CuratedDDS: A Taxonomy and a Dataset of Data-Driven Stories to Support Journalists' Inspiration

Louri G. K. Compain
louri.compain@gmail.com
Polytechnique Montreal
Montréal, Québec, Canada

Thomas Hurtut
thomas.hurtut@polymtl.ca
Polytechnique Montreal
Montréal, Québec, Canada

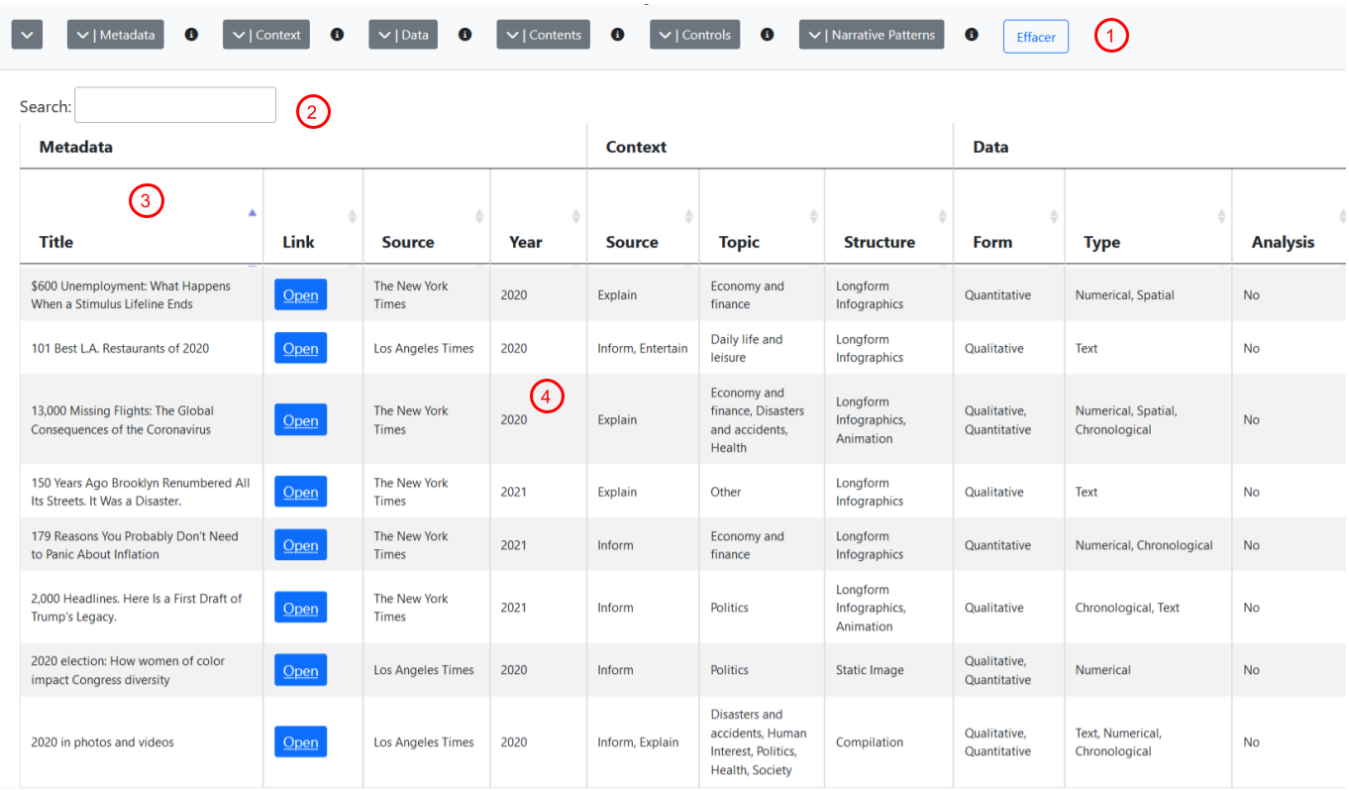

Figure 1: Interface of CuratedDDS, a tool designed to enhance Data Driven Stories (DDS) ideation through a curated archive of examples annotated according to a specialized taxonomy. This interface facilitates exploration, discoverability and selection via several features: (1) Dropdown menus in the top bar display taxonomy elements as responsive filters; (2) A search bar enables keyword exploration; (3) A dual-level table header correlates taxonomy elements with annotations for clarity, and reorganization; (4) The main table visually presents the examples alongside their annotations, allowing for easy access and analysis.

Abstract

So-called Data Driven Stories (DDS) are an increasingly popular online narrative format that combines text, media such as photos, and data visualizations. This format engages readers effectively due to its visual appeal and the trustworthiness conferred by data journalism. However, compared to more traditional formats, DDS require an thorough ideation and design process, about its narrative structure, and the ancitipated visualizations, which brings up additional challenges. In this paper, we present CurratedDDS, a tool developed to support DDS designers by offering a curated archive of DDS examples, annotated using a proposed DDS taxonomy, and an exploratory interface. To evaluate the efficiency and impact of CurratedDDS for the target audience, we conducted an evaluation in the form of a case study with a group of 6 media professionals.

CFP '24, October 25-27, 2024, Boston, MA

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

CCS Concepts

• **Applied computing** → *Document searching*; • **Human-centered computing** → Visualization design and evaluation methods; • **Information systems** → *Content analysis and feature selection*; **Multimedia information systems**.

Keywords

Communication, Data Driven Stories,Data Journalism, Exemples, Ideation, Inspiration, Workflow Design, Journalism

ACM Reference Format:

Louri G. K. Compain and Thomas Hurtut. 2018. CuratedDDS: A Taxonomy and a Dataset of Data-Driven Stories to Support Journalists' Inspiration. In *Proceedings of C+J 2024: The 2024 Computation + Journalism Symposium (CFP '24).* ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

The evolution of online journalism has facilitated the emergence of Data Driven Stories (DDS), which are news articles that prioritize data collection, analysis, and visualization over traditional media such as text and photos. DDS add analytical depth and increase audience engagement through visually compelling narratives that attract more attention than text-heavy articles [6, 21]. Narratives in DDS not only engage audiences more deeply but also foster personal connections and prompt action more effectively than bare statistics [4]. Additionally, DDS builds trust, aids memory, enhances pattern recognition, and simplifies complex information [5]. However, a significant challenge for DDS designers is the additional workload from creating narrative formats, including both the embedded data visualizations, interactions, and overall narrative structure of those stories. The ideation phase is thus a crucial component of this design process.

Access to examples of existing DDS articles is a vital resource for ideation. Examples can serve as a means of team communication, help identify common designs or new ideas, and aid in understanding the target audience [3]. Additionally, they promote more diverse and creative ideas [4]. However, accessing examples is not always straightforward. Designers use a range of tools from general creative design to open image publication, indicating a lack of specialized tools for finding DDS inspiration [22]. Our initial discussions with two Canadian media outlets confirmed this observation and highlighted a significant challenge in example search: recalling and locating previously consulted media from memory.

In this paper, we introduce CurratedDDS, a specialized tool designed to support DDS journalists by providing access to an archive of examples selected and annotated according to a proposed taxonomy of DDS content and objectives. We evaluate CuratedDDS by conducting a case study on a panel of 6 journalists. Our contributions are threefold: A novel taxonomy for DDS, derived from the adaptation and expansion of several existing taxonomies; A dataset of 780 online DDS articles, selected and annotated in accordance with our taxonomy; An interactive tool, available online, that facilitates the exploration of this dataset (see Figure 1). sectionRelated work This section first reviews the existing works dedicated to the classification and caracterization of DDSs and then discusses the existing literature concerning ideation and examples in design, especially for DDSs.

1.1 Classifying DDS

Several taxonomies have been developed to characterize Data-Driven Stories (DDS) across various contexts [1, 2, 11, 13, 15–20, 23]. These taxonomies offer different perspectives, ranging from data comics and live presentations to more focused subsets like Storymaps, utilizing approaches from narrative devices to quantifying text and visuals. Most taxonomies contain 4 to 20 dimensions; however, one extends to 75 dimensions, with the majority featuring fewer than 10 dimensions. The foundational taxonomy by Segel [16] from 2010 is frequently cited, with several others building on its concepts such as reader-driven and author-driven storytelling and the narrative visualization genre.

Parsons [12] highlights a disconnect between academic research and practical application in data visualization, noting that academic guidelines are often overlooked by practitioners. The identified taxonomies are divided into those aimed at future research or tool development for practitioners and those intended for direct reference by practitioners. In our approach, we align with the first group, aiming to develop a tool that incorporates taxonomy elements for practitioner use.

1.2 Consuming DDS Examples During Ideation Phase

When designing Data-Driven Stories (DDS), ideation is essential, particularly for integrating visual elements. Effective DDS design begins with a clear understanding of the problem, which guides the visualization process [14]. This preliminary phase requires creativity in defining problems and devising visual components, as designers seek innovative approaches [10]. They often utilize examples from previous projects as ideation tools, drawing inspiration to form a domain-specific collection of techniques and styles [3, 9]. Designers access examples through various platforms, including specialized websites, blogs, and broader platforms like Pinterest and Google. However, these resources typically lack the capability to filter or search using specialized technical terminology [3, 7, 8]. This limitation can impede the ability to find precisely relevant examples.

2 Proposed Methodology

In this paper, we introduce a methodology designed to assist journalists during the ideation phase of their work with Data-Driven Stories (DDS). This methodology consists of three core elements: a novel taxonomy for DDS, a curated archive of DDS examples annotated according to this taxonomy, and a tool that enables users to effectively explore the archive.

2.1 Taxonomy

The taxonomy we propose integrates insights from eleven prior taxonomies [1, 2, 11, 13, 15–20, 23], extracting a comprehensive list of 155 dimensions, including 75 specifically from a taxonomy focusing on storymaps [1].

We refined this extensive list to 25 dimensions that are applicable to DDS and allow for categorization, excluding those only relevant

for specific contexts like cartography or live presentations. The list was further streamlined by merging overlapping dimensions, reducing the total to 20. To enhance usability, we added titles and sources as identifiers and adjusted certain terminologies for clarity, such as renaming "interactive cues" to "tutorial." After organizing these into six main groups to improve navigability, the finalized taxonomy now comprises 22 dimensions, systematically presented in Table 1.

2.2 Curated Archive

The archive was constructed by selecting DDS examples from the annual best compilations of major online newspapers, initially focusing on the most recent three years—2020 to 2022—as a manageable scope that also reflected the latest advancements in DDS quality. Initially, we gathered 9 newspapers, 55 compilations, and 2,891 examples. Given the extensive nature of this collection, we streamlined our efforts to 6 newspapers and 14 compilations, totaling 780 examples, which took about 300 hours to annotate. This selection not only managed the annotation workload but also ensured good coverage across all taxonomy dimensions. We chose a reverse chronological order to prioritize recent improvements in DDS production and to maximize the dataset's sustainability. The comprehensive list of sources and annotations are detailed in Table 2. This curated set represents a strong foundation, but there is potential to further enrich the dataset with ongoing annotations.

2.3 CurratedDDS

The proposed interface, CuratedDD[1], is designed to help users navigate the archive efficiently using the developed taxonomy as a filtering mechanism. Each element of the taxonomy is presented in a user-friendly form within the interface, where users can select specific values to filter the examples that correspond to these criteria. To reduce user effort and enhance interaction, modifications to the filter settings trigger an automatic query, updating the displayed results instantaneously.

To address the complexity and enhance the clarity of the taxonomy, which includes 6 groups, 22 dimensions, and 109 values, we implemented dropdown menus for each group. Each dimension within a group can be expanded or collapsed to reveal or hide its respective values. This design choice helps manage screen space and prevent information overload. Pop-ups are used to provide definitions and clarify the more complex or ambiguous terms within the taxonomy, ensuring users understand the filtering options.

The filtered results are displayed in a detailed table format. Given the extensive data involved, the table extends horizontally beyond the typical screen width. To maintain readability and ease of use, we incorporated a horizontal scroll feature. This allows users to view all data without overwhelming them with a cluttered or densely packed interface, facilitating a clear and comprehensive overview of the filtered information.

[1]CuratedDDS and the archive are accessible at this URL

3 Evaluation

3.1 Protocol

We recruited six volunteers from two French-speaking Canadian media outlets, encompassing data journalists, software engineers, and web developers, all of whom regularly design DDSs as part of their work. We conducted a case study where participants used CuratedDDS over two weeks within their usual routines[2]. Initially, we discussed potential uses for CuratedDDS, including searching for project examples and broadening their professional knowledge. After the trial, we conducted semi-structured interviews to gather qualitative feedback from the participants.

3.2 Results and Discussion

Taxonomy. Participants appreciated the depth of the taxonomy, particularly valuing metadata, context, data, and content over controls and narrative patterns. This deeper focus helped them utilize broad filtering criteria effectively, complemented by their expert judgment for precise selections. Although the comprehensive structure—comprising 6 groups, 20 dimensions, and 109 values—was initially a bit overwhelming, it was acknowledged as a valuable resource for in-depth example descriptions and enhanced comprehension.

Vocabulary mismatches also surfaced; terms used in the taxonomy sometimes diverged from the language familiar to practitioners, necessitating the use of explanatory pop-ups in CuratedDDS, which, while helpful, added an extra layer of effort. One participant highlighted that while most of their colleagues are journalists who have developed a keen interest in data, only a few have formal training in data journalism. This leads to many practitioners encountering concepts in ways that differ from academic definitions, potentially creating a disconnect in understanding.

Archive. The archive received positive reviews for its diversity and the relevance of its content, with participants valuing the inclusion of examples from sources they trust and regularly consult. This resource helped them uncover previously unnoticed DDS examples from familiar media outlets, enhancing their understanding and appreciation of available content. However, the reliance on predominantly US sources was noted as a limitation, with participants expressing a need for more linguistic and cultural diversity to better reflect the global context of news media. Suggestions for more frequent updates were made to keep the archive timely and reflective of the latest developments in data visualization.

CurratedDDS. Five out of six participants observed that CuratedDDS enhanced the accessibility of examples and simplified the search process compared to their conventional methods. Notably, four participants found that using CuratedDDS altered their approach to searching for examples in a positive way, leading them to discover new and diverse examples. Additionally, the same number of participants appreciated how the tool enabled them to uncover

[2]The ethical committee of Polytechnique Montreal granted the permission to conduct this multi stage user research on all individuals involved in this study on November 1st 2023, ID CER-2324-30-D. Written informed consent was obtained from all involved participants.

Table 1: The proposed taxonomy aims at characterizing DDSs for the purpose of filtering them. It has 6 groups, each divided in dimensions. Every dimension can take one or several values for a particular DDS.

Group	Dimension	Values
Metadata	Title	Unique value
	Publication date	Unique vallue
	Source	Unique vallue
Context	Goal	Entertain, explain, inform, persuade, comfort
	Topic	Other; arts, culture, entertainment and media; conflict, war and peace; crime, law and justice; disaster, accident and emergency incident; economy, business and finance; education; environment; health; human interest; labour; lifestyle and leisure; politics; religion; science and technology; society; sport; weather
	Structure	Other, animation, compilation, slideshow, static image, longform infographic, multimedia, storymap
Data	Form	Qualitative, quantitative
	Type	Numerical, network, spacial, temporal, text, other
	Analysis	Yes, no
Contents	Visualizations	Other, flow map, dot map, proportionnal symbol map, choropleth , heatmap, navigation map, cartogram, bar chart, bubble chart, proportional shape chart, circular chart, instance diagram, parallel coordinates, sankey diagram, flow chart, area chart, radar chart, histogram, infographi, line, 3D model, dot plot, organigram, pictogram, range chart, network, table, mosaic chart
	Medias	Image, audio, text, video
	Tutorial	None, explicit, implicit
Controls	Users degree of freedom	Author-drive, martini glass, interactive slideshow, drill-down, reader-driven
	Sequence	No sequence, continuous linear, step-by-step linear, non-linear
	Transitions	No transitions, fading, swiping, tweening, panning
	Progression indicators	Progress bar, section header buttons, timeline, map with progression, breadcrumbs, checklist
	Interactions	No interactions, abstraction/elaboration, comparision, connection, encoding, exploration (navigation), filter, reconfiguration, selction
Narrative patterns	Argumentation	None, comparison, concretization, repetition
	Framing	None, defamiliaraization, familiarization, physical metahor, breaking conventions
	Engagement and Empathy	None, speed-up/slow-down, breaking the 4th wall, make a guess, exploration, humans behind the dots, rhetorical question, gradual reveal

Table 2: Sources used to build the archive.

p23eml	
Title	Source
The New York Times, 2022: The Year in Visual Stories and Graphics	Link
The New York Times, 2021: The Year in Visual Stories and Graphics	Link
The New York Times, 2020: The Year in Visual Stories and Graphics	Link
The Washington post; 2022: The year in graphics	Link
Bloomberg, 2020 in Graphics	Link
Bloomberg, 2021: The Year in Graphics	Link
Bloomberg, 2022: The Year in Graphics	Link
The Wall Street Journal, 2020 The Year in Graphics	Link
The Wall Street Journal, 2021 The Year in Graphics	Link
The Wall Street Journal, 2022 The Year in Graphics	Link
Propublica, 2022: Our Year in Visual Journalism	Link
Propublica (2021) ProPublica's Year in Visual Journalism	Link
LA Times, 2020: The Year in Data and Graphics	Link

innovative presentation techniques. However, the tool's complex interface and extensive taxonomy options presented a bit of a learning curve.

4 Conclusion and Future Work

This paper proposed a methodology intended to improve designers' access to Data Driven Stories (DDS) examples, facilitating creative ideation in DDS production. We developed a comprehensive taxonomy, based on a synthesis of existing scholarly work, which was then used to organize an archive and to design the CuratedDDS online exploration interface. While our evaluation revealed that these resources benefitted practitioners in their work, it also identified key areas for enhancement. To address these, we propose refining the taxonomy using a bottom-up approach that better captures the nuances of practitioners' needs. Additionally, we suggest employing innovative methods such as crowdsourcing to diversify the archive's content, ensuring it includes more current and varied examples. Implementing AI-driven techniques for scraping and automated annotation could also streamline the process of updating the archive, allowing for the inclusion of content from a broader range of news sources. Finally, we recommend expanding the capabilities of the CuratedDDS tool to include advanced functionalities like generative AI, which could tailor examples to user-specific queries, thereby enhancing user experience and resource utility.

Acknowledgments

This research was funded by an NSERC Alliance Grant ALLRP-561132-20, in collaboration with Le Devoir and PROMPT Innov. We acknowledge and appreciate the help and contribution of Le Devoir as well as all the participants in different stages of this study

References

[1] Noé Abraham and Cortés Landaverde. 2018. *A conceptual framework for interactive cartographic storytelling.* Ph. D. Dissertation. http://essay.utwente.nl/85868/

[2] Benjamin Bach, Moritz Stefaner, Jeremy Boy, Steven Drucker, Lyn Bartram, Jo Wood, Paolo Ciuccarelli, Yuri Engelhardt, Ulrike Koeppen, and Barbara Tversky. 2018. *Narrative design patterns for data-driven storytelling.* AK Peters/CRC Press, 107–133.

[3] Hannah K. Bako, Xinyi Liu, Leilani Battle, and Zhicheng Liu. 2022. Understanding How Designers Find and Use Data Visualization Examples. *IEEE Transactions on Visualization and Computer Graphics* 29 (2022), 1–11. https://doi.org/10.1109/tvcg.2022.3209490

[4] Nancy Duarte. 2019. *Data story : explain data and inspire action through story.* Ideapress Publishing.

[5] Yu Fu and John Stasko. 2023. More Than Data Stories: Broadening the Role of Visualization in Contemporary Journalism. *IEEE Transactions on Visualization and Computer Graphics* (01 2023), 1–20. https://doi.org/10.1109/tvcg.2023.3287585

[6] Jonathan Gray, Lucy Chambers, and Liliana Bounegru. 2012. *The data journalism handbook: How journalists can use data to improve the news.* " O'Reilly Media, Inc.".

[7] Yan Holtz. 2022. Dataviz Inspiration | Hundreds of chart examples. https://www.dataviz-inspiration.com/

[8] Robert Kosara. 2016. A roundup of Year-End News Graphics roundups. https://eagereyes.org/blog/2016/a-roundup-of-year-end-roundups

[9] Tarja-Kaarina Laamanen and Pirita Seitamaa-Hakkarainen. 2014. Interview Study of Professional Designers' Ideation Approaches. *The Design Journal* 17, 2 (2014), 194–217.

[10] Andrew Vande Moere and Helen Purchase. 2011. On the role of design in information visualization. *Information Visualization* 10, 4 (2011), 356–371.

[11] Adegboyega Ojo and Bahareh Heravi. 2018. Patterns in award winning data storytelling. *Digital Journalism* 6, 6 (2018), 693–718. https://doi.org/10.1080/

21670811.2017.1403291

[12] Paul Parsons. 2021. Understanding data visualization design practice. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2021), 665–675.

[13] Paul Parsons. 2022. Visualization Design Patterns - InfoVis:Wiki. https://infovis-wiki.net/wiki/Visualization_Design_Patterns

[14] Paul C Parsons. 2023. Design Cognition in Data Visualization. In *Visualization Psychology*. Springer, 197–215.

[15] Robert E. Roth. 2020. Cartographic Design as Visual Storytelling: Synthesis and Review of Map-Based Narratives, Genres, and Tropes. *The Cartographic Journal* 58 (09 2020), 1–32. https://doi.org/10.1080/00087041.2019.1633103

[16] Edward Segel and Jeffrey Heer. 2010. Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics* 16 (11 2010), 1139–1148. https://doi.org/10.1109/tvcg.2010.179

[17] Florian Stalph. 2017. Classifying Data Journalism A Content Analysis of Daily Data-driven Stories. *Journalism Practice* 12 (10 2017), 1332–1350. https://doi.org/10.1080/17512786.2017.1386583

[18] Charles D Stolper, Bongshin Lee, Henry Riche, and John Stasko. 2016. Emerging and recurring data-driven storytelling techniques: Analysis of a curated collection of recent stories. https://www.microsoft.com/en-us/research/publication/emerging-and-recurring-data-driven-storytelling-techniques-analysis-of-a-curated-collection-of-recent-stories/

[19] Charles D Stolper, Bongshin Lee, Nathalie Henry Riche, and John Stasko. 2018. *Data-driven storytelling techniques: Analysis of a curated collection of visual stories*. AK Peters/CRC Press, 85–105.

[20] Alice Thudt, Jagoda Walny, Theresia Gschwandtner, Jason Dykes, and John Stasko. 2018. *Exploration and explanation in data-driven storytelling*. AK Peters/CRC Press, 59–83.

[21] Wibke Weber, Martin Engebretsen, and Helen Kennedy. 2018. Data stories: Rethinking journalistic storytelling in the context of data journalism. *Studies in communication sciences* 2018, 1 (2018), 191–206.

[22] Ziming Wu, Qianyao Xu, Yiding Liu, Zhenhui Peng, Yingqing Xu, and Xiaojuan Ma. 2021. Exploring Designers' Practice of Online Example Management for Supporting Mobile UI Design. *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction* (09 2021). https://doi.org/10.1145/3447526.3472048

[23] Z. Zhao and N. Elmqvist. 2023. The Stories We Tell About Data: Surveying Data-Driven Storytelling Using Visualization. *IEEE Computer Graphics and Applications* 43, 04 (jul 2023), 97–110. https://doi.org/10.1109/MCG.2023.3269850

Received 16 June 2024