# A Visual Exploratory System for Data Facts in Data-driven Articles and Its Use in Business Reports

CHEN XUEYI[1,a)] Ken WAKITA[1,b)]

**概要**：For lots of people working in different fields, reports describing a variety of data facts are authored and analyzed by lots of people every day. To get insights efficiently from such kind of data-rich document, people naturally take notes by using visual marks such as highlights, underlines, and circles traditionally. Recently, as business reporting goes electronic gradually, there is a surge of attention and interest in effective support for such kind of report analysis with the power of computer science. In this paper, we design a visual-guided method focusing on facilitating business report authoring, especially, aiming to help build interactive business reports to convey data facts via the non-programming interface. The proposed method aims to simplify the authoring of interactive business reports, lower the technical barrier, as well as empower people to comprehend the key information effectively. A use case generated by our system based on real business reports is presented to show the usage and capability of the implemented architecture.

## 1. Introduction

Data-driven storytelling (DDS) [19] technique has attracted a lot of attention recently, sparked numerous exploratory studies on related concepts and design principles, tool implementations, discussions on its effectiveness, etc [24]. As one of the prominent forms in the design space of DDS, data-driven articles [23], have been constantly discussed over the past years due to their significance. With the characteristic of combining visualizations with narrative text, data-driven articles provide a potentially efficient way for data explanation for the general public.

Despite its potential advantages in encouraging reading at the cognitive level [3], the creation of such data-driven articles is never easy as the authoring workflow spans a variety of disciplines including data analysis, visualization, human-computer interaction, storytelling, etc. While works focusing on supporting the authoring of data-driven articles flourished and have been proven to be potentially efficacious, the steep learning curve keeps such tools hard to be applied. Programming skills and script knowledge are often required to author with these systems, which brings technique barriers for non-technique users. Intelligence algorithm approaches are recently widely used to automate the process of both data analysis (e.g., visualization recommendation based on machine learning approaches) and storytelling, mitigate the burden of such users [17]. However, the ideal trade-off between the automated process (e.g., text generation and data facts extraction) and the preservation of human involvement (e.g, hand-crafted storytelling and manual insights recommendation during data analysis based on domain knowledge and experiences of the analysts), remains to be discussed [25].

These challenges remain to be tough to solve while adapting data-driven article authoring techniques to business intelligence (BI). Besides, the demands of the BI domain are still not fully taken into consideration among the existing tools and solutions for interactive business report creation [27].

To fill this gap, we propose a visual exploratory system prototype to adopt designs for supporting data-driven article authoring in the field of BI. We stress the significance of human involvement including hand-crafted metrics without the requirement of technical knowledge, to ensure the quality of the generated report in an easy way,

[1] School of Computing, Tokyo Institute of Technology
[a)] xueyichen_980221@outlook.com
[b)] wakita@is.titech.ac.jp

aiming to support the authoring workflow of interactive business reports.

## 2. Related Work

We get insights into the possible challenges and chances of this work by surveying existing literature from 2 perspectives.

### 2.1 Data-Driven Articles

There is a recent surge of attention and interest in Data-driven articles that combine visualization with narrative text, which are also known as "magazine style narrative visualizations (MSNVs)", "narrative visualization" [21], "visual data stories" [16], etc. Though the widely discussed benefits of such synergetic association between text and visualization in promoting active reading, encouraging engagement, as well as supporting understanding, integrating, and remembering the information [6, 10, 3], the authoring of such data-driven remains to be challenging. The workflow acquires multiple skills, including design, data analysis, visualization and storytelling, therefore cooperation among domain experts is often required during the process.

To fill such a gap, tools focusing on improving the authoring experience of data-driven articles have emerged recently. While helpful, technique barriers, including required programming skills, are usually high and thus keep non-technical users from using them [8, 14], especially in the case of creating articles supporting customized interactions between text fragments and charts [23]. Idyll Studio [9] supports WYSIWYG-style editing for interactive document authoring, to a certain extent, but the component's implementation requires graphics coding. Besides, VizFlow [23] is a tool that allows authors to build interactive documents that make use of visual linking between text fragments and charts. It reports, though, that the tool is targeted at more technically experienced authors.

To address this problem, a recent research trend is the introduction of intelligence algorithms in the pipeline, aiming to automate such processes to some extent and reduce the burden on non-technical authors. For example, Calliope [22] is proposed to create data-driven articles from spreadsheets automatically. Erato [24] provides users with key-frames to describe the story. Nevertheless, both tools are designed for relatively short texts. Traditionally, due to the limitation of NLP techniques, the quality of the generated content is hard to ensure. It may be desirable to rely on human editors' skills to polish the quality of the text. However, placing human editors in the process of automated digital text generation requires further study. The customization of interaction between text fragments and charts in the generated article is also limited in the workflow. Thus many works strike a balance between efficiency brought by the intelligence algorithms and the quality and readability ensured by human involvement.

### 2.2 Visualization in BI

Traditionally, dashboards are the most popular tools for visualizations in BI [11, 13], which consist of visual component collections aggregated on a single view. Tools for BI dashboard construction [15, 5] have been discussed exhaustively over the past decades with the support of tools for visualization design, including libraries like D3.js [7], Vega-lite [20], and commercial tools like Tableau Stories[18] and Microsoft BI [12], etc. Recently, there are also efforts toward applying intelligence algorithms to dashboard creation, such as MultiVision [26], which selects data columns and makes recommendations of charts in dashboards based on a deep learning approach.

Nevertheless, so far, data-driven articles that combine storytelling with visualization yet remain seldom discussed in BI [1]. Recently, CODAS [27] is proposed to facilitate authoring by transforming dashboards into interactive business reports. While helpful, the system takes Vega-lite charts as inputs, and thus requires related technical knowledge like scripting. Besides, tabular data presentation is seldom taken considered in such tools. Though the process could be time-consuming, accessing and exploring the underlying data could be important for authors in business analysis. As discussed by Bartram et. al [4], analysts need to get in contact with the underlying data throughout the analysis process as a way to deepen their perception and comprehension of the data facts. For authors of data-driven articles, accessing tabular data enables them to build trust in the system by having greater control over the content they create and more self-confidence in the stories they want to tell during the interactive authoring process. Badam et. al [2] proposed Elastic Documents that enhance document reading by coupling text, visualization, and tabular components, aiming to support comprehension and insights-finding during the reading. But the system does not support authoring and is not aiming at applying such techniques to the BI domain, thus lacks consideration of customization and structured creation of the article, and also

the demands of business analysis, including re-usability.

## 3. Problem Statements and Design Goals

This section describes the technical issues that remain in the past research and the feedback gained from the discussion with the experts in business data analysis. From this observation, the design goals of the system are inspired.

- *For the scenario that applying techniques facilitating the authoring of data-driven articles to the domain of BI, potential challenges are still not fully defined, and the opportunities are not thoroughly investigated.* Traditionally, dashboards are the most popular tools for visualizations in BI and were thoroughly discussed and get supported by commercial tools, while data-driven articles containing large-scale narrative text that combine storytelling with visualization remain seldom discussed in this specific domain.
- *Technical barriers, including the requirements of programming skills, are high and thus keep non-technical users unable to benefit from using these tools.* This happens especially when the addition of the interactive capability to the digital document is desired to make more sense of the document by, for example, adding visual links between text fragments and charts. Meanwhile, much research work seeks an ideal trade-off between relying on intelligence algorithm-based solutions for lowering technological barriers and the involvement of human experts to ensure the quality of the article via hand-crafted metrics.
- *In particular, when considering applying data-driven article authoring techniques to BI, i.e., in the case of interactive business reports authoring, the unique demands of this specific domain are not fully considered among the existing tools and solutions.* While necessary for business analysis, tabular data presentation is seldom taken considered in such system design. The reusability of the generated article for data updating and design customization is also limited.

Three design goals are compiled based on the problem statement and the review of the survey on related work, which guides the development of our system and is discussed throughout the rest of the paper to justify the usability and contribution.

**DG1** *Give an investigation about the hidden challenges and opportunities for adopting techniques facilitating data-driven article writing into the BI domain.* This is based on a literature survey and the data provided by TEIKOKU DATA BANK, LTD (TDB).

**DG2** *Lower the technological barriers based on the design of easy-to-use non-programming components for authoring.* This aims to encourage the involvement of authors via interactive authoring approaches, ensuring the quality and readability of the narrative text content in the created business report. Data facts classified as *References* and *Claims* are allowed to be defined while authoring, to support the customization of interactions between text fragments and charts in the presented GUI-based interface.

**DG3** *Support interactive authoring while meeting the specific needs of the BI domain by providing 3 tight coupling components: Text, Visualization, and Table components allowing the visual exploratory all the way to the underlying data.* Facilitate design customization to construct a narrative structure, and decouple the data and the content to enhance the reusability of the created components and authored documents, helping with dynamic data updates.

## 4. System Design and Implementation

We propose the system design with the guidance of design goals to fulfill the requirements driven by the problem statements (DG1). Figure 3 shows the pipeline of the system.

### 4.1 System Design

The prototype is a web-based editor to facilitate interactive data-driven article authoring. With the raw tabular data, users are able to develop a narrative structured data story using the three basic components supported by the system without programming knowledge (DG2).

Given the dataset (as CSV files), with the specification file (as the JSON) which is prepared by the system for parsing the configuration of the structured arrangement of the article dynamically, three types of basic components: Visualization Component, Table Component, and Text Component would be generated by the system for further use (DG3). Next, users are allowed to arrange these basic components into Cell(s) components in single or double column-layout to fulfill their design customization demands, which would be used to further be arranged into Section components in top-to-bottom linear order, and eventually be modularized into the final artifact. During the authoring process, users are free to upload tabular data in CSV forms, create new cell components and ar-

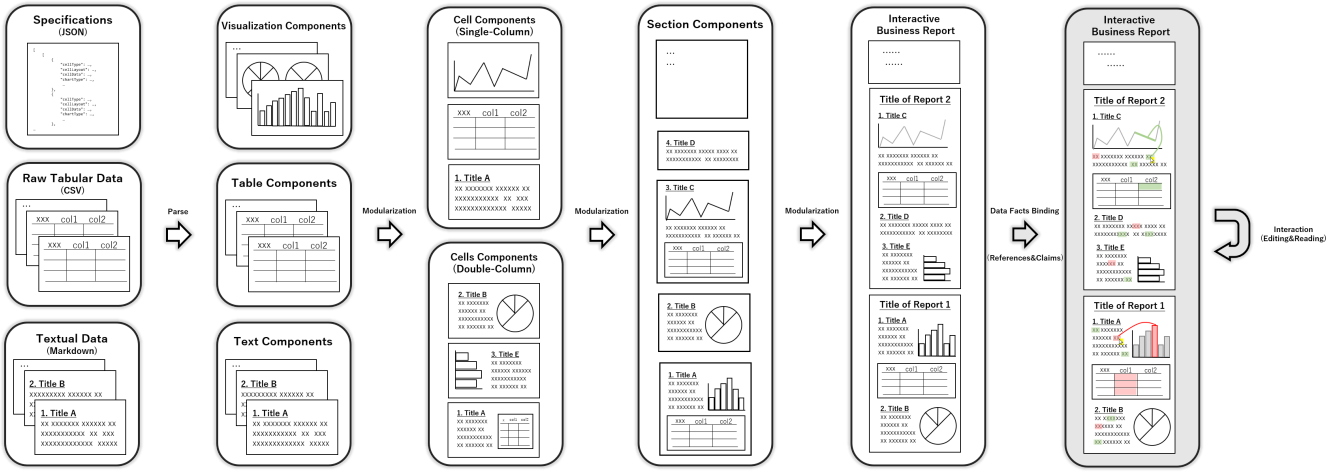


**Fig. 1** The pipeline of our system.

range them into the editing document via the GUI-based interface, as shown in Figure 3 (DG2, DG3).

Two types of Data facts are defined by the system: *References* which simply refer to the data in the table, and *Claims* which are based on statistical reasonings of the data in the table. According to the definition, users are able to assign data facts among the created report and create binding relations for the same data fact presented in different formats (i.e., among the different types of components) via the GUI-based interface. Given the assigned data facts embedded in text, visualization, and (or) table, the tight-coupling among these components would be represented as effects like visual links and highlights triggered by interaction (e.g., hover), so as to support insights finding and comprehension (DG2).

### 4.2 Implementation

A client-server architecture is followed during the development of our system. The server-side is implemented with Python Flask. Raw tabular data and the specifications of the document including layouts and textual data are structured for reuse and update (DG3). Preliminary analysis is handled by the pandas library. The client-side is realized with HTML and JavaScript. Markdown is used to able manual storytelling with the support of the GUI-based interface provided in Text components for non-technical users. Thus the quality of the text content could be ensured (DG2). D3.js and Vue.js are used for visualization construction in Visualization components, which could be used in non-programming interaction-based approaches (DG2). By creating a library of Visualization components using these techniques, the tool is able to ensure the reusability of the authored report, and decouple the raw tabular data from the content (DG3).

## 5. Use Case

A use case (Figure 5) is generated with our system based on the real case of business analysis offered by TDB, so as to show the applicability of the design of this work. Figure 5(a) is a report surveying the Kinki companies' awareness of new coronavirus infections, which includes three sections with four visualization charts and two tabular presentations.

As shown in Figure 5(a), data facts including *References* and *Claims* are assigned manually by the users and are highlighted in green and red in the text components. While being assigned, these data facts presented as text fragments would be asked to be bounded with the elements in both visualization and table components in a GUI-based interface, which support further interactions among different components (DG2). Figure 5(b) shows some of the cases of interaction relying on the tight-coupling of text, visualization and table components. Specifically, when a *Reference* or *Claim* represented in a text fragment is hovered with the mouse by the users who intend to dig into it, this colored text fragment would serve as AOI (area-of-interests) while corresponding parts presenting the same data facts assigned by the author in visualization and the table would be highlighted immediately. Visual links are displayed in the form of B ´ezier curve, as a further enhancement of visual aids. Besides, as shown in Figure 5(c), new cells could be added to the linear arrangement of the article during the process of authoring via non-programming interaction, so as to decouple the data and facilitate the creation of a narrative structure in an easy way even for the non-programming


**Fig. 2** The use case created by our system. (a) The authored business report with assigned data facts (colored in red referring to *Claims* and green referring to *References*) via GUI-based interactions. (b) An interaction case building visual links between referenced text fragments, visualization and table. (c) The support of adding Cell components dynamically to facilitate flexible style customization and data updates.

users.

## 6. Conclusion and Future Work

In this work, we present a visual exploratory system for data facts in data-driven articles and show its usability in business reports authoring. The system consists of non-programming interfaces to encourage the involvement of authors. We plan to organize a user study including semi-structured interviews in the next step, within which a detailed evaluation will be conducted.

## 7. Acknowledgment

**謝辞** The authors would like to thank the experts from TEIKOKU DATA BANK, LTD (TDB) who support this work for their time and feedback.

## References

[1] Christos Amyrotos. “Adaptive Visualizations for Enhanced Data Understanding and Interpretation”. In: *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*. 2021, pp. 291–297.

[2] Sriram Karthik Badam, Zhicheng Liu, and Niklas Elmqvist. “Elastic documents: Coupling text and tables through contextual visualizations for enhanced document reading”. In: *IEEE transactions on visualization and computer graphics* 25.1 (2018), pp. 661–671.

[3] Oswald Barral, Sébastien Lallé, and Cristina Conati. “Understanding the effectiveness of adaptive guidance for narrative visualization: a gaze-based analysis”. In: *Proceedings of the 25th international conference on intelligent user interfaces*. 2020, pp. 1–9.

[4] Lyn Bartram, Michael Correll, and Melanie Tory. “Untidy data: The unreasonable effectiveness of tables”. In: *arXiv preprint arXiv:2106.15005* (2021).

[5] Rahul C Basole. “Visual business ecosystem intelligence: Lessons from the field”. In: *IEEE computer graphics and applications* 34.5 (2014), pp. 26–34.

[6] Michelle A Borkin et al. “Beyond memorability: Visualization recognition and recall”. In: *IEEE transactions on visualization and computer graphics* 22.1 (2015), pp. 519–528.

[7] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. “$D^3$ data-driven documents”. In: *IEEE*

*transactions on visualization and computer graphics* 17.12 (2011), pp. 2301–2309.

[8] Matthew Conlen and Jeffrey Heer. “Idyll: A markup language for authoring and publishing interactive articles on the web”. In: *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. 2018, pp. 977–989.

[9] Matthew Conlen et al. “Idyll studio: A structured editor for authoring interactive & data-driven articles”. In: *The 34th Annual ACM Symposium on User Interface Software and Technology*. 2021, pp. 1–12.

[10] Graham Dove and Sara Jones. “Narrative visualization: Sharing insights into complex data”. In: (2012).

[11] Micheline Elias, Marie-Aude Aufaure, and Anastasia Bezerianos. “Storytelling in visual analytics tools for business intelligence”. In: *IFIP Conference on Human-Computer Interaction*. Springer. 2013, pp. 280–297.

[12] Alberto Ferrari and Marco Russo. *Introducing Microsoft Power BI*. Microsoft Press, 2016.

[13] Danyel Fisher, Steven Drucker, and Mary Czerwinski. “Business Intelligence Analytics [Guest editors’ introduction]”. In: *IEEE Computer Graphics and Applications* 34.5 (2014), pp. 22–24.

[14] Nam Wook Kim et al. “Datatoon: Drawing dynamic network comics with pen+ touch interaction”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019, pp. 1–12.

[15] Ana Lavalle et al. “Visualization requirements for business intelligence analytics: a goal-based, iterative framework”. In: *2019 IEEE 27th International Requirements Engineering Conference (RE)*. IEEE. 2019, pp. 109–119.

[16] Bongshin Lee et al. “More than telling a story: Transforming data into visually shared stories”. In: *IEEE computer graphics and applications* 35.5 (2015), pp. 84–90.

[17] Dominik Moritz et al. “Formalizing visualization design knowledge as constraints: Actionable and extensible models in draco”. In: *IEEE transactions on visualization and computer graphics* 25.1 (2018), pp. 438–448.

[18] Daniel G Murray. *Tableau your data!: fast and easy visual analysis with tableau software*. John Wiley & Sons, 2013.

[19] Nathalie Henry Riche et al. *Data-driven storytelling*. CRC Press, 2018.

[20] Arvind Satyanarayan et al. “Vega-lite: A grammar of interactive graphics”. In: *IEEE transactions on visualization and computer graphics* 23.1 (2016), pp. 341–350.

[21] Edward Segel and Jeffrey Heer. “Narrative visualization: Telling stories with data”. In: *IEEE transactions on visualization and computer graphics* 16.6 (2010), pp. 1139–1148.

[22] Danqing Shi et al. “Calliope: Automatic visual data story generation from a spreadsheet”. In: *IEEE Transactions on Visualization and Computer Graphics* 27.2 (2020), pp. 453–463.

[23] Nicole Sultanum et al. “Leveraging text-chart links to support authoring of data-driven articles with vizflow”. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 2021, pp. 1–17.

[24] Mengdi Sun et al. “Erato: Cooperative Data Story Editing via Fact Interpolation”. In: *IEEE Transactions on Visualization and Computer Graphics* (2022).

[25] Aoyu Wu et al. “Ai4vis: Survey on artificial intelligence approaches for data visualization”. In: *IEEE Transactions on Visualization and Computer Graphics* (2021).

[26] Aoyu Wu et al. “MultiVision: Designing Analytical Dashboards with Deep Learning Based Recommendation”. In: *IEEE Transactions on Visualization and Computer Graphics* 28.1 (2021), pp. 162–172.

[27] Zhuohao Zhang et al. “CODAS: Integrating Business Analytics and Report Authoring”. In: (2022).