Secondly, we asked them to express their agreement on the ease perceived in their experience with the MELODY platform (Likert scale: 1=strongly disagree, 5= strongly agree). Results are shown in Table 2.

Likert	Ease of use	Harmony in the presentation of components	Inconsistency of components	Cumbersome to use	Confidence in usage
1	8%	0%	**19.2%**	16%	23.1%
2	16%	3.8%	**69.2%**	**36%**	**23.1%**
3	**36%**	**42.3%**	7.7%	**36%**	**30.8%**
4	**32%**	**38.5%**	3.8%	12%	15.4%
5	8%	15.4%	0%	0%	7.7%

Table 2 The perceived usability of the platform leans toward a general appreciation, although confidence needs to grow through prolonged usage.

Results of the questionnaire show a general reticence at this point of the course to answer with a strong opinion when verbalising their experience. We can assume this is due to the lack of prior experience with similar platforms, let alone with the technologies at hand, and therefore they cannot compare their current experience with previous ones. Nonetheless, the majority of students do not consider the platform inconsistent (88.4%, likert 1–2) and cumbersome (52%, likert 1–2). In turn, they mildly judge positively its easiness (40%, likert 4–5) and the harmony of components presented (53.9%, likert 4–5). However, confidence is still low at this point of the course (46.2% likert 1–2).

Student feedback. We collected open–ended answers during the final examination with respect to the perceived satisfaction of the learning experience. We were able to collect insights from people that created 14 data stories (27%). We classified answers according to four topics relevant to our research question, namely: (1) communication aspects enabled by the platform, (2) query and feedback aspects related to data sources, (3) general reflections on the whole experience, and (4) usability aspects of MELODY. Secondly, we performed manual sentiment annotation of answers, which were categorised as positive, neutral, and negative, and we qualitatively examined the answers to extract specific recurring problems.

Figure 11 shows the sentiment extracted from feedback grouped into the four main topics. In summary, overall negative perception is due to problems faced with incomplete or inconsistent data sources, which are deemed a common element of frustration. Students frequently encountered data quality or availability issues, which required them to refine their queries iteratively. Technical and usability issues in the platform, which did not give adequate feedback on errors they may have made in the construction of the query, and the steep learning curve also emerged as areas of frustration. Positive feedback is mostly related to MELODY's ability to facilitate communication through its diverse visualisation tools. Students found the platform effective for structuring and presenting data stories, as it enabled clear and engaging storytelling. Several students also reported that visualising charts enabled them to spot inconsistencies in data sources, facilitating the iterative refinement of queries in a positive query–feedback loop, therefore improving the quality of their analysis. Neutral responses include suggestions for improving MELODY's authoring workflow and tools, in order to simplify navigation, the editorial flow (e.g. more saving options) and add features (e.g. add timelines).

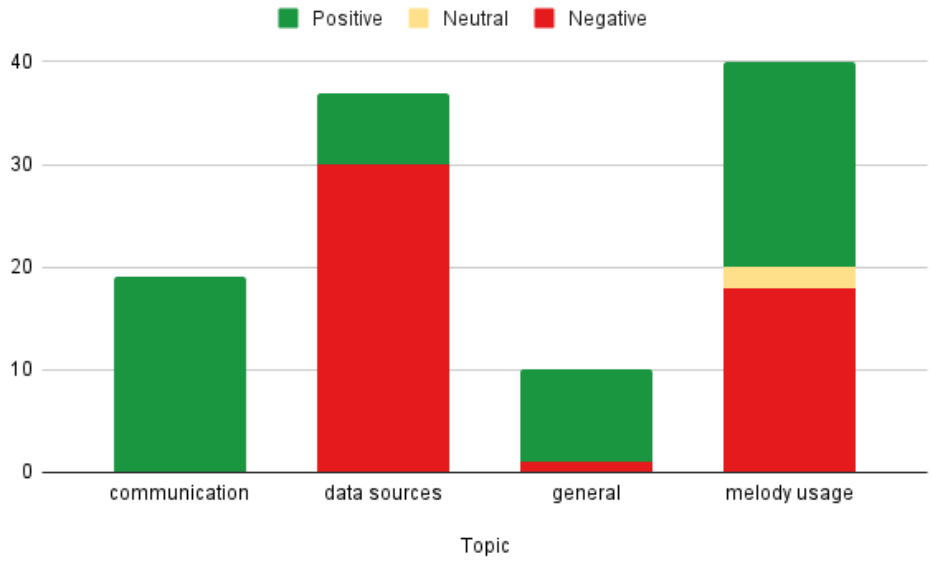

Figure 11 Trouble–shooting affects the learning experience. Sentiment analysis of students' feedback.

6. Discussion and Conclusion

This study provides an exploratory analysis of how students engage with MELODY to construct data–driven stories. In particular, the evaluation addresses students' perception of their learning experience and how the usage of a WYSIWYG tool supported them in their journey. Our findings reveal key patterns in how students use visual and textual elements to structure narratives, highlighting both opportunities and challenges in leveraging digital storytelling for Semantic Web education.

Data literacy and visualisation. The patterns observed in element usage analysis reveal important insights into how students engage with visual and textual storytelling tools. We noticed most stories are characterised by similar patterns, namely: having an introductory text, giving some figures at the beginning of the story using counters, alternating charts and explanatory, concise, texts. These practices suggest an emerging understanding of the narrative structure, where text and visuals complement each other to create a cohesive story. Still, a significant group of stories does not regularly present explanatory descriptions before or after each chart (Figure 5). This leads us to assume that the rhetorical aspects are sometimes overlooked by students overwhelmed by technical aspects. Summarisation skills are yet to be developed by most students, therefore a significant number of stories does not provide conclusions or actionable insights at the end of the story. Nonetheless, text components are generally short, showing awareness of communication modalities on the web. All in all, we can conclude that **such behaviours are evidence of an approach to Exploratory Data Analysis (EDA)**, characterised by (1) a high number of charts to showcase peculiarities of a data source with respect to a topic, and (2) little space for reflection, which is instead prominent in hypothesis–driven analysis, where the narrative is a walkthrough evidence collected to demonstrate a phenomenon and presented via visualisations [34]; [18].

The linearity of data stories seems to be embraced in most projects, where tables and interactive charts are widely used to integrate the rhetorical presentation with visual aids. Notably, not even text searches break such a linearity, since 19/29 stories include the text search component in the middle of the story rather than at the end. An iterative text search component can lead to several concatenations of results, therefore allowing for a wilder exploration of the data. While we would have expected stories to use it as the main exploratory tool, students seem to be more comfortable with the design of a traditional linear story, and elements breaking the linearity are either integrated (and limited), or avoided. This tendency reflects MELODY's design as a hybrid narrative environment: while it predominantly supports an author–driven linear structure, it incorporates elements of reader–driven interaction through components like filters and text searches. Such a combination is consistent with typologies of narrative visualisation structures that recognise the blending of linear exposition with limited interactive exploration, as discussed in prior work on storytelling strategies in visualisation environments [31].

Results of the usability survey tell us about the **relation between complexity, usefulness, and actual usage** of interactive components. On the one hand, the preference for simpler visualisation formats, such as tables and doughnut charts, suggests that students prefer visual tools that they feel intuitive and accessible in the design phase, and that are considered less error–prone in the understanding [16]; [7]. Indeed, those visualisations that do not require extensive data literacy skills neither to design nor to appreciate results are deemed very useful.

On the other hand, complex techniques (e.g. scatterplots) which require more sophisticated data analysis skills are perceived as less useful (Figure 10). For the same reason, we would have expected text searches to be underrated and underrepresented too, due to their complexity in the design phase (the iterative queries require a clear understanding of the usage of variables

inherited from previous queries to build the next result set). Instead, despite deemed complex components, both the perception of usefulness (Figure 10) and the actual usage of text searches in our population are interestingly satisfactory (Figure 5). To this extent, we can argue that **novices prefer easy–to–understand solutions that simplify the exploration to final users and that are perceived as useful, regardless of their design complexity**. This may also be significant evidence of a preference for exploratory tools based on natural language (i.e. the search box) that offer iterative, follow–your–nose, journeys.

SPARQL proficiency. Most queries use a limited number of triple patterns (mostly 2), which is evidence of little analytical approaches to dive into specificities of the chosen data source while privileging presentational aspects. The distribution of types of triple patterns confirms prior studies on most recurring triple patterns in large data sources [37]; [5]; [1], and suggests novices adopt a few exploratory approaches while looking for (1) relations with known objects (VCC), e.g. all entities uniquely related to term from a controlled vocabulary or to an entity that restricts the range of the analysis, and (2) relations between unknown entities (VCV). Again, the prevalence of exploratory triple patterns supports the hypothesis that **queries are designed for EDA** rather than for knowledge discovery [5]. The abundant usage of filtering methods shows an understanding of "post–processing" operations that are often needed to **retrieve clean results**. Likewise, the usage of OPTIONAL reveals those situations where **incomplete data sources** have to be appropriately addressed, showing evidence of frustration elements later confirmed by students' feedback.

Evaluation of the learning experience. The sentiment analysis performed over students' feedback highlights two important aspects related to learning tools like MELODY. The wide appreciation of MELODY communication features, lead us to believe that novices are looking for **tools that foster creative dissemination** and engagement with data (Figure 11). However, challenges posed by incompleteness, inconsistency, and difficulty to access data sources, emphasise the importance of developing more effective error–handling and feedback systems that can minimise frustration typical of trial–and–error approaches to query SPARQL endpoints [25]. A similar consideration applies to the negative correlation between perceived usefulness and complexity of components, like the text search (Figure 10). In fact, students seem to measure the satisfaction of their whole experience by positively judging the results of their work, which is perceived as useful, but show concerns on the way their task was performed, underlying little trust towards the correctness of data sources and results.

Pedagogical implications. The feedback analysis highlights the dual nature of using tools like MELODY for educational purposes. The platform's strengths lie in its ability to make abstract data tangible and enable critical and creative engagement through visual storytelling. Notably, students reported that visualising data helped them identify inconsistencies, facilitating query refinement and fostering essential problem–solving skills [12]. However, working with real–world data proved to be an iterative and sometimes frustrating process, particularly due to incomplete or inconsistent datasets and technical usability limitations. These findings highlight an important reality: while visualisation tools facilitate engagement with data, they also require students to navigate challenges that mirror real–world data analysis scenarios–a valuable aspect of learning data literacy.

A broader pedagogical implication emerges from the balance between textual and visual storytelling. While students showed an initial understanding of how these elements can complement one another, many narratives relied heavily on visuals, with text playing a minor role. Research suggests that combining textual explanations with visuals enhances both comprehension and engagement [22]; [39]. This observation points to the need for pedagogical frameworks that guide students in blending descriptions and visuals effectively. Future iterations

of MELODY–based courses could explore strategies to encourage students to develop more integrated storytelling techniques, ensuring a more balanced use of text and visuals.

Validity of our results and limitations. According to students' answers on their background (Figure 2 The population of the study is homogeneously formed by novices, with little or no knowledge of both data visualisation and SPARQL.), we can argue that the population of our study has no prior experience with WYSIWYG tools for data visualisation as well as with Semantic Web technologies, and that the test we conducted was their first time adopting such solutions. This hypothesis is confirmed by results of the usability perception, where no strong opinions (neither positive, nor negative) were expressed by participants on many aspects (Table 2). Such a reticence is likely due to their inexperience and the impossibility to judge by comparison.

To this extent, we can speculate about the impact of MELODY in students' learning journey without confounding variables such as previous knowledge of the system, previous knowledge of similar systems, and/or the technologies at hand. While we have not performed yet an A/B test to confirm our hypothesis that MELODY causally affects data visualisation skills, nor can we compare our results with similar experiments as baseline (that, to the best of our knowledge, are missing in the literature), we can argue that the nature of the population of our study is a sufficient condition to argue about the paramount role of MELODY in their learning experience. However, we are aware that such results cannot be generalised for other WYSIWYG platforms. Moreover, this study was conducted within a single master's course, while further experiments should confirm the validity of results with broader and less homogeneous populations. Future research will explore several key areas:

- A follow–up study will compare data stories created without MELODY, assessing whether MELODY unequivocally provides unique benefits in developing data literacy and visualisation skills in a Semantic Web environment.
- Expanding the research to students with varying levels of expertise will provide insights into how MELODY supports different learning trajectories, from beginners to advanced users.
- Investigating how students' storytelling and visualisation skills evolve over time through repeated exposure to MELODY will help refine its role in Semantic Web education.

In conclusion, this study demonstrates how tools like MELODY can bridge the gap between learning Semantic Web technologies and developing data and visualisation literacy skills. By empowering students in the production of meaningful, engaging, useful, data–driven narratives, novices are motivated to explore more sophisticated technical solutions, whether these are complex SPARQL queries or visual components. While challenges related to platform usability and workflow were identified, such difficulties also reflect the reality of working with real–world datasets. Future work will focus on expanding the research on how visualisation and storytelling impact data literacy in different learning contexts.

7. Acknowledgements

This work was partially funded by Project PE 0000020 CHANGES – CUP J33C22002850006, NRP Mission 4 Component 2 Investment 1.3, funded by the European Union – NextGenerationEU.

8. References

[1] Arias, Mario, Javier D. Fernández, Miguel A. Martínez–Prieto, and Pablo de la Fuente. 2011. 'An Empirical Study of Real–World SPARQL Queries'. In *1st International Workshop on Usage Analysis and the Web of Data (USEWOD2011) in the 20th International World Wide Web Conference (WWW2011), Hyderabad, India, March 28th, 2011*. New York, NY, USA: Cornell University Library, arXiv.org. https://arxiv.org/abs/1103.5043v1.

[2] Barber, John F. 2016. 'Digital Storytelling: New Opportunities for Humanities Scholarship and Pedagogy'. Edited by Ray Siemens. *Cogent Arts & Humanities* 3 (1). https://doi.org/10.1080/23311983.2016.1181037.

[3] Bhargava, Rahul, Erica Deahl, Emmanuel Letouzé, Amanda Noonan, David Sangokoya, and Natalie Shoup. 2015. 'Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data'. Data–Pop Alliance White Paper Series. Harvard Humanitarian Initiative, MIT Media Lab and Overseas Development Institute. https://dspace.mit.edu/handle/1721.1/123471.

[4] Bowler, Leanne, and Charlie Shaw. 2024. 'Trends in Data Literacy, 2018–2023: A Review of the Literature'. *Information Research an International Electronic Journal* 29 (2): 198–205. https://doi.org/10.47989/ir292822.

[5] Buil–Aranda, Carlos, Martín Ugarte, Marcelo Arenas, and Michel Dumontier. 2015. 'A Preliminary Investigation into SPARQL Query Complexity and Federation in Bio2RDF'. In *Proceedings of the 9th Alberto Mendelzon International Workshop on Foundations of Data Management, Lima, Peru, May 6 – 8, 2015*, edited by Andrea Calì and Maria–Esther Vidal. Vol. 1378. CEUR Workshop Proceedings. CEUR–WS.org. http://ceur–ws.org/Vol–1378/AMW_2015_paper_37.pdf.

[6] Burrows, Toby, Laura Cleaver, Doug Emery, Eero Hyvönen, Mikko Koho, Lynn Ransom, Emma Thomson, and Hanno Wijsman. 2022. 'Medieval Manuscripts and Their Migrations: Using SPARQL to Investigate the Research Potential of an Aggregated Knowledge Graph'. *Digital Medievalist* 15 (1). https://doi.org/10.16995/dm.8064.

[7] Cleveland, William S., and Robert McGill. 1984. 'Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods'. *Journal of the American Statistical Association* 79 (387): 531–54. https://doi.org/10.2307/2288400.

[8] Cross, Victoria L., Megan N. Imundo, Courtney M. Clark, and Melissa Paquette–Smith. 2024. 'Is There a Main Effect? Improving Data Literacy Using Practice Examples and Peer Collaboration'. *Psychology Learning & Teaching* 23 (2): 172–88. https://doi.org/10.1177/14757257231220057.

[9] Daquino, Marilena, Albert Meroño Peñuela, Christophe Guillotel–Nothmann, and Paul Mulholland. 2022. 'D1.9: Polifonia Web Portal – 1st Version (V1.0)', September. https://zenodo.org/records/7116711.

[10] Daquino, Marilena, Giulia Renda, and Marco Grasso. 2024. 'D1.10: Polifonia Web Portal – 2nd Version'. https://zenodo.org/records/11120387.

[11] D'Ignazio, Catherine, and Rahul Bhargava. 2018. 'Creative Data Literacy: A Constructionist Approach to Teaching Information Visualization'. *Digital Humanities Quarterly* 12 (4). https://dspace.mit.edu/handle/1721.1/123473.

[12] Evenstein Sigalov, Shani, and Rafi Nachmias. 2023. 'Investigating the Potential of the Semantic Web for Education: Exploring Wikidata as a Learning Platform'. *Education and Information Technologies* 28 (10): 12565–614. https://doi.org/10.1007/s10639–023–11664–1.

[13] Gershon, Nahum, and Ward Page. 2001. 'What Storytelling Can Do for Information Visualization'. *Communications of the ACM* 44 (8): 31–37. https://doi.org/10.1145/381641.381653.

[14] Grasso, Marco, Marilena Daquino, and Giulia Renda. 2023. 'MELODY User Survey'. Zenodo. https://doi.org/10.5281/zenodo.10442429.

[15] Gray, Jonathan, and Liliana Bounegru, eds. 2024. *The Data Journalism Handbook 2. Towards a Critical Data Practice*. European Journalism Centre and Google News Initiative.

[16] Heer, Jeffrey, and Michael Bostock. 2010. 'Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design'. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 203–12. CHI '10. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/1753326.1753357.

[17] Hogan, Aidan. 2020. 'The Semantic Web: Two Decades On'. Edited by Pascal Hitzler and Krzysztof Janowicz. *Semantic Web* 11 (1): 169–85. https://doi.org/10.3233/SW–190387.

[18] Jebb, Andrew T., Scott Parrigon, and Sang Eun Woo. 2017. 'Exploratory Data Analysis as a Foundation of Inductive Research'. *Human Resource Management Review* 27 (2): 265–76. https://doi.org/10.1016/j.hrmr.2016.08.003.

[19] Lo Duca, Angelica, and Kate McDowell. 2025. 'Using the S–DIKW Framework to Transform Data Visualization into Data Storytelling'. *Journal of the Association for Information Science and Technology* 76 (5): 803–18. https://doi.org/10.1002/asi.24973.

[20] Lodi, Giorgia, Luigi Asprino, Andrea Giovanni Nuzzolese, Valentina Presutti, Aldo Gangemi, Diego Reforgiato Recupero, Chiara Veninata, and Annarita Orsini. 2017. 'Semantic Web for Cultural Heritage Valorisation'. In *Data Analytics in Digital Humanities*, edited by Shalin Hai–Jew, 3–37. Cham: Springer International Publishing. https://doi.org/10.1007/978–3–319–54499–1_1.

[21] McDowell, Kate. 2021. 'Storytelling Wisdom: Story, Information, and DIKW'. *Journal of the Association for Information Science and Technology* 72 (10): 1223–33. https://doi.org/10.1002/asi.24466.

[22] McDowell, Kate, and Matthew J. Turk. 2024. 'Teaching Data Storytelling as Data Literacy'. *Information and Learning Sciences* 125 (5/6): 321–45. https://doi.org/10.1108/ILS–06–2023–0068.

[23] Miller, A. 2019. 'Data Visualization as Participatory Research: A Model for Digital Collections to Inspire User–Driven Research'. *Journal of Web Librarianship* 13 (2): 127–55. https://doi.org/10.1080/19322909.2019.1586617.

[24] Ojo, Adegboyega, and Bahareh Heravi. 2018. 'Patterns in Award Winning Data Storytelling: Story Types, Enabling Tools and Competences'. *Digital Journalism* 6 (6): 693–718. https://doi.org/10.1080/21670811.2017.1403291.

[25] Parkin, Louise, Brice Chardin, Stéphane Jean, and Allel Hadjali. 2022. 'Explaining Unexpected Answers of SPARQL Queries'. In *Web Information Systems Engineering – WISE 2022*, edited by Richard Chbeir, Helen Huang, Fabrizio Silvestri, Yannis Manolopoulos, and Yanchun Zhang, 136–51. Cham: Springer International Publishing. https://doi.org/10.1007/978–3–031–20891–1_11.

[26] Raffaghelli, Juliana Elisa. 2021. '"Datificazione" e istruzione superiore: verso la costruzione di un quadro competenziale per una rinnovata Digital Scholarship'. *Excellence and Innovation in Learning and Teaching – Open Access*. https://doi.org/10.3280/exioa0–2021oa11132.

[27] Renda, Giulia. 2025. 'A Component–Based Analysis of Data Stories Involving Semantic Web Data and Tools'. Zenodo. https://zenodo.org/records/15211624.

[28] Renda, Giulia, and Marilena Daquino. 2024. 'MELODY Data Stories and Anonymous Student Feedback Dataset'. Zenodo. https://doi.org/10.5281/zenodo.14499992.

[29] ——. 2025. 'Learning the Semantic Web with Tools for Information Visualisation and Data Storytelling'. In *IRCDL 2025. Information and Research Science Connecting to Digital and Library Science*, edited by Marcello Cornia, Giorgio Maria Di Nunzio, Donatella Firmani, Stefano Mizzaro, Giuseppe Serra, Sara Tonelli, and Alessandro Tremamunno, 3937:9. Udine: CEUR Workshop Proceedings. https://ceur–ws.org/Vol–3937/.

[30] Renda, Giulia, Marilena Daquino, and Valentina Presutti. 2023. 'Melody: A Platform for Linked Open Data Visualisation and Curated Storytelling'. In *Proceedings of the 34th ACM Conference on Hypertext and Social Media*, 1–8. HT '23. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3603163.3609035.

[31] Segel, Edward, and Jeffrey Heer. 2010. 'Narrative Visualization: Telling Stories with Data'. *IEEE Transactions on Visualization and Computer Graphics* 16 (6): 1139–48. https://doi.org/10.1109/TVCG.2010.179.

[32] Shreiner, Tamara L. 2019. 'Students' Use of Data Visualizations in Historical Reasoning: A Think–Aloud Investigation with Elementary, Middle, and High

School Students'. *The Journal of Social Studies Research* 43 (4): 389–404. https://doi.org/10.1016/j.jssr.2018.11.001.

[33] Stokes, Chase, Vidya Setlur, Bridget Cogley, Arvind Satyanarayan, and Marti A. Hearst. 2023. 'Striking a Balance: Reader Takeaways and Preferences When Integrating Text and Charts'. *IEEE Transactions on Visualization and Computer Graphics* 29 (1): 1233–43. https://doi.org/10.1109/TVCG.2022.3209383.

[34] Thudt, Alice, Jagoda Walny, Theresia Gschwandtner, Jason Dykes, and John Stasko. 2018. 'Exploration and Explanation in Data–Driven Storytelling'. In *Data–Driven Storytelling*. A K Peters/CRC Press.

[35] Vance, Eric A., David R. Glimp, Nathan D. Pieplow, Jane M. Garrity, and Brett A. Melbourne. 2022. 'Integrating the Humanities Into Data Science Education'. *Statistics Education Research Journal* 21 (2): 9–9. https://doi.org/10.52041/serj.v21i2.42.

[36] Vázquez–Ingelmo, Andrea, Alicia García–Holgado, Sonia Verdugo–Castro, Roberto Therón, and Francisco José García–Peñalvo. 2024. 'Data Visualization and Domain Knowledge: Insights through Focus Groups of Researchers in Spain'. *Computers in Human Behavior* 155 (C). https://doi.org/10.1016/j.chb.2024.108162.

[37] Warren, Paul, and Paul Mulholland. 2018. 'Using SPARQL – The Practitioners' Viewpoint'. In *Knowledge Engineering and Knowledge Management*, edited by Catherine Faron Zucker, Chiara Ghidini, Amedeo Napoli, and Yannick Toussaint, 485–500. Cham: Springer International Publishing. https://doi.org/10.1007/978–3–030–03667–6_31.

[38] Yan, Chunlai, Huan Wang, and Xuegang Luo. 2024. 'Knowledge Mapping of Data Literacy: A Bibliometric Study Using Visual Analysis'. *Journal of Librarianship and Information Science*. https://doi.org/10.1177/09610006241285512.

[39] York, Christopher. 2017. 'Exploratory Data Analysis for the Digital Humanities: The Comédie–Française Registers Project Analytics Tool'. *English Studies* 98 (5): 459–82. https://doi.org/10.1080/0013838X.2017.1332024