Exploring data-driven narratives in Digital Humanities web-based projects: features and impact

Tommaso Battisti[1], Marilena Daquino[2]
[1] University of Bologna, Italy - tommaso.battisti5@unibo.it
[2] University of Bologna, Italy - marilena.daquino2@unibo.it

ABSTRACT (ENGLISH)
Narrative approaches as an aid to information visualisation practices help to present complex information clearly and enable communication to a broader audience, serving as an interpretative framework with positive outcomes in the Cultural Heritage domain. Several studies discuss narrative design spaces for information visualisation, producing different taxonomies and classifications. Nonetheless, existing contributions mainly relate to data journalism or domain-agnostic projects, while surveys accounting for narrativity in the Digital Humanities are restricted to specific contexts and limited to the surface of these aspects. This work extends previous contributions by addressing narrative design choices in 186 web-based Digital Humanities projects that leverage information visualisation techniques. Projects are classified according to narrativity and narrative strategies, the humanities domain, and visualisation type. Findings show limited use of narrative approaches with a strong reliance on the use of storytelling tools and frequent homogeneous and similar solutions, while humanities domains show interesting and diverging patterns.
Keywords: Data-driven narratives; Information visualisation; Narrative design spaces; Survey

ABSTRACT (ITALIANO)
Esplorare le narrazioni guidate dai dati in progetti di Digital Humanities sul web: caratteristiche e impatto.
Gli approcci narrativi come supporto alle pratiche di visualizzazione dell'informazione aiutano a presentare informazioni complesse in modo chiaro e permettono la comunicazione a un pubblico più ampio, fungendo da quadro interpretativo con esiti positivi nel campo del Patrimonio Culturale. Diversi studi discutono aspetti di progettazione narrativa per la visualizzazione dell'informazione, producendo diverse tassonomie e classificazioni. Tuttavia, i contributi esistenti si riferiscono principalmente al campo del data journalism o a progetti senza relazione con domini specifici, mentre le indagini che tengono conto della narratività nelle Digital Humanities sono limitate a contesti precisi e affrontano solo superficialmente questi aspetti. Questo studio amplia i contributi precedenti analizzando le scelte di progettazione narrativa in 186 progetti di Digital Humanities che utilizzano tecniche di visualizzazione dell'informazione sul web. I progetti sono classificati in base alla narratività e alle strategie narrative, al dominio umanistico e al tipo di visualizzazione. I risultati mostrano un utilizzo limitato di approcci narrativi, con una forte dipendenza dall'uso di strumenti di storytelling e soluzioni frequentemente simili e omogenee, mentre i domini umanistici mostrano modelli interessanti e divergenti.
Parole chiave: Narrazioni guidate dai dati; Visualizzazione dell'informazione; Progettazione narrativa; Indagine

1. INTRODUCTION
Information visualisation offers significant opportunities to present and investigate complex information within the field of Digital Humanities (DH). Visual representations supported by computation tools and interaction are key for amplifying cognition (Card et al., 1999). In this context, the Web represents a powerful means to enhance data presentation and knowledge dissemination. Meanwhile, storytelling emerges as an additional aid to convey information more efficiently (Gershon & Page, 2001), with practical advantages for the Cultural Heritage domain, such as improving the presentation, exploration, and interpretation of complex information (Renda et al., 2023) and shaping stronger connections with the audiences, enhancing user experience and dissemination (Shan et al., 2022).
Several studies contributed to the definition of design spaces for data-driven narratives. Among these, (Segel & Heer, 2010) define different narrative genres, tactics, and structures, distinguishing author-driven and reader-driven approaches, with hybrid potentials in the middle. Similar contributions and further extensions followed (Hullman & Diakopoulos, 2011; Stolper et al., 2016; Roth, 2021; Zhao & Elmqvist, 2023). From a different perspective, (McKenna et al., 2017) defined and studied factors involved in the data-driven narratives' flow, capable of affecting the reading experience. Most of these studies

mainly surveyed diverse online sources, generally related to data journalism or more domain-agnostic works.
In the DH, various contributions propose solutions exploiting explorative and narrative approaches, which encompass generous interfaces (Whitelaw, 2015), workflows and platforms for the creation of Cultural Heritage visualisation-based stories (Liem et al., 2023), and authoring tools for linked open data-driven stories (Renda et al., 2023). Nonetheless, only a few surveys on DH projects exist and describe the application of information visualisation techniques in the field, while narrative aspects are even rarely considered. These are also limited to specific sub-contexts of applications such as Cultural Heritage collection interfaces (Windhager et al., 2019), text analysis support tasks (Jänicke et al., 2017), musicology works (Khulusi et al., 2020), and DH publications using visualisations as part of the research process (Panagiotidou et al., 2023). While having diverse scopes, interestingly, only (Windhager et al., 2019) address narrativity, limiting the analysis to an account of narrative approaches intended as "curated paths", without surveying deeper narrative design choices.
The purpose of this study is to broaden the scope of previous surveys encompassing the wider spectrum of the DH. In addition, it seeks to compensate for the lack of comprehensive reports on data-driven narrative design choices so far unexplored within the field.

2. METHODS

The present work analyses a corpus of 186 web-based DH projects leveraging information visualisation techniques with a particular focus on narrative strategies.
The corpus was developed by relying on diverse sources: 1) the existing classification by (Windhager et al., 2019) and related authors of interfaces, which were secondly surveyed to find additional material (33 projects); 2) the "projects" sections of the websites of the Italian (AIUCD)[1] and European (EADH)[2] DH associations, extending the inquiry to related research centres and institutions (104 projects); and 3) the "Best DH data visualisation" category of the Digital Humanities Awards[3] website (49 projects).
Selection criteria to include projects in the survey are the following: 1) must be web-based dissemination projects; 2) outcomes are potentially produced through visualisation tools or storytelling software but are not tools or software themselves; 3) free from access barriers like registration forms; 4) significantly reliant on visualisation techniques; and 5) accessible at the time of assessment.
The 186 projects composing the final corpus were classified through direct observation and interaction with the interfaces, evaluating only visual solutions and not accounting for any kind of textual description. This empirical approach minimises potential bias deriving from inconsistencies between documentation and available features of the systems during evaluation.
Our final classification framework reuses different taxonomies and schemas that were selected and adapted from literature. It was developed and refined using an iterative process inspired by (Segel & Heer, 2010) and consisting of testing different column configurations over samples of projects during classification until a suitable result was achieved, according to which we re-classified all instances. Even if two annotators evaluated the final dataset by reviewing a sample of instances, as recognised by (Segel & Heer, 2010), using a custom taxonomy over diverse design dimensions inevitably introduces some degree of subjectivity.
Features relevant to the current study include:
Narrativity. To identify projects incorporating information visualisation within narratives, whether linear author-driven or user-driven (Segel & Heer, 2010). Hybrid projects combine both narrative and non-narrative approaches.
Domain. Building on (Terras et al., 2016: 138) and (Schreibman et al., 2004), a categorical variable distinguishes the humanities field of the project, including: 1) history and archaeology 2) art and art history 3) language and literature, including linguistics, philology, narrative and literary studies 4) music and musicology 5) multimedia and performing arts 6) philosophy and religion, and 7) other, for other domains or collections with no unique focus.
Visualisation techniques. Adapted from (Windhager et al., 2019), a Boolean value indicates the presence of the following visualisation types: 1) plot 2) cluster or set 3) map, including a further identification for statistical symbol maps (i.e., data points are statistical charts) 4) network 5) hierarchical

[1] Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD): https://www.aiucd.it/progetti/.

[2] European Association for Digital Humanities (EADH): https://eadh.org/projects.

[3] http://dhawards.org/.

diagram 6) treemap 7) word cloud 8) bars, any bar-based chart; 9) line chart 10) area chart 11) pie chart 12) 3D plot 13) proportional area 16) timeline, and 16) other, miscellaneous. Each column corresponds to its type and includes also the related stacked layouts and variations (e.g., a bar chart and a radial bar chart).

Narrative flow factors. A set of columns adapted from (McKenna et al., 2017) identifies relevant patterns in the design of narratives that involve: 1) story layout 2) visualisation role 3) story progression 4) navigation input 5) navigation progress, and 6) level of control over text, visualisations, and animated transitions. It is worth noticing that projects with multiple narratives can potentially report multiple design choices for the same flow factor. Moreover, while we use the term narrative since we refer to both authorial linear data stories and more user-directed narratives, the term "story" was used for factor names to match the original taxonomy.

The final dataset (Battisti, 2024) was analysed via a Jupyter Notebook (Battisti, 2025) to ensure reproducibility, while the following research questions guided our analysis:

- **RQ1**. How many projects use narrative approaches to information visualisation?
- **RQ2**. How can we characterise data-driven narratives in DH projects according to visualisation use and narrative flow factors?
- **RQ3**. How do humanities domains use narrativity?

3. RESULTS AND DISCUSSION

Overall, DH dissemination projects present only limited use of narrative approaches, the emergence of patterns linked to humanities domains, but also uncritical reuse of storytelling tools without additional customisation, impeding further valorisation of research material's unique characteristics.

RQ1. The great majority of the projects in our corpus are non-narrative (76,9%;143/186). Most of the remaining 23,1% of projects containing narratives rely on the complementary strengths of both narrative and non-narrative sections (12,9%; 24/186), while fully narrative projects are only 10,2% of the total (19/186). Narrativity is slightly higher than in (Windhager et al., 2019) survey on Cultural Heritage collection interfaces, but the count is still limited compared to the communication benefits that narrative approaches are claimed to bring to information visualisation (Gershon & Page, 2001; Renda et al., 2023), especially considering their role for dissemination.

RQ2. Considering visualisation use, the narrative context causes differences in visualisation frequency rankings. Most relevant, it widens the gap between maps—the most frequent solution, employed by 65,5% of the projects (122/186)—and networks, adopted by 37,6% of projects overall (70/186). In projects using at least a narrative element, the prevalence of maps rises to 87%, and the adoption of networks falls to 18,6%. The latter's underrepresentation may be an echo of limitations related to readability in dense configurations and connections' inadequacies in representing data measures compared to other visual variables (Koponen & Hildén, 2019). Nonetheless, narrative approaches and guided explorations could serve as an aid to enhance network exploration and analysis (Li et al., 2023).

Moreover, while most non-narrative projects (54,6%; 78/143) employ multiple visualisations, the majority of fully narrative projects rely on a single technique (57,9%; 11/19). Due to the presence of at least two sections (one narrative and one non-narrative) almost all hybrid projects use multiple visualisations (87,5%; 21/24). Concerning flow factors, narratives in DH dissemination projects show a particular preference. They are almost always linear and author-driven (95,4%; 41/43), with a preference for slideshow layouts (44,2%; 19/43) followed by document layouts (37,2%; 16/43), two projects with more narrative sections using different layouts, and six using hybrids or more complex formats. Despite slideshows being more frequent, document layouts account for more visualisations. Among the few exceptions, we found timelines (five occurrences in slideshows and three in documents) and maps (18 in slideshows and 10 in documents). The latter is also the most used visualisation in slideshow-document hybrid layouts.

Narrative progress is assigned to a button or click approaches 44,2% of the time (19/43) while 32,6% use the scroll (14/43) reflecting the division between slideshow and document layouts. 20,9% use multiple techniques (9/43), while only one uses a slider approach. Narrative progression is communicated by the visualisation itself in 25,6% of projects (11/43), followed by textual elements (11,6%; 5/43), multiple techniques (7%; 3/43) and only one by dots. Interestingly, most projects do not communicate user position within the narrative (53,5%; 23/43). The problem affects mainly document layouts that rely on the webpage scrollbar as an implicit signifier for user position. However, since a webpage may not reflect

the actual length of a narrative, the practice is in contrast with the idea that signifiers and feedback should ideally match user expectations to avoid misinterpretations during interaction (Norman, 2013).
Regarding the control of narrative elements, discrete and continuous text control is equally employed, with a project using them both. Differently, the majority of control type over visualisations is continuous (72,1%; 31/43), while discrete control affects most animations (65,1%; 28/43). 23,3% of projects (10/43) have no animation control.
In most narratives, visualisations and texts play an equal role in telling stories (53,5%; 23/43). 18,6% of narratives entrust visualisations with the main role (8/43), while only 11,6% of projects use them as support for text (5/43). The remaining projects use multiple solutions. In relation to the visualisation, maps and networks are mostly used with an equal role (71%; 22/31 and 62,5%; 5/8 respectively) and only 1/11 of timelines have a subordinate role. It is worth noticing that we excluded six projects containing both multiple roles and visualisations that impeded correct tracking between the two variables.
Relevant patterns also include 84.2% of slideshows with equal role visualisations (16/19), which correspond to 69.6% of equal role visualisations used in slideshows (16/23); 50% of document layout narratives with a subordinate role for visualisations (8/16), corresponding to the use of all the subordinated visualisations within document narratives (8/8); and all the other layouts without category providing visualisations with a driving role (2/2), which correspond to 40% of main-role visualisations used in other non-categorised layouts. Other patterns involving the layout include 94.7% of slideshows using button/click input (18/19) corresponding to the same amount also by shifting perspective; 87.5% of documents using scroll input (14/16) corresponding to the use of all scroll inputs within document layouts (14/14); and all document-slideshow hybrids using hybrid scroll-click navigation (4/4).
The most common control patterns are discrete text, continuous visualisation, and continuous animation control (37,2%; 16/43), which is always used by slideshows; and continuous text, continuous visualisation, and no animation control (20,9%; 9/43), which is always used by document layouts.
Association rules generated through the Apriori algorithm[4] also underline the connection between document layout, scroll input, continuous text control, and slideshows, as well as discrete text control, button input, and visualisation's equal role.
Overall, the preference for single visualisation techniques in fully narrative projects, the higher number of slideshows and their use of most maps and timelines, as well as more complex patterns involving narrative flow factors witness the frequent use of storytelling tools such as StoryMapJS[5] or TimelineJS[6]. Such tools typically rely on a single visualisation technique and are employed for most narratives using slideshows (66,7%; 12/18)[7]. On the one hand, these tools offer a technically convenient way to develop narratives for dissemination projects. However, they almost always lack further customisation, failing to adapt to the specificities of the research material and producing always similar and homogeneous results that inhibit the valorisation of the uniqueness of the working data.
RQ3. Overall, two domains accommodate most projects: "language and literature" (39,8%; 74/186) and "history and archaeology" (34,9%; 65/186). Given the greatly higher number of non-narrative projects, all domains present a preference for non-narrative approaches. However, by shifting perspective, while "language and literature" is the most popular domain among non-narrative projects (43,3%; 62/143) followed by "history and archaeology" (30,8%; 44/143), the latter category is the most present within narrative projects (48,8%; 21/43), with "language and literature" following at 27,9% (12/43). The gap enlarges considering only fully narrative projects, where "history and archaeology" count 57,9% of projects (11/19) and "language and literature" only 21,1% (4/19).
The narrative characterisation of the domains also comes with related and already discussed patterns. Therefore, while all domains use multiple visualisations—except for "music and musicology" which includes one single- and one multi-visualisation project— "history and archaeology" shows a more balanced approach (33/65 multi-visualisation to 32/65 single-visualisation). Noticeably it also accounts for most narrative projects, and a higher use of maps, including nearly half of all map occurrences in the dataset (46,7%; 57/122).

[4] The documentation of the Python algorithm used during the analysis can be accessed at https://rasbt.github.io/mlxtend/api_subpackages/mlxtend.frequent_patterns/.

[5] https://storymap.knightlab.com/.

[6] https://timeline.knightlab.com/.

[7] The number refers to narrative projects only. Nonetheless, these tools are often used also in non-narrative contexts to present items according to relevant metadata.

The divergent preferences corresponding to the two domains reflect distinct dissemination approaches, suggesting that disciplinary characteristics shape visualisation strategies and project design.
Limitations. Given the dimension of our corpus and the reliance on selected sources, we claim to have derived representative conclusions about the impact of narrativity in DH web-based dissemination projects. Nonetheless, the analysis of flow factors and visualisation use in narrative contexts suffer from the limited dimension of the narrative subset, which leads to interesting consideration but no further generalisable results. A similar line of reasoning involves the analysis of domains and their narrative characterisation since it was possible to derive conclusions only for the most frequent categories. Furthermore, the analysis focuses on projects rather than on single narrative sections leading to the impossibility of precisely understanding whether specific visualisations are used in narrative or non-narrative contexts within hybrid projects. Finally, also the presence of multiple narrative sections within single projects led to the impossibility of tracking visualisation roles when occurring both multiple visualisations and roles.

4. CONCLUSIONS

The present study analysed narrative practices in DH web-based dissemination projects leveraging information visualisation techniques, assessing the number of narrative approaches, how these vary according to the humanities domain, the use of visualisations in different narrative contexts, and narrative design patterns involving narrative flow factors. We extended both the context and the focus of previous humanities-related studies that almost never and only partially surveyed narrative aspects related to information visualisation techniques. The work also serves as a point of departure for further critical reflection on the role of storytelling tools and the relation between technical needs for simple and reusable solutions, and custom possibilities to enhance and valorise research material.

ACKNOWLEDGEMENTS

Project funded by the European Union – NextGenerationEU under the National Recovery and Resilience Plan (NRRP), Investment I.4.1 - Borse PNRR Patrimonio Culturale.

REFERENCES

Battisti, T. (2024). Classification of web-based Digital Humanities projects leveraging information visualisation techniques [Dataset]. Zenodo. https://doi.org/10.5281/ZENODO.14192758

Battisti, T. (2025). Data-driven insights into web-based Digital Humanities projects leveraging information visualization techniques: An analysis of the 'Classification of web-based Digital Humanities projects leveraging information visualisation techniques' dataset [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.14824189

Card, S. K., Mackinlay, J., & Shneiderman, B. (1999). Readings in information visualization: using vision to think. Morgan Kaufmann.

Gershon, N., & Page, W. (2001). What storytelling can do for information visualization. Communications of the ACM, 44(8), 31–37. https://doi.org/10.1145/381641.381653

Hullman, J., & Diakopoulos, N. (2011). Visualization Rhetoric: Framing Effects in Narrative Visualization. IEEE Transactions on Visualization and Computer Graphics, 17(12), 2231–2240. https://doi.org/10.1109/TVCG.2011.255

Jänicke, S., Franzini, G., Cheema, M. F., & Scheuermann, G. (2017). Visual Text Analysis in Digital Humanities. Computer Graphics Forum, 36(6), 226–250. https://doi.org/10.1111/cgf.12873

Khulusi, R., Kusnick, J., Meinecke, C., Gillmann, C., Focht, J., & Jänicke, S. (2020). A Survey on Visualizations for Musical Data. Computer Graphics Forum, 39(6), 82–110. https://doi.org/10.1111/cgf.13905

Koponen, J., & Hildén, J. (2019). Data visualization handbook (first edition). Otavan kirjapaino.

Li, W., Schöttler, S., Scott-Brown, J., Wang, Y., Chen, S., Qu, H., & Bach, B. (2023). NetworkNarratives: Data Tours for Visual Network Exploration and Analysis. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 1–15. https://doi.org/10.1145/3544548.3581452

Liem, J., Kusnick, J., Beck, S., Windhager, F., & Mayr, E. (2023). A Workflow Approach to Visualization-Based Storytelling with Cultural Heritage Data. 2023 IEEE 8th Workshop on Visualization for the Digital Humanities (VIS4DH), 13–17. https://doi.org/10.1109/VIS4DH60378.2023.00008

McKenna, S., Henry Riche, N., Lee, B., Boy, J., & Meyer, M. (2017). Visual Narrative Flow: Exploring Factors Shaping Data Visualization Story Reading Experiences. Computer Graphics Forum, 36(3), 377–387. https://doi.org/10.1111/cgf.13195

Norman, D. A. (2013). The design of everyday things (Revised and expanded ed). Basic Books.

Panagiotidou, G., Lamqaddam, H., Poblome, J., Brosens, K., Verbert, K., & Vande Moere, A. (2023). Communicating Uncertainty in Digital Humanities Visualization Research. IEEE Transactions on Visualization and Computer Graphics, 29(1), 635–645. https://doi.org/10.1109/TVCG.2022.3209436

Renda, G., Daquino, M., & Presutti, V. (2023). Melody: A Platform for Linked Open Data Visualisation and Curated Storytelling. Proceedings of the 34th ACM Conference on Hypertext and Social Media, 1–8. https://doi.org/10.1145/3603163.3609035

Roth, R. E. (2021). Cartographic Design as Visual Storytelling: Synthesis and Review of Map-Based Narratives, Genres, and Tropes. The Cartographic Journal, 58(1), 83–114. https://doi.org/10.1080/00087041.2019.1633103

Schreibman, S., Siemens, R. G., & Unsworth, J. (Eds.). (2004). A companion to digital humanities. Blackwell Pub.

Segel, E., & Heer, J. (2010). Narrative Visualization: Telling Stories with Data. IEEE Transactions on Visualization and Computer Graphics, 16(6), 1139–1148. https://doi.org/10.1109/TVCG.2010.179

Shan, X., Wang, D., & Li, J. (2022). Research on Data Storytelling Strategies for Cultural Heritage Transmission and Dissemination. In C. Stephanidis, M. Antona, S. Ntoa, & G. Salvendy (Eds.), HCI International 2022 – Late Breaking Posters (pp. 344–353). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-19679-9_43

Stolper, C. D., Lee, B., Riche, N. H., & Stasko, J. (2016). Emerging and recurring data-driven storytelling techniques: Analysis of a curated collection of recent stories. (2016).

Terras, M., Nyhan, J., & Vanhoutte, E. (2016). Defining Digital Humanities: A Reader. Routledge.

Whitelaw, M. (2015). Generous Interfaces for Digital Cultural Collections. Digital Humanities Quarterly, 9(1), 1–16.

Windhager, F., Federico, P., Schreder, G., Glinka, K., Dork, M., Miksch, S., & Mayr, E. (2019). Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges. IEEE Transactions on Visualization and Computer Graphics, 25(6), 2311–2330. https://doi.org/10.1109/TVCG.2018.2830759

Zhao, Z., & Elmqvist, N. (2023). The Stories We Tell About Data: Surveying Data-Driven Storytelling Using Visualization. IEEE Computer Graphics and Applications, 43(4), 97–110. https://doi.org/10.1109/MCG.2023.3269850