Data Affordances in Digital Humanities Information Visualization Practices

Giulia Renda[1,*,†], Marilena Daquino[1,†]

[1]Department of Classical Philology and Italian Studies, University of Bologna, Bologna, Italy

Abstract

Data storytelling in information visualization combines analysis, design, and narrative to communicate insights effectively. In this study, we examine 33 student projects from a master's course at the University of Bologna (2021–2025) that use cultural heritage datasets, often integrating multiple sources including Linked Open Data. We investigate how visualization practices reflect dataset characteristics, support exploratory or explanatory storytelling, and inform digital library design. While many metrics assess data quality, few consider how data affordances in visualizations signal dataset richness and integration potential. To address this gap, we introduce four metrics: Interaction Level (IL), Creator Complexity Index (CCI), Exploratory Data Analysis score (EDS), and Explanatory score (XPS), quantifying interactivity, authorial effort, and narrative orientation. Our analysis explores three research questions: (1) how visualization practices reveal data quality and complexity; (2) the relationship between data integration and exploratory or explanatory storytelling; and (3) implications for digital library policy and design.

Keywords

Data Affordances, Information Visualization, Cultural Heritage Data

1. Introduction

Data storytelling has become a central practice in information visualization, combining analytical rigor with narrative techniques to communicate complex insights effectively [1, 2, 3, 4]. In academic and professional contexts, students and practitioners are increasingly expected not only to analyze and visualize data, but also to craft coherent narratives that guide audiences through their findings. Understanding how these narratives are constructed, and how visualization practices interact with data characteristics, is essential to both advancing the pedagogy of information visualization and informing digital library design.

It is becoming standard practice for digital libraries to provide datasets that adhere to FAIR principles (Findable, Accessible, Interoperable, and Reusable) as these qualities facilitate data integration and support rich, multi-source analyses. While many metrics and automated monitoring platforms [5] exist to assess dataset quality and FAIRness, few studies have explored data quality indirectly through the affordances it provides in visualizations created by practitioners. In this sense, the way data is used in visualization—its ability to support integration, interactivity, and complex narrative construction—can serve as a meaningful indicator of its quality [6, 7].

In this study, we examine the visualization and storytelling practices of 33 student projects produced between 2021 and 2025 in a Digital Humanities master's course on Information Visualization at the University of Bologna. Students were tasked with creating data-driven stories around cultural heritage datasets, integrating multiple sources, including Linked Open Data, and developing narratives that address research questions within broad thematic areas such as art history, photography, and gender

IRCDL 2026: 22nd Conference on Information and Research Science Connecting to Digital and Library Science, February 19-20, 2026, Modena, Italy

*Corresponding author.

† Marilena Daquino was responsible for Section 1 (Introduction) and 6 (Conclusion), while Giulia Renda contributed to Sections 2 (Related Work), 3 (Materials and methods), 4 (Results) and 5 (Discussion).

giulia.renda3@unibo.it (G. Renda); marilena.daquino2@unibo.it (M. Daquino)

0000-0003-2990-0021 (G. Renda); 0000-0002-1113-7550 (M. Daquino)

studies. These projects provide a unique lens to investigate how visualization choices reveal data characteristics, support exploratory or explanatory storytelling, and inform digital library design.

We focus on three research questions. First, we explore how visualization practices reveal the quality and complexity of digital library data, asking whether more interactive or structurally complex visualizations correspond to richer, more heterogeneous datasets. Second, we investigate the relationship between data integration and the narrative orientation of data stories, distinguishing between exploratory and explanatory approaches. Finally, we consider the implications of these findings for digital library policy and design, aiming to translate observed practices into actionable guidelines for data curation and visualization support. Rather than establishing predictive relationships, this work proposes a structured framework for describing narrative and visual complexity in digital storytelling projects and provides an exploratory empirical analysis of how these dimensions co-occur in practice.

To quantify these aspects, we develop a set of metrics capturing both chart complexity and narrative strategy: the Interaction Level (IL), the Creator Complexity Index (CCI), the Exploratory Data Analysis score (EDS), and the Explanatory score (XPS). Together, these measures allow us to systematically characterize visualization effort, interactivity, and storytelling orientation, providing insight into how students leverage data integration and visualization techniques to construct meaningful narratives and, indirectly, how data quality enables these practices.

Results suggest that neither chart complexity nor interactivity systematically increases with dataset integration, and that visualization choices often reflect narrative intent more than the number of sources used. While exploratory stories show a modest inclination toward multi-source integration, explanatory ones can be effectively constructed from a single dataset. These findings indicate that visualizations cannot be treated as direct proxies for data quality, but they do reveal which data attributes—such as completeness, interlinking, or spatial enrichment—most strongly support common visualization tasks. This provides nuanced, evidence-based guidance for the curation and design of FAIR digital library datasets.

2. Related Work

Research on Cultural Heritage (CH) data has established Linked Open Data (LOD) as a paradigm for data dissemination [8, 9, 5]. Despite its wide usage, the complexity of graph data often hinders reuse in dissemination projects, prompting calls for "Linked Open Usable Data" [10]. On the one hand, prior work has extensively focused on LOD compliance to Findable, Accessible, Interoperable, Reusable (FAIR) principles [11] and schema validity [8, 5]. On the other hand, recent surveys have categorized narrative design choices driven by structured data [12]. However, little attention has been paid to data affordances in dissemination projects, i.e. what practitioners can actually do with these datasets, creating a gap between abstract quality assessments and the observable, narrative-level affordances expressed through visualizations.

In particular, data integration is constrained by heterogeneity of schemas, missing or uneven granularity of fields, inconsistent identifiers, and reconciliation challenges [13, 9]. Studies repeatedly highlight how these issues limit reuse for visualization, especially for spatial, temporal, and network-based representations. Metadata availability—such as geographic coordinates, controlled vocabularies, and stable identifiers—determines which visualization types are even feasible and whether data can support intended analytical tasks [14, 15]. These constraints are particularly evident in digital library contexts, where dataset structure strongly influences whether visualizations can support comparison, aggregation, or cross-collection narratives. To mitigate these constraints and support rich, multi-source analyses, adherence to FAIR principles in digital libraries is paramount for facilitating robust data integration and ensuring long-term reuse. The technical difficulty of querying multiple sources via interfaces like SPARQL [16, 17] further adds friction to advanced integration workflows (as also observed in CH data curation contexts, [14]). While recent work discusses LOD integration pipelines and proposes quality-improvement strategies, few studies examine how visualizations produced by users surface the underlying affordances (or limitations) of integrated cultural heritage data.

Our study aims to address this gap by analysing what data-driven web projects can tell us about CH data sources. We focus on narrative solutions, interactivity, and chart complexity, as potential proxies of data quality. We introduce a set of operationalizable measures (IL, CCI, EDS, XPS) and apply them to 33 student projects, revealing how visualization design indirectly reflects dataset characteristics and digital library support for integration.

In particular, research distinguishes between reader-driven exploratory narratives and author-driven explanatory ones, describing how structure, sequencing, annotation, and interaction shape interpretation [18, 19, 4]. Recognizing that visualizations serve both exploratory and explanatory purposes, exploratory forms rely on user agency (multiple linked views, zooming, filtering), while explanatory forms foreground interpretive guidance through titles, captions, and annotated highlights [15]. In these explanatory contexts, text plays a central role in conveying context and interpretation [19, 3]. We draw on this theoretical distinction to operationalize the concepts of exploratory and explanatory narratives. We designed the EDS metric to capture the degree to which a visualization invites open-ended investigation, whereas the XPS metric reflects the presence of interpretive cues that guide the reader toward specific insights.

Interaction taxonomies classify user actions according to intent (highlighting, filtering, zooming, navigating, querying) and many works propose hierarchical schemes reflecting increasing cognitive and analytical engagement [18]. These frameworks inform our **Interaction Level (IL)** metric, which captures the strongest form of interaction provided by a chart. Chart complexity has been discussed in terms of structural complexity (chart type), cognitive load, and authorial effort, with studies noting that multi-layered or composite visualizations require more sophisticated design decisions [20]. Our **Creator Complexity Index (CCI)** builds on these insights by incorporating chart type, analytic features, annotations, and interaction, while also discounting repetitions to avoid inflated estimates of effort.

Taken together, prior work has examined data quality, narrative visualization structures, and interaction or complexity taxonomies individually. However, within the Digital Humanities, visualization's interpretive power remains often undervalued in comparison to textual analysis [21]. No existing study has brought these strands together to examine how visualization practices within cultural heritage storytelling serve as indicators of data affordances and integration potential. In particular, no previous research has proposed metrics that quantify interactivity, authorial complexity, and narrative orientation to analyse how users engage with CH data in practice.

3. Materials and methods

We analysed 33 web projects that showcase a data story produced by master students of a course on Information Visualisation, running from 2021 to 2025 at the University of Bologna [22]. Students were introduced to data analysis, visualisation, and storytelling techniques to gather, interpret and present data, choosing effective visual representations to address research questions to communicate analytical results through data-driven storytelling. While the instructor suggested broad themes, such as art history, history of photography, gender studies, students then had to formulate their own research questions within the selected theme and develop data-driven analyses and narratives. Students were required to select one or more cultural heritage datasets, with a preference for Linked Open Data ones, in order to try and integrate them.

For each project, we collected minimal demographic information, such as the number and gender of contributors, and the data sources used. Then we split each data story into free-text and chart components, further classified by two annotators. Free-text components are classified by narrative role (title, introduction, research questions, data preparation, transition, counting, summary, conclusion), while chart sections include annotations on descriptive texts (section title, chart title, free description), chart type, chart features (e.g. if a bar chart is a stacked bar chart), interaction features (e.g. filters, zoom), if the chart has any annotation that highlights salient aspects, the number of data points, whether the data visualisation is a repetition (in a different form) of a previous one, and the number of data sources

used to create it.

After an exploratory analysis to appreciate visualisation and storytelling practices, we focus on our three research questions, namely:

- RQ1. How do visualization practices reveal the quality of digital library data? We investigate whether the visualization practices reflect underlying characteristics of datasets. Specifically, we examine whether more interactive or structurally complex visualizations tend to rely on a larger number of heterogeneous data sources. We do not assume a causal relationship; instead, we explore whether visualization design can indirectly signal the presence of integrated or enriched data. In parallel, we identify which visualization types explicitly depend on data integration workflows, providing insight into how data complexity shapes visualization needs.
- RQ2. What is the relationship between data integration and the exploratory or explanatory nature of data stories? We investigate how data integration relates to the exploratory and explanatory goals of data stories. We classify each story according to its dominant orientation—exploratory (EDA) or explanatory (XPS)—and examine whether these orientations are associated with the number and heterogeneity of data sources used in the project. This allows us to assess whether certain storytelling strategies tend to rely more heavily on integrated datasets.
- RQ3. What are the implications for digital library policy and design? Building on the findings from the previous two questions, we investigate the implications for digital library policy and design. Our goal is to translate the empirical relationships observed between visualization practices, data integration, and storytelling strategies into actionable guidelines for data curation and service design.

To quantify aspects relevant to chart complexity (RQ1) and narrative approach (RQ2) we defined four scores, namely: the interaction level (IL), the creator complexity index (CCI), the Exploratory Data Analysis score (EDA), and the Explanatory score (XPS).

The IL metric quantifies how interactive a visualisation is by classifying its interaction features into six families, each representing a progressively higher degree of user engagement. These families are weighted from 1 to 3, reflecting increasing cognitive load—from simple actions (e.g., hover) to complex narrative interactions (e.g., animation, scrollytelling). For each chart, all detected interaction tokens are mapped to their family, and the interaction level is defined as the highest weight present. This approach avoids inflating scores through multiple low-level interactions while capturing the strongest form of engagement the chart affords. The final metric ranges from 0 (no interaction) to 3 (high, narrative-level interaction).

The CCI measures how demanding a chart is to produce by combining four components: chart type, which sets a structural baseline (from simple bars to complex networks or sankeys); analytic features added by the author (e.g., comparisons, trendlines), which contribute small increments; interaction level, reflecting user-driven behaviours such as filtering or animation; annotations, which add a small bonus for narrative effort. To avoid overestimating complexity, charts that visualise the same subset of data in multiple forms are marked as repetitions, and their score is reduced with a 0.7 penalty factor. The resulting CCI values provide an interpretable measure of authorial effort and can be aggregated at the project level to compare how different stories deploy chart complexity in their visual narratives.

The Exploration Score (EDS) measures how much a story encourages users to investigate the data. At chart level, it combines interaction intensity, the strength of focus controls (e.g., filters, search, zoom vs. simple hover/click), and the presence of multiple related views of the same data segment. Scores are aggregated per project (using average chart score and maximum interaction level) and normalised to 0–1. The Explanation Score (XPS) captures interpretive guidance. It detects explanatory chart annotations and "results"-oriented titles, and narrative features such as summaries and conclusions, which signal interpretive closure. These elements are also aggregated and normalised to 0–1. Using median EDS and XPS as thresholds, each project is classified into one of four categories: EDA, XPS (explanation-led), Hybrid, or Descriptive.

All four metrics are automatically computed from structured annotations describing observable visualization features rather than from subjective quality judgments. For each chart, attributes such

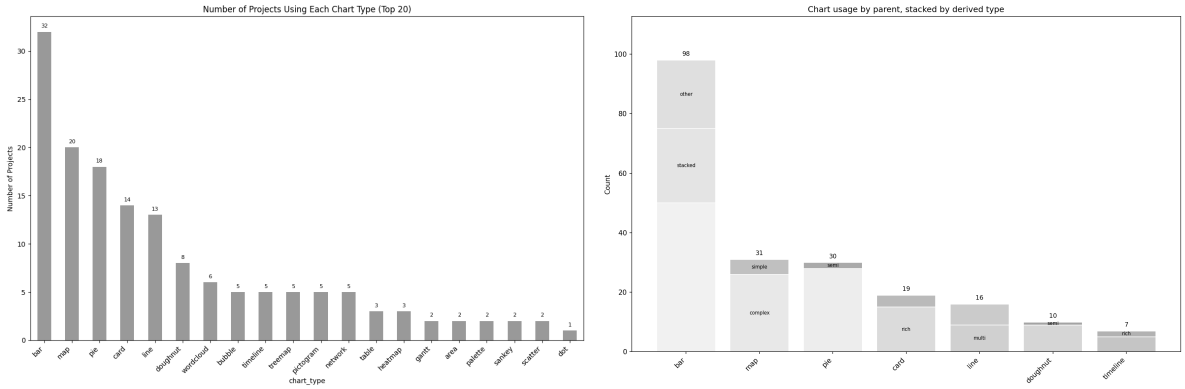

Figure 1: a) Chart type distribution in 33 projects; b) charts visual features.

as chart type, interaction mechanisms, presence of annotations, and analytic features were recorded following a predefined schema. The annotator's role was limited to identifying explicit interface and narrative elements (e.g., presence of filters, annotations, chart composition). Metric values are derived deterministically from these attributes through fixed weighting and aggregation rules, being therefore fully reproducible and not depending on interpretive scoring. For example, a chart including filtering and zoom interactions is automatically assigned a higher IL value than a static chart, while CCI increases with the presence of composite chart structures and analytic overlays.

4. Results

Projects mostly rely on well-known, easy-to-understand and implement charting solutions, such as bar charts, maps, and pie charts (Fig. 1a). Popular charts often include more sophisticated visual features, e.g. maps including filters or facets (Fig 1b).

Overall, all stories present a good alternation between charts and texts (Fig. 2), which we believe being a peculiarity of good explanatory storytelling. 19 projects (58%) present the research questions leading the story, and 21 projects (63%) present data sources and data preparation operations for the sake of the analysis – other good indicators of explanatory approaches. However, only 13 stories (40%) present a conclusion where actionable or generalisable insights are derived from the analysis, and 8 projects (24%) present a summary of insights presented in charts, without further elaborating on them.

We observed a weak but statistically significant negative correlation between chart Interaction Level (IL) and the number of data sources (Pearson $r = -0.138$, $p = 0.017$; Spearman $\rho = -0.138$, $p = 0.016$; Kendall $\tau = -0.126$, $p = 0.016$) (Fig. 3a). This indicates that, in our sample, more interactive charts tend to rely on slightly fewer data sources. While the effect is small, it suggests that high interactivity does not necessarily require extensive data integration. Instead, interactive charts may often focus on exploring or emphasizing complex aspects within a single dataset, rather than integrating multiple sources. These findings imply that, although interactivity can enhance data storytelling, it is not a straightforward indicator of dataset richness or integration. Notably, most charts (200/302) are static visualisations (Fig. 4).

In contrast to interaction level, chart complexity (CCI) of single charts exhibited no meaningful correlation with the number of sources (Pearson $r = -0.022$, $p = 0.700$; Spearman $\rho = 0.022$, $p = 0.702$; Kendall $\tau = 0.016$, $p = 0.731$) (Fig 3b), suggesting that the structural or analytic complexity of charts does not depend on dataset integration. These findings imply that high interactivity may focus on exploring or emphasizing a single dataset, while chart complexity is largely determined by authorial choices such as chart type, analytic features, and narrative intent, rather than the number of data sources integrated.

At the project level, the relationship between chart complexity and the number of data sources depends on how complexity is measured. When considering average chart complexity (mean CCI, Fig.

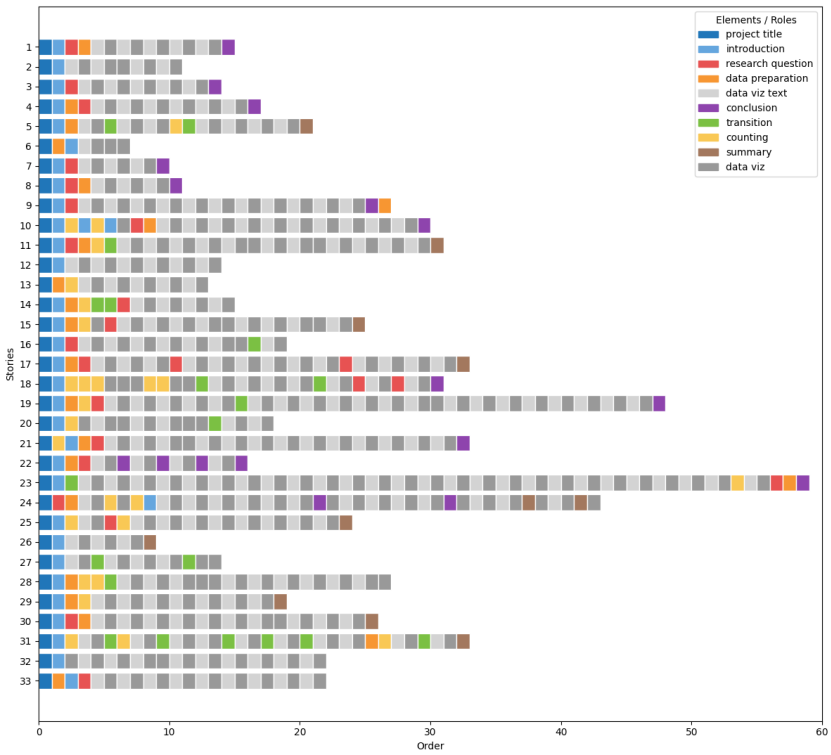

Figure 2: Sequence of text and chart components in 33 projects.

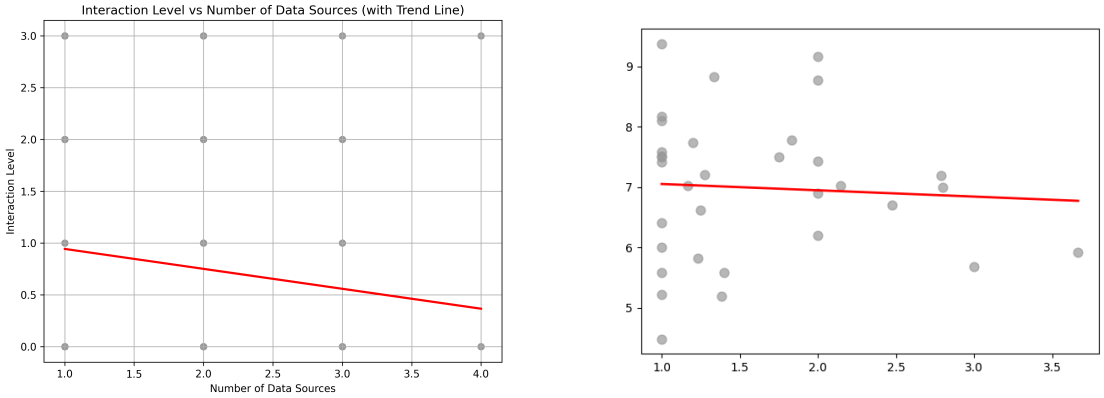

Figure 3: a) Correlation between interaction level (IL) in single charts and the number of data sources b) correlation between chart complexity (CCI) and the number of data sources.

5a), correlations with the number of sources are negligible and not statistically significant (Pearson $r = -0.063, p = 0.730$; Spearman $\rho = -0.080, p = 0.660$; Kendall $\tau = -0.059, p = 0.644$), indicating that the typical level of chart complexity in a project does not predict dataset integration. In contrast, when focusing on the most complex chart in each project (max CCI, Fig 5b), correlations are positive and somewhat stronger (Pearson $r = 0.247, p = 0.165$; Spearman $\rho = 0.331, p = 0.059$; Kendall $\tau = 0.252$, $p = 0.062$), suggesting that projects featuring at least one highly complex chart tend to integrate more data sources. Although these correlations are only marginally significant, the Pearson correlation does not reach statistical significance (p = 0.165). Therefore, complexity alone cannot be considered a reliable predictor of data integration in this sample. Nonetheless, the comparison of patterns suggests that maximum chart complexity may better capture the relationship between visualization effort and dataset integration than average complexity. In other words, while most charts in a project may be simple, the

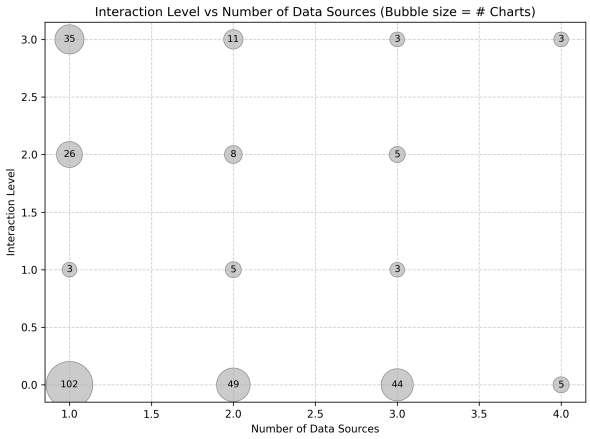

Figure 4: Bubble chart showing that most charts are static and use few data sources.

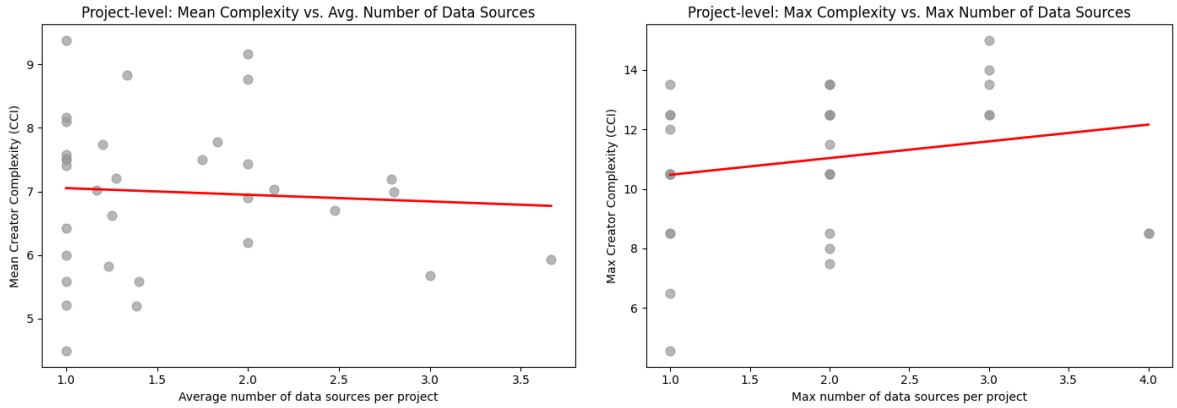

Figure 5: a) Correlation between Mean CCI in projects and number of data sources b) Correlation between highest CCI in projects and number of data sources.

presence of a highly demanding chart often reflects the need to combine and visualize heterogeneous sources, highlighting a link between authorial effort and data richness.

Notably, in our study, 73% of chart types (30/36 types) may require multiple data sources to be built (Fig. 6), and 28% types (10/36) always require multiple data sources, revealing a latent dependency on data integration across any type of chart, regardless of its complexity.

The heatmap (Fig. 7) represents the co-occurrence of data sources and chart types across the 33 projects (values are normalised in percentage). It reveals a pronounced concentration of data-source usage within a limited subset of collections, indicating that digital humanities visualization projects tend to rely on a relatively small number of well-structured or readily accessible repositories. This is evidenced by the cluster of chart types on the left of the matrix (e.g., *bar*, *pie*, *treemap*, *timeline_rich*) that exhibit multiple medium- to high-intensity cells (approximately 40–100%) across the first several sources on the x-axis. In contrast, data sources positioned toward the right margin display predominantly pale cells (0–20%), suggesting only occasional or single-project usage. This asymmetry implies that factors such as data completeness, standardization, or existing scholarly familiarity may govern the integration of particular sources into visualization workflows.

In terms of chart-type requirements, the visualization demonstrates that complex or multi-dimensional visualizations, such as *timeline_rich*, *treemap*, *network*, and *card-rich*, engage with a broader array of sources, each showing numerous mid-range intensity cells across ten or more datasets. This pattern suggests that these chart types demand heterogeneous or multi-faceted inputs and therefore draw upon diverse data repositories. Conversely, highly specialized or methodologically narrow chart types

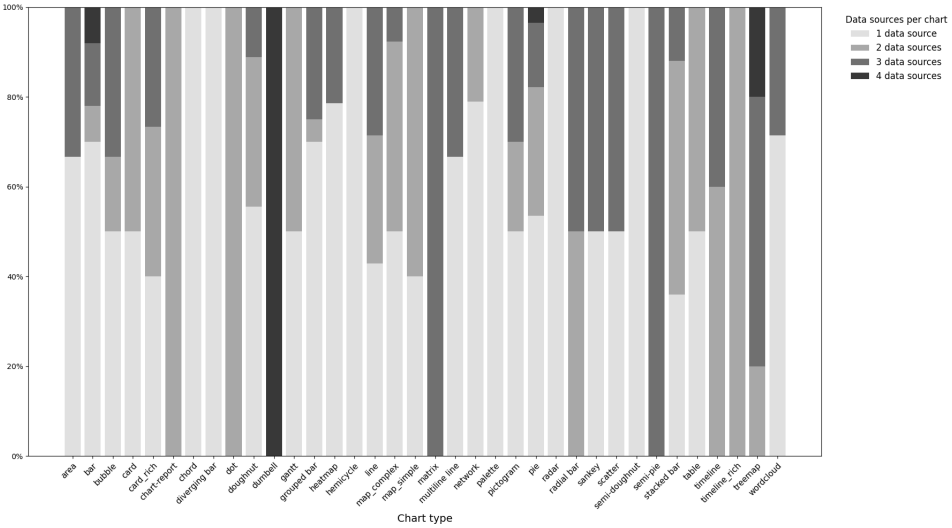

Figure 6: Number of data sources used to build specific chart types.

(including dot, chord, radial bar, and semi-doughnut) are associated with only one or two high-intensity cells (80–100%), with the remainder of the row effectively empty. This indicates that such visualizations are typically constructed from single, well-defined datasets and do not generalize across broader data ecosystems. Finally, chart types dependent on structured categorical or geographic information (e.g., *map_complex*, *network*) show moderate-intensity cells (30–50%) clustered within a few specific sources, reflecting more constrained data requirements that limit cross-source applicability.

We examined how the narrative orientation of projects relates to dataset integration (Fig. 8). Projects with a more exploratory orientation showed a weak positive correlation with the number of data sources (Pearson $r = 0.210$, $p = 0.241$; Spearman $\rho = 0.270$, $p = 0.128$; Kendall $\tau = 0.193$, $p = 0.160$), suggesting a slight tendency for exploratory stories to incorporate more datasets, although the relationship is not statistically significant. In contrast, projects with a predominantly explanatory orientation exhibited very weak and non-significant correlations with the number of sources (Pearson $r = 0.094$, $p = 0.604$; Spearman $\rho = 0.129$, $p = 0.474$; Kendall $\tau = 0.092$, $p = 0.516$), indicating that explanatory narratives do not systematically depend on integrating multiple datasets. Overall, these results suggest that while exploratory storytelling may modestly encourage the use of heterogeneous sources, explanatory storytelling can be effectively constructed with a single or few datasets, highlighting differences in how narrative strategies relate to data integration.

The project-level EDA/XPS map reveals distinct patterns in the distribution of exploration and explanation scores across the sampled projects (Fig 9).

- Most projects cluster in the upper-right quadrant, indicating above-average performance on both exploration and explanation, suggesting a subset of projects effectively balance knowledge discovery with interpretability.
- A smaller cluster appears in the lower-left quadrant, representing projects with consistently low scores on both dimensions.
- Two minor clusters are observed in the off-diagonal quadrants: one with high exploration but low explanation, and another with low exploration but high explanation, highlighting cases where projects excel in one dimension while underperforming in the other.
- Additionally, a tight grouping near the intersection of the dashed mean lines (approx. 0.4, 0.4) represents projects with average scores on both metrics. Outliers at the extremes further emphasize the heterogeneity within the dataset.

Overall, while a weak positive association is suggested between exploration and explanation, the spread across quadrants illustrates diverse project profiles, ranging from well-balanced high performers to specialized or underdeveloped cases.

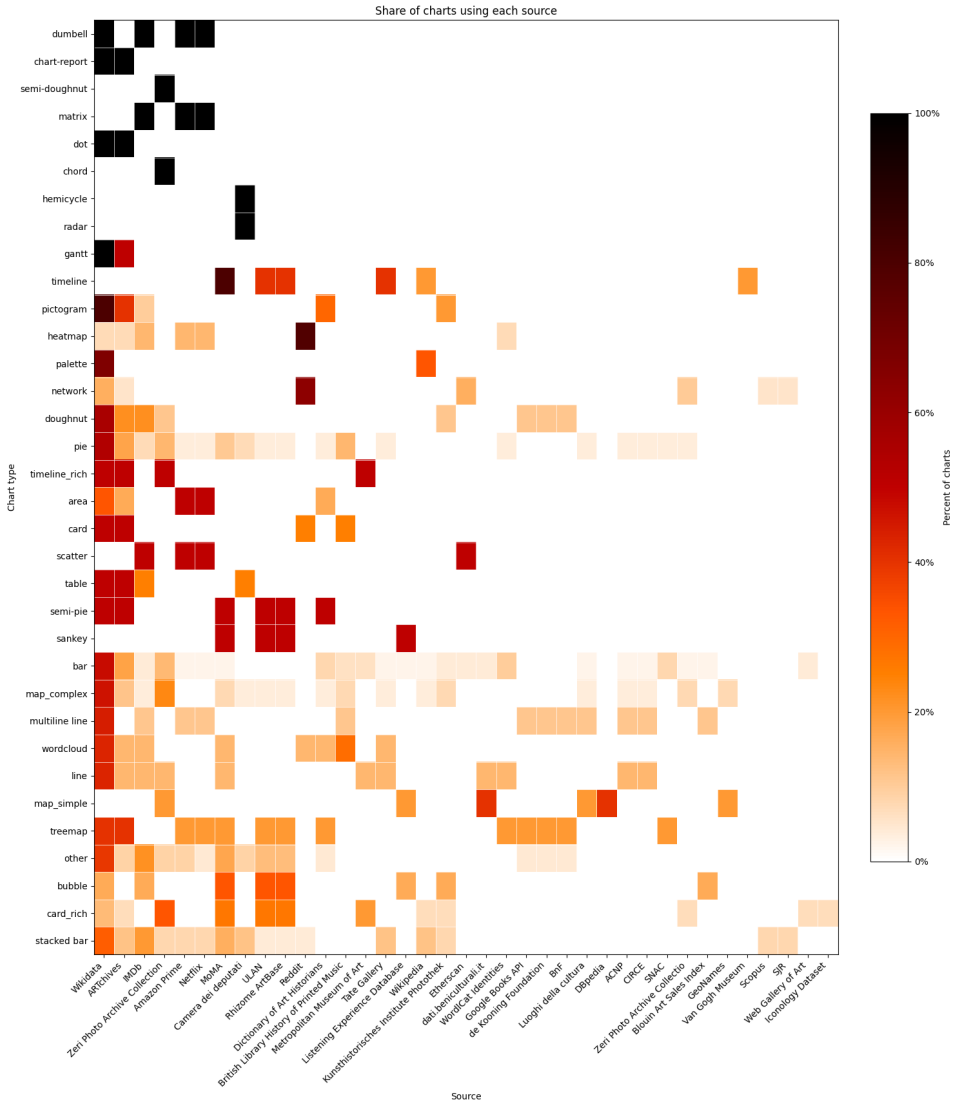

Figure 7: Normalised distribution of data sources required to build charts, grouped by type.

5. Discussion

Our findings offer several insights into how visualization practices reflect data characteristics, how narrative strategies relate to data integration, and what these patterns imply for digital library design. The heatmap analysis further reinforces these conclusions by showing how specific visualization types and data sources interact in practice.

Results suggest that visualization affordances are only weakly coupled with the underlying data richness of the projects (RQ1). Neither interaction level (IL) nor chart complexity (CCI) reliably predicted the number of integrated data sources. This pattern aligns with the heatmap, where a small cluster of sources on the left side shows high-intensity usage (approx. 60–100%) across many chart types, while the majority of sources display only pale, low-intensity cells (approx. 0–20%). Despite this clear asymmetry in data-source preference, most chart types (especially common ones such as *bar*, *line*, and *map*) appear across a wide array of projects regardless of how many sources those projects integrate. Interaction level even showed a weak negative correlation with dataset heterogeneity, suggesting that interactive features were often used to explore internal structure within a single dataset rather than to navigate multiple integrated ones. Similarly, structural or analytic chart complexity appears to stem more from authorial design choices than from data availability, a point reflected in the heatmap's observation that

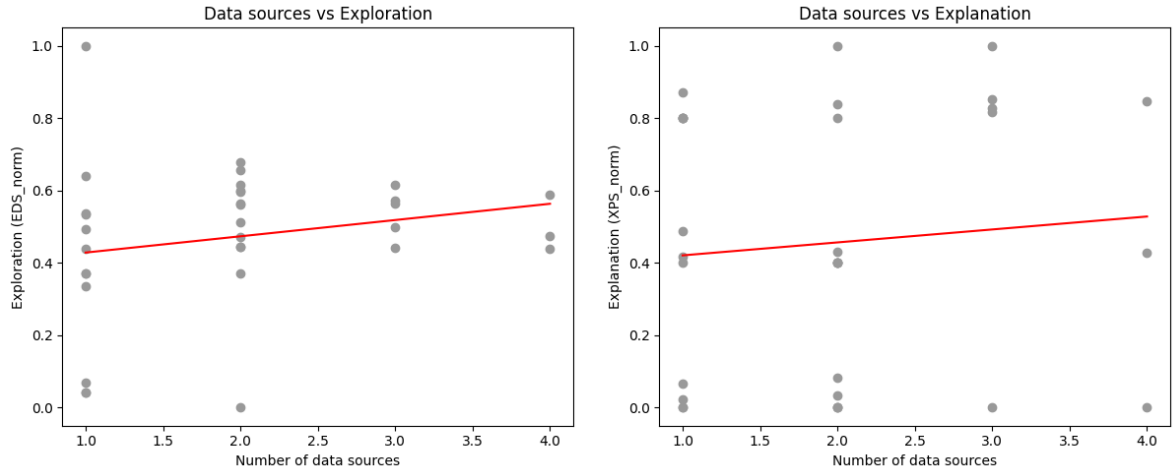

Figure 8: a) correlation between exploratory approach in projects and number of sources b) correlation between explanatory approach in projects and number of sources.

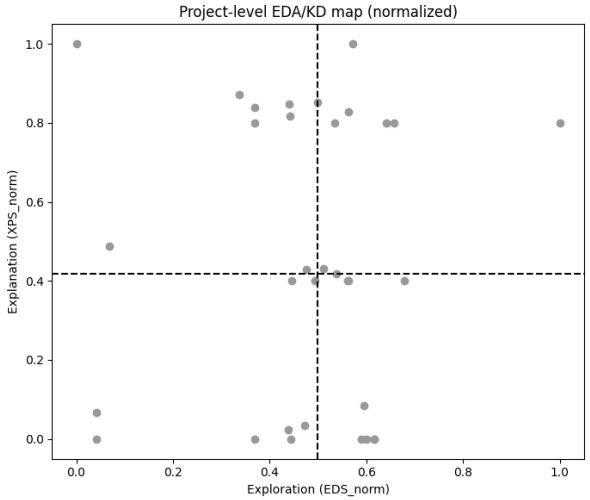

Figure 9: Classification of projects based on exploratory or explanatory narrativity.

specialized visualizations (e.g., *dot*, *chord*, *radial bar*) typically rely on one or two highly used sources rather than on broadly integrated data. This aligns with recent prototype development efforts where high interactivity is achieved through the deep exploration of single, well-structured collections rather than broad integration [23].

Yet the project-level analysis nuances this picture. While the average chart in a project rarely reflects data integration effort, the most complex chart often does. This suggests that data integration manifests not as a general elevation of visual complexity across a project, but rather in the creation of one or two strategically complex visualisations, typically those requiring reconciliation, enrichment, or combination of heterogeneous sources. The heatmap corroborates this: complex chart types such as *timeline_rich*, *treemap*, *network*, and *card-rich* show mid-range intensities (approx. 40–60%) across ten or more sources, indicating that these visualizations selectively appear where integration is most likely. In this sense, visualization complexity functions less as a continuous signal of data quality and more as a local indicator of data integration effort.

These findings challenge assumptions within the digital library community that richer datasets will automatically translate into more sophisticated or interactive visualisations. Instead, they show that visualization complexity is a selective outcome, emerging only where authors intentionally leverage integrated data to address specific analytical needs.

The relationship between data integration and narrative strategy reveals further subtleties (RQ2). Exploratory-oriented projects (high EDS) showed a modest tendency to use more data sources, whereas explanatory-oriented projects (high XPS) did not. This aligns with the distinctive aims of exploratory and explanatory storytelling: exploration typically benefits from heterogeneity, allowing users to navigate multiple dimensions of data, while explanation often relies on a curated subset of sources that support a clear argumentative arc. The heatmap mirrors this distinction: exploratory-friendly chart types (*map_complex*, *network*, *timeline_rich*) exhibit more dispersed, medium-intensity cells across numerous sources, whereas explanatory staples (e.g., *bar*, *pie*) show high-intensity usage concentrated in just a few main sources. The EDA/XPS map underscores this division. The most advanced projects excel in both exploration and explanation, balancing interactive discovery with strong narrative closure [4]. These hybrid projects frequently occupy the upper-right quadrant and tend to draw on richer data ecosystems. In contrast, purely exploratory dashboards and purely explanatory essays each rely on different types (and amounts) of data integration, showing that narrative strategy mediates how data is mobilized rather than simply reflecting the number of sources available. Together, these patterns suggest that data integration is not a prerequisite for strong explanatory storytelling, but it can enhance exploratory depth and produce more versatile narrative designs.

Across both research questions, a central implication emerges: visualization practices expose not only data quality but also the extent to which digital libraries support integration (RQ3). Students often resorted to simple charts because many cultural heritage datasets lack essential fields (e.g., geographic coordinates, timestamps, reconciled identifiers) or provide them inconsistently. A similar trend has also been observed in surveys of DH projects [12]. The heatmap further illustrates how chart types requiring structured metadata—such as map-based and network visualizations—appear only where specific fields are present, as evidenced by their moderate-intensity cells clustered around a limited set of well-prepared sources [15]. When structured metadata is available, students are more likely to design complex or integrative visualisations. Importantly, this highlights a novel perspective on data quality in digital libraries: even when multiple sources are available, the ability of data to support interactive or complex visualizations can serve as an indirect indicator of its FAIRness and integration potential, providing insights that traditional metrics alone may not capture.

Given this, digital libraries aiming to support rich visual storytelling should prioritize:

- Self-contained datasets for common visualization tasks, ensuring that widely used metadata (locations, dates, entity identifiers) is complete and reliable without requiring external reconciliation.
- Interoperable identifiers and crosswalks, enabling frictionless integration of heterogeneous sources and supporting more advanced exploratory narratives.
- Clear documentation of data granularity and completeness, helping creators anticipate which visualisations are feasible.
- Tools or APIs that streamline integration workflows, lowering the barrier to constructing high-complexity, multi-source visual stories [8, 24].

Finally, the fact that visualization complexity often appears in isolated moments rather than across entire projects suggests that digital library resources should be designed to support these "high-effort points"—the specific visualisations that demand integration—rather than assuming a uniform trajectory of complexity throughout a narrative.

Overall, the study indicates that visualization practices can serve as indirect indicators of data quality, but only when interpreted through the lens of narrative intent and authorial strategy. Digital libraries that understand this relationship can better anticipate user needs and design services that truly support rich, integrated, and interpretable data storytelling.

Limitations. This study is based on projects developed within a single master's course, which inevitably reflects specific pedagogical goals, tools, and instructional constraints. As a result, the observed visualization practices should not be interpreted as representative of professional or large-scale digital humanities projects. Rather, the dataset provides a controlled environment to explore early design tendencies and emerging narrative strategies. Future work will extend the analysis to professional and publicly available visualization projects.

6. Conclusion

This study analysed 33 student data stories to understand how visualization practices relate to dataset characteristics, narrative strategies, and digital library design. Using four metrics—Interaction Level (IL), Creator Complexity Index (CCI), Exploratory score (EDS), and Explanatory score (XPS)—we assessed how authors engage with cultural heritage data and how storytelling styles emerge from data affordances.

Findings show that interactivity and chart complexity do not consistently increase with the number of data sources. Interactive and complex charts often rely on well-structured single datasets, while only the most demanding visualizations (high CCI) show a modest link to data integration. Exploratory narratives tend to use slightly more heterogeneous sources, whereas explanatory narratives can be built from one or two coherent datasets. These results suggest that visualization design reflects authorial choice more than raw dataset richness.

For digital library design, the key implication is that data should provide strong affordances for common visualization tasks—such as geolocation, stable identifiers, and interlinking—so users can create both exploratory and explanatory narratives without heavy integration work. Improving metadata completeness and interoperability can directly enhance the storytelling potential of cultural heritage datasets.

Overall, examining data stories offers a valuable perspective on how data quality manifests in practice and how digital libraries can better support analysis, creativity, and narrative communication. This work highlights the value of examining data storytelling as a lens onto dataset affordances. By observing how authors use and struggle with data, we gain insight into how digital library infrastructures can better support analysis, creativity, and narrative communication. The proposed metrics and findings can inform both pedagogical practices in visualization education and the ongoing development of FAIR-aligned data services within cultural heritage institutions.

Acknowledgments

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.

Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT in order to: Grammar and spelling check, Paraphrase and reword. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] H. Shao, R. Martinez-Maldonado, V. Echeverria, L. Yan, D. Gasevic, Data Storytelling in Data Visualisation: Does it Enhance the Efficiency and Effectiveness of Information Retrieval and Insights Comprehension?, in: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, CHI '24, Association for Computing Machinery, New York, NY, USA, 2024, pp. 1–21. URL: https://dl.acm.org/doi/10.1145/3613904.3643022. doi:10.1145/3613904.3643022.

[2] K. McDowell, Storytelling wisdom: Story, information, and DIKW, Journal of the Association for Information Science and Technology 72 (2021) 1223–1233. doi:10.1002/asi.24466.

[3] R. Bhargava, E. Deahl, E. Letouzé, A. Noonan, D. Sangokoya, N. Shoup, Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data, Technical Report, Harvard Humanitarian Initiative, MIT Media Lab and Overseas Development Institute, 2015. URL: https://dspace.mit.edu/handle/1721.1/123471.

[4] E. Segel, J. Heer, Narrative Visualization: Telling Stories with Data, IEEE Transactions on Visualization and Computer Graphics 16 (2010) 1139–1148. URL: https://ieeexplore.ieee.org/document/5613452. doi:10.1109/TVCG.2010.179.

[5] G. Tuozzo, M. A. Pellegrino, A. Lieto, CHeCLOUD—the Cultural Heritage Linked Open Data Cloud, in: I. Celino, O. Hassanzadeh, A. Bernstein, N. Noy, G. Cheng, S. Wang, S. Ferrada, T. Soulard, K. Kozaki, H. Takeda, A. L. Gentile (Eds.), Joint Proceedings of Industry, Doctoral Consortium, Posters and Demos of the 24th International Semantic Web Conference (ISWC-C 2025), volume 4085 of *CEUR Workshop Proceedings*, CEUR, Nara, Japan, 2025, pp. 395–401. URL: https://ceur-ws.org/Vol-4085/#paper66, iSSN: 1613-0073.

[6] S. Evenstein Sigalov, R. Nachmias, Investigating the potential of the semantic web for education: Exploring Wikidata as a learning platform, Education and Information Technologies 28 (2023) 12565–12614. URL: https://doi.org/10.1007/s10639-023-11664-1. doi:10.1007/s10639-023-11664-1.

[7] T. Burrows, L. Cleaver, D. Emery, E. Hyvönen, M. Koho, L. Ransom, E. Thomson, H. Wijsman, Medieval Manuscripts and Their Migrations: Using SPARQL to Investigate the Research Potential of an Aggregated Knowledge Graph, Digital Medievalist 15 (2022). URL: https://journal.digitalmedievalist.org/article/id/8064/. doi:10.16995/dm.8064.

[8] E. Hyvönen, Digital Humanities on the Semantic Web: from Infrastructure to Practical Applications, AI-Based Knowledge Discovery, and Web of Wisdom, in: W.-T. Balke, K. Golub, Y. Manolopoulos, K. Stefanidis, Z. Zhang (Eds.), Linking Theory and Practice of Digital Libraries, Springer Nature Switzerland, Cham, 2025, pp. 3–8. doi:10.1007/978-3-032-05409-8_1.

[9] G. Lodi, L. Asprino, A. G. Nuzzolese, V. Presutti, A. Gangemi, D. R. Recupero, C. Veninata, A. Orsini, Semantic Web for Cultural Heritage Valorisation, in: S. Hai-Jew (Ed.), Data Analytics in Digital Humanities, Springer International Publishing, Cham, 2017, pp. 3–37. URL: https://doi.org/10.1007/978-3-319-54499-1_1.

[10] J. A. Raemy, Linked Open Usable Data for cultural heritage: perspectives on community practices and semantic interoperability, Ph.D. thesis, University of Basel, 2024. URL: https://hdl.handle.net/20.500.14716/148352.

[11] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data 3 (2016) 160018. URL: https://www.nature.com/articles/sdata201618. doi:10.1038/sdata.2016.18, publisher: Nature Publishing Group.

[12] T. Battisti, M. Daquino, Exploring data-driven narratives in Digital Humanities web-based projects: features and impact, in: S. Rebora, M. Rospocher, S. Bazzaco (Eds.), Diversità, Equità e Inclusione: Sfide e Opportunità per l'Informatica Umanistica nell'Era dell'Intelligenza Artificiale, AIUCD, Verona, 2025, pp. 18–23. doi:https://doi.org/10.6092/unibo/amsacta/8380.

[13] P. Warren, P. Mulholland, Using SPARQL – The Practitioners' Viewpoint, in: C. Faron Zucker, C. Ghidini, A. Napoli, Y. Toussaint (Eds.), Knowledge Engineering and Knowledge Management, Springer International Publishing, Cham, 2018, pp. 485–500. doi:10.1007/978-3-030-03667-6_31.

[14] J. Lian, Y. Zhao, X. Li, Q. Zhu, An Investigation of Searching Behavior for Open Datasets: Insights from Cultural Heritage Data Curation Competitions, Proceedings of the Association for Information Science and Technology 62 (2025) 1007–1012. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/pra2.1330. doi:10.1002/pra2.1330, _eprint: https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/pra2.1330.

[15] F. Windhager, P. Federico, G. Schreder, K. Glinka, M. Dörk, S. Miksch, E. Mayr, Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges, IEEE Transactions

on Visualization and Computer Graphics 25 (2019) 2311–2330. URL: https://ieeexplore.ieee.org/document/8352050. doi:10.1109/TVCG.2018.2830759.

[16] A. Hogan, The Semantic Web: Two decades on, Semantic Web 11 (2020) 169–185. doi:10.3233/SW-190387.

[17] C. Buil-Aranda, M. Ugarte, M. Arenas, M. Dumontier, A Preliminary Investigation into SPARQL Query Complexity and Federation in Bio2RDF, in: A. Calì, M.-E. Vidal (Eds.), Proceedings of the 9th Alberto Mendelzon International Workshop on Foundations of Data Management, Lima, Peru, May 6 - 8, 2015, volume 1378 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2015. URL: http://ceur-ws.org/Vol-1378/AMW_2015_paper_37.pdf.

[18] V. Cavaller, Dimensional Taxonomy of Data Visualization: A Proposal From Communication Sciences Tackling Complexity, Frontiers in Research Metrics and Analytics 6 (2021). URL: https://www.frontiersin.org/journals/research-metrics-and-analytics/articles/10.3389/frma.2021.643533/full. doi:10.3389/frma.2021.643533, publisher: Frontiers.

[19] A. Thudt, J. Walny, T. Gschwandtner, J. Dykes, J. Stasko, Exploration and Explanation in Data-Driven Storytelling, in: Data-Driven Storytelling, A K Peters/CRC Press, 2018. Num Pages: 25.

[20] K. Lin, S. S.-t. Ru, D. N. Rapp, H. Guan, C. Xiong Bearfield, What Makes a Visualization Visually Complex?, in: Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, CHI EA '25, Association for Computing Machinery, New York, NY, USA, 2025, pp. 1–7. URL: https://doi.org/10.1145/3706599.3719983. doi:10.1145/3706599.3719983.

[21] E. M. Champion, Digital humanities is text heavy, visualization light, and simulation poor, Digital Scholarship in the Humanities 32 (2017) i25–i32. URL: https://doi.org/10.1093/llc/fqw053. doi:10.1093/llc/fqw053.

[22] G. Renda, Data affordances in Digital Humanities information visualisation practices — Reproducible Analysis, 2025. URL: https://zenodo.org/records/17910955. doi:10.5281/zenodo.17910955, publisher: Zenodo.

[23] S. Im, D. Park, E. Kim, J. Lee, Development and comparison of a prototype for visualizing artifact information for museum professionals, Information Visualization 24 (2025) 246–268. URL: https://doi.org/10.1177/14738716251324639. doi:10.1177/14738716251324639, publisher: SAGE Publications.

[24] A. Velhinho, M. Alves, P. Almeida, L. Pedro, Designing a Collaborative Storytelling Platform to Enrich Digital Cultural Heritage Archives and Collective Memory, in: A. Marcus, E. Rosenzweig, M. M. Soares (Eds.), Design, User Experience, and Usability, Springer Nature Switzerland, Cham, 2024, pp. 142–159. doi:10.1007/978-3-031-61351-7_10.