Data narrative crafting via a comprehensive and well-founded process

Faten El Outa[1[0000−0002−8153−4252]], Patrick Marcel[1[0000−0003−3171−1174]], Veronika Peralta[1[0000−0002−9236−9088]], Raphaël da Silva[2], Marie Chagnoux[3], and Panos Vassiliadis[4[0000−0003−0085−6776]]

[1] University of Tours, Blois, France. firstname.lastname@univ-tours.fr
[2] Rue89 Strastbourg, France. raphipons@gmail.com
[3] University Paris 8, Paris, France. marie.chagnoux@univ-paris8.fr
[4] University of Ioannina, Ioannina, Greece. pvassil@cs.uoi.gr

Abstract. Data narration is the activity of crafting narratives supported by facts extracted from data exploration and analysis, using interactive visualizations. While data narration has recently attracted much attention, the process of crafting data narratives is loosely documented and has not yet been formally described. In this article, we propose a comprehensive and well-founded process to fill this need. It aims at (i) supporting the complete cycle of data narration, from the exploration of data to the visual rendering of the narrative, (ii) being flexible enough to cover a wide range of crafting practices, and (iii) being well founded upon with a conceptual model of the domain.

Keywords: Data narrative crafting, Data journalism, Process

1 Introduction

Data narratives are receiving increasing interest from several research communities (e.g., visualization, data management, computer-human interfaces) [2] and many application domains (e.g. journalism, business, e-government, health). They are largely used by journalists, scientists, and other communicators, to convey striking messages to a given audience. In addition, the crafting of a data narrative includes a variety of activities, including the analysis of data, the drawing of relevant messages from data, the structuring of messages into a coherent story and its visual rendering. Despite this diversity of activities, sometimes even conducted by different people with varied professions and skills, there is no framework, workflow, or tool for supporting the crafting of data narratives.

In an effort to clarify the concepts of data narratives, we recently defined a data narrative *as a structured composition of messages that (a) convey findings over the data, and, (b) are typically delivered via visual means in order to facilitate their reception by an intended audience*, and we proposed a conceptual model describing and structuring the key concepts around data narratives [12]. This model (described in Section 2) is organized in 4 layers: factual, intentional, structural and presentational, which reflect the transition from raw data to the visual rendering of the story. With this definition and model in mind, our aim in this paper is to contribute with a study of the dynamic aspects of data narrative crafting. Like many works in the literature (e.g., [8, 10, 5]), we postulate that the different forms of data narration can be described by a comprehensive process encompassing the various activities ranging from data exploration to the rendering of the narrative. A formal description of this process will benefit novice data narrators, like e.g., non technical data journalists, and will be instrumental to the development of tools for supporting advanced data narrators.

Accordingly, we reviewed the literature around the process of crafting data narratives, and we conducted a survey with data journalists in order to understand how they craft a data narrative. As an outcome of the former, we found that globally the research communities agreed in the fact that the crafting process includes three main phases: (i) the *analyzing* phase that handles the activities of exploring data, retrieving findings and formulating messages learned from data, (ii) the *structuring* phase that includes the activities to organize the plot of the narrative in an understandable way and, (iii) the *presenting* phase that covers the activities to convey the structured messages visually. However, our bibliographical study revealed the absence of a comprehensive and well-founded process that covers the main activities of the crafting process, specially those dealing with user intentions and their tight relation to data analysis. From the survey, we observed the crafting workflows regularly followed by 18 data journalists, and we contrasted them to the literature. It turned out that journalists follow the same three phases, mostly in a linear way, attaching less attention to the structuring phase, while spending more time in the analyzing phase.

These considerations from the literature study and the survey with data journalists enabled us to identify the activities (and their chaining) for crafting data narratives. Based on those, we propose a comprehensive and well-founded process that (i) covers the whole cycle of data narrative crafting, from exploration of the data to the visual presentation of the narrative, (ii) accommodates a wide range of practices observed on the field, and, (iii) is founded on a conceptual model of the domain that clarifies the concepts involved in the process [12].

The scope of our method targets the population of data journalists or any other data enthusiast that constructs data narratives out of existing data. The reason for proposing the method is exactly the observed discrepancy between literature and practice, with omissions of important parts from both sides. *Thus one significant contribution of our work is the explicit treatment of all the steps that should be involved in the process.* Secondly, apart from providing a methodological guidance, our method can enable *the support of the process via tooling.* Indeed, there is a lack of integrated tools covering the whole crafting process and recommending actions to less-experienced narrators. In particular, an application that would automatically document the data exploration and narration crafting is desperately needed by data workers, who spend hours to document their work. This is important for reproducibility, transparency, and linkage, and requires a conceptual model and a process that are both consensual.

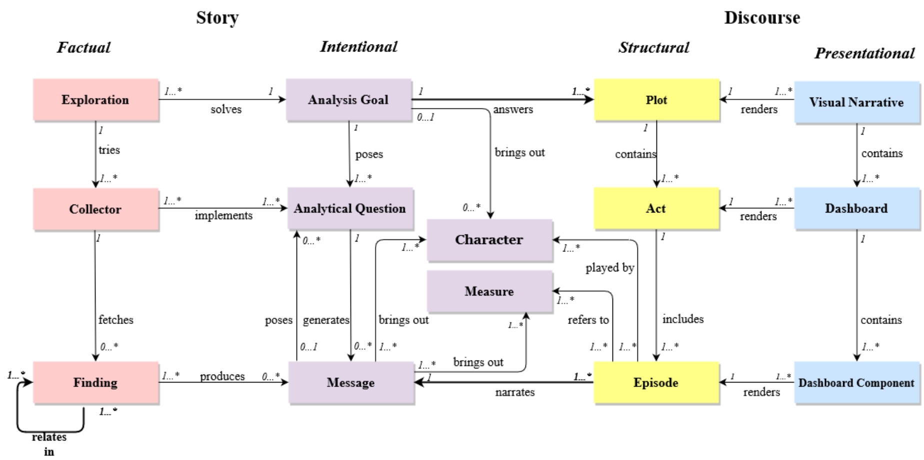


Fig. 1. The conceptual model for data narratives (relations in bold were extended w.r.t. the original version in [12])

The paper is organized as follows: Section 2 recalls the key concepts of the conceptual model proposed in [12]. Section 3 reviews the related work for the processes of crafting a data narrative, and Section 4 discusses the survey we conducted with data journalists. The proposed process is detailed in Section 5. Section 6 concludes and draw research directions.

2 A conceptual model for data narratives

We recently proposed in [12] a conceptual model of data narratives providing a principled definition of the key concepts of the domain, along with their relationships, and clarifying their role and usage (see Fig. 1).

This model is based on 4 layers following Chatman's organisation [4], who defined narrative as a pair of (a) *story* (content of the narrative), and, (b) *discourse* (expression of it). In our model, the *factual* layer handles the *exploration* of facts (i.e., the underlying data), via a set of *collectors* that allow for manipulating facts with varied tools and fetching *findings*, in an objective way, while the *intentional* layer models the subjective substance of the story, identifying the *messages*, *characters* and *measures* the narrator intends to communicate, and tracing how they are obtained through *analytical questions*, according to an *analysis goal*. As to the discourse, the *structural* layer models the structure of the data narrative, its *plot* being organized in terms of *acts* and *episodes*, while the *presentational* layer deals with its rendering, that is communicated to the audience through visual artifacts (*dashboards*[5] and *dashboard components*).

[5] We use the term dashboard since it is general enough to accommodate various types of visualizations (e.g. a Business Intelligence dashboard, an infographics, a section in a python notebook, a section in a blog or web page).

The interested reader is redirected to [12] for a deeper presentation of the model. Here, we will highlight the main decisions behind the model that are necessary for grasping its essence. Importantly, it should be noted that the concept of *message* is the model's corner stone, which is clearly evidenced by the way we have related message to the other concepts. A specific message is rooted in the facts analyzed, conveying essential findings, potentially raising new analytical questions. The message allows introducing episodes, the building blocks of the discourse. Each episode of the discourse is specifically tied to a message which it aims to convey. The relationship between messages and episodes is the basis for structuring stories that address analysis goals, narrated by structured discourses (with cohesive acts being the backbone of the narrative structure) and dashboards their presentational counterpart.

3 Related work

In this section, we review the works describing the internals of the data narration process, as well as the tools that automate (part of) the crafting process.

3.1 Global data narration processes

Data narration is a complex process, at the crossroads of several domains: data exploration, data visualization, data management, etc. Despite the many contributions in each of these areas, few works offer comprehensive workflows describing the entire data narration process. The first attempt to model data narration processes come from the visualization community. For example, Kosara and Mackinlay [9] proposed a two-phases process: First, narrators collect information and *explore* their interrelationships, pointing to key facts, and then, they *tie* those facts together into a story. Chen et al. [5] surveyed early proposals and concluded that their crafting processes are composed of two main phases: (a) *visual analytics*, which requires seeing all aspects of complex data, explore their interrelationships, and is supported by multiple coordinated views and sophisticated interaction techniques, and (b) *storytelling*, which is meant to convey only interesting or important information (i.e. findings) extracted through the analysis, presented in a simple and easily understandable way.

To bridge the gap between these two phases, Chen et al. proposed an intermediate one, called *data synthesis* [5]. In this phase, the narrator assembles and organizes the findings to be communicated, to represent explicitly the essential relationships between them, building a compelling narrative. Lee et al. [10] also identified three main phases: *explore data* to retrieve findings, *make a story* to turn findings into a sequence of narrative pieces to build the plot of the narrative, and *tell a story* to materialize the plot in a visual manner. The authors stress the importance for the data narrator to go back and forth between the exploration and the story-making phases. More recently, Duangphummet et al. [6] proposed a protocol consisting of the following phases: *conceptualization* of the data narrative domain, targeted audience and distribution channel, *data preparation* to deliver data that is relevant to the use, *realization* to deliver a storyline with detailed content and an initial form of key visualizations, *visualization design* to redesign the visualizations and create visualization prototypes, and finally, the *visualization development* where technical requirements are defined, and the key visualizations for target devices are developed and deployed.

In addition to [10], many works underline the importance of moving between the data narrative crafting phases. For instance, Wang et al. [19] ran a workshop on data comics, organized by an interdisciplinary team with expertise in data visualization, graphic design, data comics, and illustration. They observed that to create stories, students require to *move back and forth between the story, visualizations, and the data*.

Besides the previously described works proposing global crafting processes, some works describe subprocesses, focusing on the necessary activities to be conducted. Without being exhaustive, we mention here some major contributions. Battle and Heer [1] identified three ways to start a data narrative: having a precise idea in mind, having a vague idea refined during data exploration, or having no idea before exploring the data. Weber et al. also point that the crafting process starts by either an idea, a problem or a question [20]. Notably, many works underline the importance of different story structures and different kinds of interactivity in data narration [13, 20]. In particular, Weber et al. [20] encourage to use non-linear structures and set up interactivity. Many works specifically deal with the phase of structuring the narrative [5, 18, 10, 14].

Finally, very few works highlight the importance of intentional aspects. Thudt et al. [16] stress that subjective perspectives can be introduced at every step of visualization creation: during data collection and processing, visual encoding, and when refining the presentation. In the context of OLAP cube exploration, Vassiliadis et al. [17] propose a set of intentional operators to express high-level analytical intentions and automate their translation to database queries.

3.2 Automated data narration

Many recent works addressed the automatic generation of data narratives, providing another source of insights on how this process is perceived.

Wang et al. [18] conducted a qualitative analysis of 245 infographics examples to explore the infographics design space in terms of structures, sheet layouts, fact types, and visualization styles. Based on those, the authors propose a system for supporting a fact sheet generation pipeline consisting of three phases: (i) fact extraction, (ii) fact composition, and (iii) presentation synthesis. Shi et al. [15] proposed Calliope, a system that can automatically generate visual data stories with facts arranged into a logical sequence. It consists of two main modules: (i) the story generation engine, for generating, choosing and organizing the facts that will participate in the narrative, and (ii) the story editor, that visualizes the data story (generated as a series of visualization charts) and allows the users to change it based on their preferences. Shi et al. [14] described the workflow for crafting data videos, consisting of 4 phases: (i) *collecting a series of data facts* around a certain topic, (ii) *constructing a storyline* as an assembly of these data facts into a sequence, (iii) *choosing data visualizations* for the data facts and deciding how to animate them by drawing a storyboard, and finally, (iv) *realizing the storyboard* via a design software in which the narrator edits and combines the animated visualizations until a coherent data video is accomplished.

In CineCubes [7], Gkesoulis et al. detail the process of crafting a data movie in the form of a powerpoint presentation, to answer a specific user's need described by a query. First, an introductory act is built with the initial query, and two subsequent acts are used to put context. These acts contain visualizations highlighting important facts, as well as text and audio describing these facts. A summary act concludes with all the important highlights of the previous acts.

In all these works, the proposed phases are consistent with those described in the previous subsection. Being a mostly automatic generation, the construction is linear in the sense that there is no back and forth movements between phases. In addition, they target a specific domain or data format and organize stories accordingly to pre-established patterns. In particular, we highlight the absence of intentions, that are, at best, modeled via an initial query or a topic.

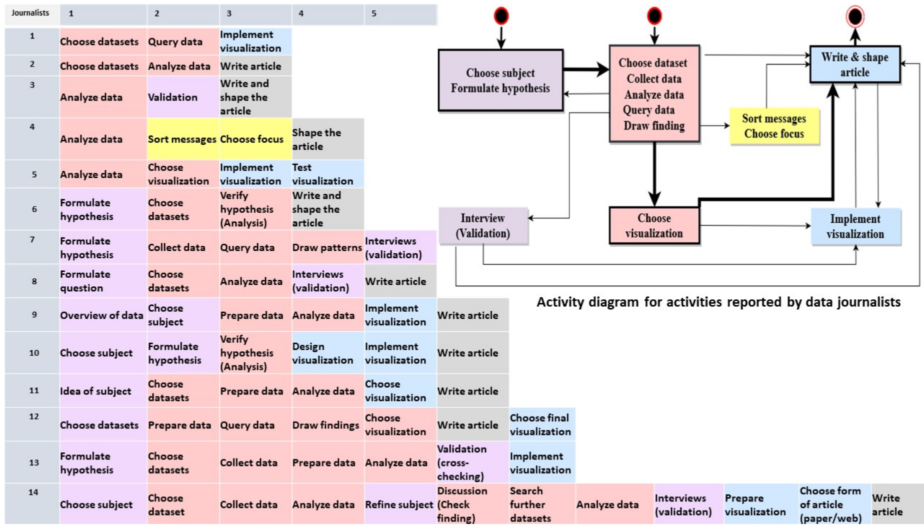
Lessons learned. Most of the works describing the data narration process agree on the 3 general phases of exploration (to retrieve findings), structuring (organizing the information gathered into narrative pieces) and presentation (crafting visual artifacts). Automated data narration is still in its infancy, mainly applying rigid patterns and lacking the necessary flexibility of moving between the 3 phases. One of the key findings is that the intentional layer of the model presented in Fig. 1 is largely absent from the works reviewed. This means the substance of the story, i.e., the **composition** of story elements (analytical questions and hypothesis, messages, etc.) as pre-processed by the author's cultural code [4] is ignored. We claim that this absence is regrettable; if data narrations are to be shared, reused, their crafting process documented, then this intentional layer deserves more attention.

4 Data journalist practices

A preliminary study, in the form of a survey [3] (in French), investigates the professional practices of data journalists.

The survey consisted of 32 questions[6], answered by 18 data journalists from 14 French regions, who have worked on a big variety of topics, including elections, environment, cinema, terrorism, paradise-papers, real estate. For nearly 50% of them, data narration is at the core of their professional activity, and is occasional or marginal for the others. Concerning training, 56.3% studied social sciences, 18.8% studied sciences and 24.9% graduated from law or journalist schools. One of the journalists works for the International Consortium of Investigative Journalists (ICIJ), 5 of them work for the national press, and the 12

[6] `https://drive.google.com/drive/folders/1zDzP_ndSlQUJCbtFMVzJDnIbyXK1D2_l?usp=sharing` (in French). Note that for some questions more than one answer was possible, and that journalists could leave the questions unanswered.



Journalists	1	2	3	4	5							
1	Choose datasets	Query data	Implement visualization									
2	Choose datasets	Analyze data	Write article									
3	Analyze data	Validation	Write and shape the article									
4	Analyze data	Sort messages	Choose focus	Shape the article								
5	Analyze data	Choose visualization	Implement visualization	Test visualization								
6	Formulate hypothesis	Choose datasets	Verify hypothesis (Analysis)	Write and shape the article								
7	Formulate hypothesis	Collect data	Query data	Draw patterns	Interviews (validation)							
8	Formulate question	Choose datasets	Analyze data	Interviews (validation)	Write article							
9	Overview of data	Choose subject	Prepare data	Analyze data	Implement visualization	Write article						
10	Choose subject	Formulate hypothesis	Verify hypothesis (Analysis)	Design visualization	Implement visualization	Write article						
11	Idea of subject	Choose datasets	Prepare data	Analyze data	Choose visualization	Write article						
12	Choose datasets	Prepare data	Query data	Draw findings	Choose visualization	Write article	Choose final visualization					
13	Formulate hypothesis	Choose datasets	Collect data	Prepare data	Analyze data	Validation (cross-checking)	Implement visualization					
14	Choose subject	Choose dataset	Collect data	Analyze data	Refine subject	Discussion (Check finding)	Search further datasets	Analyze data	Interviews (validation)	Prepare visualization	Choose form of article (paper/web)	Write article

Fig. 2. Sequence of activities reported by journalists

remaining work for the regional press. Regarding their general working habits, 75% of them work alone. They usually work on open data (72.2%) and more specifically on data from public institutions (44.4%). They consume from minutes to months during the data narration and use different tools during data exploration, such as spreadsheets (93.8%), scripts (50%), notebooks (18.8%), powerBI-like tools (31.3%) and some machine learning tools (28.6%).

Two main questions were asked on their data narration practices. For the first one, "How does a data story's subject emerge?", multiple answers were possible. The answers showed that the goal, or subject, of an article emerges from: an idea to be confirmed by data (68%), a dataset which needs exploration to reveal important facts (68%), a refinement of the subject while exploring the dataset (48%). The second, open question was: "What is the general workflow you apply for data narrative crafting?". Fig. 2 sketches the answers provided by 14 of the 18 journalists, where activity names summarize journalists' descriptions of their main activities[7], rows correspond to journalists and column numbers reflect the sequence of activities. We color these activities according to the layers of the conceptual model: factual (pink), intentional (purple), structural (yellow) and presentational (blue). Gray-colored cells indicate that the activity may overlap structural and (more probably) presentational tasks. In addition, activities concerning the checking of findings and the validation of messages (namely interviews, validation or cross-checking), aiming at transforming a factual object into an intentional one, are in between the factual and intentional layer. Similarly, visualizations are used both in the factual layer, to understand data and

[7] Since the question was open, we homogenized the answers and grouped them into few categories.

retrieve findings, and in the presentational layer, to choose the most suitable one for communicating findings to the audience in a visual manner.

We have abstracted these sequences in the form of an activity diagram (top-right corner of Fig. 2). Most frequent paths are highlighted by larger arrows.

Lessons learned. Fig. 2 shows that many activities under different names aim towards the same action, and that different paths can be followed by journalists when crafting a data narrative. *The figure also shows a preponderance of activities from the factual and the intentional layer.* The activity diagram shows that journalists enter the workflow either in the factual layer, i.e., by exploring a dataset, or by the intentional layer, i.e., having at least a vague idea of the subject. After this, the workflow becomes mostly linear, with some movements between the factual and intentional layers. Usually, data journalists start writing their articles once the analysis phase is over, and there is no backtrack once the presentational layer is entered.

Notably, the journalists attached little importance to the activities of the structural layer. At the exception of one of them, structuring activities are either hidden in writing activities or even not mentioned explicitly. Precisely, many of them agree that while data exploration usually takes long, visual storytelling can be extremely fast, potentially done on the fly, with some of them actually not even involved in the writing of the article. For those that mention it, the activity "write article" includes several hidden details concerning the organization of messages that should be communicated, the visual presentation and communication of the analysis results.

Overall, we can say that there is a chasm between what practitioners do and what literature suggests – and in fact, there are deficits in both sides. On the one hand, compared to what is reported in the literature, the work of the data journalists is over-emphasizing the intentional part and under-investing on the structural and the presentational part. On the other hand, when it comes to the literature, the presented methodologies overemphasize presentation and (to some extent) structuring, and pay much less attention to the intentional part. A process that gracefully hosts all aspects of narrative construction would facilitate narratives that are more complete and intuitive.

5 A process for crafting data narratives

From the literature review and the survey with journalists, we synthesize a set of requirements for a comprehensive data narration process and we propose a process that fulfills these requirements.

5.1 Requirements

First of all, we note the absence in the literature of a whole workflow for crafting data narratives, including all the activities identified in Section 3 and Section 4. Fig. 3 depicts the activities as phrased in the literature (in gray boxes) and

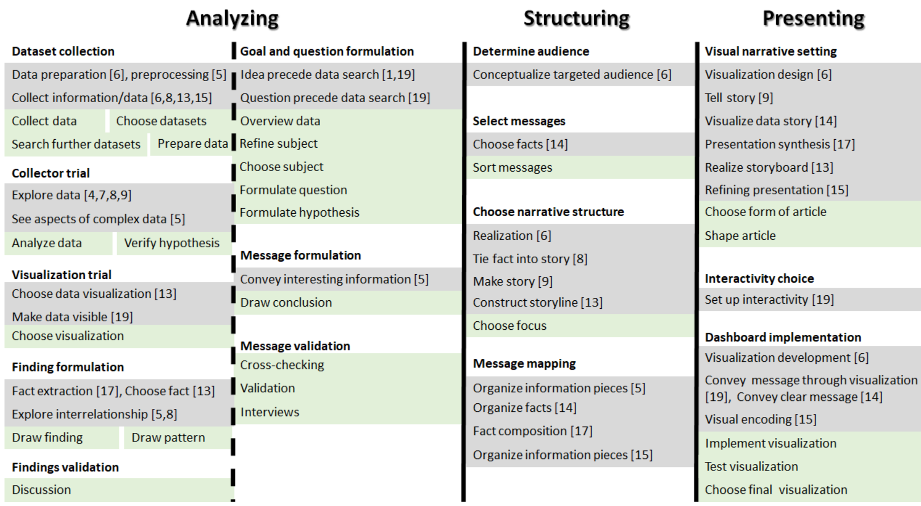


Fig. 3. The main activities for crafting data narratives identified from the literature and a survey with data journalists (AQ abbreviates Analytical Question)

by journalists (in green boxes). We group those referring to the same task and propose new names (the bold ones in Fig. 3) which are consistent with the conceptual model of Fig. 1.

In more details, most authors [5, 10, 18, 15, 6, 14, 9, 19] agree that data narration process includes three main phases: (i) *analyzing*, (ii) *structuring*, and (iii) *presenting*. The survey reveals that the data journalists agreed with the literature, especially on the phases (i) and (iii). In Fig. 3, activities are grouped according to these phases. We remark that activities pertaining to the factual and intentional layers of the conceptual model are mixed in phase (i). In addition, while the literature rarely mentions the activities pertaining to the intentional layer, these activities are often pointed by data journalists. Furthermore, as we explained in [12], the substance of a story, representing the narrator's intention in reporting the story, is a constituent of the data narrative [4]. Conversely, while the journalists did not attach much importance to the activities of the structural phase, this aspect is emphasized in the literature. Finally, as noted in [19, 10], the narrator should have the possibility to move freely back and forth between the different phases of data narration. However, this movement should not prevent that different groups of activities could be conducted by different persons with different profiles. These groups of activities, identified by layers in the conceptual model [12], should be as isolated as possible.

To summarize, a comprehensive workflow for crafting data narratives should satisfy the following requirements:

- (R_1) cover the activities and the paths identified by the survey with data journalists, which are depicted in Fig. 2,
- (R_2) cover the activities of the three phases identified from the literature,

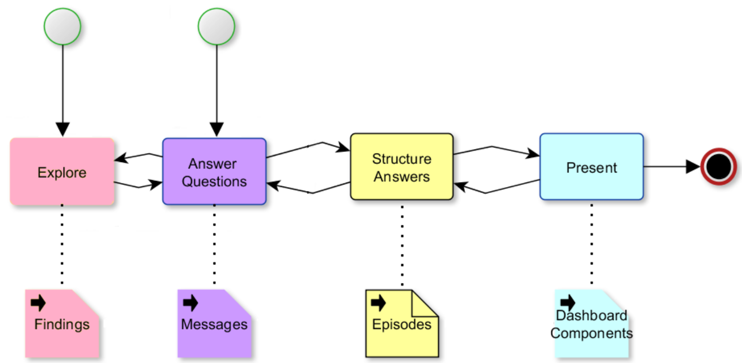


Fig. 4. The process of data narrative crafting

– (R_3) allow the free back and forth transition between phases,
– (R_4) clearly delineate the different layers of the conceptual model [12] within its activities.

5.2 The process of crafting data narratives

In this subsection, we propose a comprehensive process for the crafting of data narratives that covers the activities and paths proposed in the literature and reported by journalists (requirements R_1 and R_2), while also being founded upon and coherent with the conceptual model (R_4) and allowing the back and forth movement between its phases (R_3).

The phases of the process are illustrated in Fig. 4. All phases are accompanied by the resulting outcomes, which are exactly the basic constituents of our conceptual model (R_4). We retain the same coloring (pink for factual exploration, purple for intentional question-answering, yellow for the structuring of the answers of the intentional questions into a plot, and blue for presentation). Observe that the factual and intentional layers of the conceptual model are well differentiated here, contrarily to the literature that mix them into one phase.

Consistently with Fig. 2, the process flexibly starts either with the existence of a data set, which is to be explored for findings, or with the emergence of an initiating question, that begs to be answered. This flexibility is important in the sense that prescribing a specific starting point for the collection of findings from the data is not what actually the practitioners do.

The following paragraphs describe the activities pertaining to each phase, including the activities abstracted from the literature and survey results (shown in Fig. 3), and some new activities that intent to cope with missing tasks.

Note that such activities should not be considered as steps to be executed sequentially. Conversely, many activities can be initiated and executed in parallel. Arrows in Fig. 5 indicate a *depends on* relationship. For example, message validation depends on message formulation, as it is necessary to formulate messages before validating them. In addition, at any time, it is possible to come back to

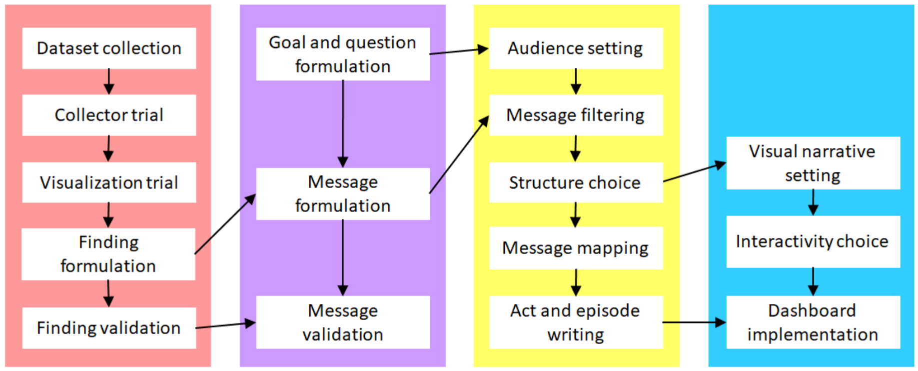


Fig. 5. Activities for data narrative crafting

previously executed activities (e.g. to rewrite messages or formulate new ones). Backtrack arrows are omitted for clearness.

Exploration The exploration phase, handling the factual layer, concerns several activities: (i) dataset collection, concerning source selection, data extraction, integration and preprocessing, (ii) trial and reuse of several collectors (i.e. querying, profiling and mining tools) and (iii) trial of diverse visualizations (crosstabs, graphics, clusters, etc.) for collecting findings, then, (iv) finding formulation, concerning the storage of findings and their relationships, and (v) finding validation, which is typically done via statistical tests, but also by discussing and crosschecking with additional data sources (as done by data journalists) and confronting with the state of the art (as done by data scientists [11]). Note that some findings may lead to additional analysis, triggering more collectors and visualisations, or even the collection of more datasets. The exploration phase is time-consuming (data journalists measure it in days or even in months). Then many activities are frequently performed asynchronously.

Question-answering This phase, neglected in the literature, handles the intentional layer and concerns activities for (i) formulating goals and questions, (ii) drawing messages from findings, and (iii) validating messages. It supports explicitly the data narrator intention, as its proposed activities help in formulating an analysis goal and a set of analytical questions that reflect their intention.

Furthermore, to cope with literature lacks (evidenced in Fig. 3), we propose a message formulation activity, concerning the derivation of messages from findings, and the identification of characters and measures (the relevant constituents of messages [4]) to be highlighted to the audience.

Remember the distinction of outcomes: A *finding* is a highlight, (or equivalently, a pattern) annotating a dataset, or a subset of it. A finding can be a typical pattern (like e.g., an association rule, or a path of a decision tree) or the verification of rejection of a hypothesis for the data. A *message*, on the other hand, is the answer to the intentional question that exploits a finding to label a character with respect to other characters or a measure. For example, based on data findings, one can answer questions like:

- By comparing $Daily\ Infections$ in $France$ to $EU\ Average$, we find that they are $similar$. Here, the character is the entity $France$, which is an instance of the concept $Country$, and we label its measure $DailyInfections$ with respect to another peer character, $EU\ Average$.
- By correlating the concept $News\ Authenticity$ to the concept $Media\ outlet$, we find a non-significant correlation, rejecting the hypothesis that the outlet can solely determine the existence of fake news.

The internals of the process, detailed in Fig. 5, allow the flexibility of exploring several paths, that can be chained according to narrator's habits and specificities of the task on hand, alternating data analysis, finding derivation and message writing, but also allowing for the validation of findings and messages, or even the expression of new analytical questions.

In any case, the identification of such messages and their structuring is a task that is practically absent from the related literature, significantly present in the everyday work of practitioners, and structured in our model for the first time.

Structuring The structuring phase, the missing part in the data journalists processes, handles the structural layer, describing acitivities for organizing the plot of the narrative in terms of acts and episodes [7] (adopted from the classical structure of plays). Plot setting starts by (i) determining the audience, (ii) eventually selecting a subset of messages for such audience, and (iii) choosing an appropriate narrative structure. Then, (iv) messages are mapped to acts and episodes. In more details, these activities allow the arrangement of the thoughts of the data narrator into different layers: an act which is a major piece of information, and is composed of several episodes that are of lesser importance on their own [12]. Remember that the result of the structuring is an *episode*, which is the annotation of a message (which has a simple structure and labeling) with comments on the context, significance, essence, etc., in other words with the content that makes the message interpretable by human beings.

Also, observe in Fig. 5, the existence of a specific activity to make the actions of writing acts and episodes explicit. Such activities can be performed before or at the same time as choosing visual means.

Presentation Finally, the presentation phase handles the presentational layer, and includes activities for (i) setting the type of visual narratives, (ii) setting the interactivity mode, and (ii) implementing dashboards for conveying acts and episodes to the audience. Such activities carry on the visualization level and build for each act an associated dashboard and present the narration in a complete visual narrative. Remember that *dashboard components* are representations of episodes in (typically) a visual form of communication, including text, figures, charts, data plots, or any other means to convey the message.

5.3 Discussion

The purpose of this paper is to support data storytelling via a method based on a conceptual model, that fits in all possible domains where storytelling is applicable. One of the main drivers for the method was to bridge the observed gap between literature and practice, with omissions of important parts from both sides. As a result, the method allows the structuring of the overall process in phases and facilitates valid translations between important intermediate results that are necessary to construct a data narrative, in a way that is flexible, realistic and adequately structured.

The intended users of the method are data journalists and data enthusiasts. We ran various preliminary experiments to empirically assess its potential of being adopted by various data workers. We organized a challenge[8] where data enthusiasts (among which journalists, students, social workers) were mixed with data scientists, aiming at producing data narratives using the open data of a French city. Interestingly, all teams started with a vague idea of the topic they wanted to treat, which was refined after many iterations among data collection, data analysis and question formulation. All of them used a unique timeline for structuring their narratives, which were rendered with varied forms. In another experiment, 44 students in BI were asked to craft a data narrative using a dataset they were familiar with, while having no experience in data narratives crafting. Students were observed during crafting, and some of them, especially those less skilled, were asked to indicate the sequence of activities they realized. This helped them to start, particularly having to write down the analytical questions that guided the data analysis, and to write down messages and early think about structuring. Finally, some of the authors of the present paper crafted a data narrative about tuberculosis, targeted for epidemiologists and public health decision-makers in Gabon. The whole crafting process is described in [11]. We highlight the importance of goal setting and message formulation activities, both of them being validated by many experts with different profiles. In particular, in scientific context, messages are not only validated by statistical tests but also confronted to other data sources and similar works of the state of the art and should pass risk assessment tests.

6 Conclusion

In this paper, we proposed a process for crafting data narratives, that covers the whole cycle of data narration, from data exploration to the visual presentation of the narrative. Backed by a literature review and a survey with data journalists, it accommodates a wide range of practices observed on the field, via clearly delineated activities, while being well founded upon a conceptual model of the domain [12]. We believe that these two models, static and dynamic, can serve as a stepping stone for future research in the area of data narration.

Extending the proposal with tool support for guiding the narrator along the process and (semi-)automating some tasks, is a clear path for future work.

[8] Sponsored by French CNRS https://www.madics.fr/event/titre1617704707-3351/#madona.

References

1. Battle, L., Heer, J.: Characterizing exploratory visual analysis: A literature review and evaluation of analytic provenance in tableau. Comput. Graph. Forum **38**(3), 145–159 (2019)
2. Carpendale, S., Diakopoulos, N., Riche, N.H., Hurter, C.: Data-driven storytelling (dagstuhl seminar 16061). Dagstuhl Reports (2016)
3. Chagnoux, M.: La datavisualisation, double point d'entrée du datajournalisme dans la PQR (in french). Interfaces numériques **9**(3) (2020)
4. Chatman, S.: Story and Discourse: Narrative Structure in Fiction and Film. Cornell University Press (1980)
5. Chen, S., et al.: Supporting story synthesis: Bridging the gap between visual analytics and storytelling. TVCG (2018)
6. Duangphummet, A., Ruchikachorn, P.: Visual data story protocol: Internal communications from domain expertise to narrative visualization implementation. In: VISIGRAPP (2021)
7. Gkesoulis, D., Vassiliadis, P., Manousis, P.: Cinecubes: Aiding data workers gain insights from OLAP queries. Inf. Syst. **53**, 60–86 (2015)
8. Kosara, R.: An argument structure for data stories. In: Kozlíková, B., Schreck, T., Wischgoll, T. (eds.) EuroVis (2017)
9. Kosara, R., Mackinlay, J.D.: Storytelling: The next step for visualization. Computer **46**(5), 44–50 (2013)
10. Lee, B., et el.: More than telling a story: Transforming data into visually shared stories. IEEE Computer Graphics and Applications **35**(5), 84–90 (2015)
11. Ondzigue Mbenga, R., et. al: A data narrative about tuberculosis pandemic in Gabon. In: DARLI-AP (2022)
12. Outa, F.E., Francia, M., Marcel, P., Peralta, V., Vassiliadis, P.: Towards a conceptual model for data narratives. In: ER. pp. 261–270 (2020)
13. Segel, E., Heer, J.: Narrative visualization: Telling stories with data. TVCG **16**(6), 1139–1148 (2010)
14. Shi, D., Sun, F., Xu, X., Lan, X., Gotz, D., Cao, N.: Autoclips: An automatic approach to video generation from data facts. Comput. Graph. Forum **40**(3), 495–505 (2021)
15. Shi, D., et al.: Calliope: Automatic visual data story generation from a spreadsheet. TVCG **27**(2), 453–463 (2021)
16. Thudt, A.F., Perin, C., Willett, W., Carpendale, S.: Subjectivity in personal storytelling with visualization. Information Design Journal **23**(1), 48–64 (2017)
17. Vassiliadis, P., Marcel, P., Rizzi, S.: Beyond roll-up's and drill-down's: An intentional analytics model to reinvent OLAP. Inf. Syst. (2019)
18. Wang, Y., et al.: Datashot: Automatic generation of fact sheets from tabular data. TVCG **26**(1), 895–905 (2020)
19. Wang, Z., Dingwall, H., Bach, B.: Teaching data visualization and storytelling with data comic workshops. In: CHI. ACM (2019)
20. Weber, W., Engebretsen, M., Kennedy, H.: Data stories: Rethinking journalistic storytelling in the context of data journalism. Stud Commun Sci **18**, 191–206 (2018)