

Un modèle conceptuel de narration de données

Faten El Outa*, Patrick Marcel*, Veronika Peralta*

* Laboratoire LIFAT, Université de Tours
prenom.nom@univ-tours.fr

Résumé. La narration de données est l'activité de production de narrations étayée par des faits extraits de l'analyse de données, à l'aide de visualisations potentiellement interactives. Malgré l'intérêt croissant de plusieurs communautés, il n'y a pas de définition consensuelle ni de modèle formel de narration de données. Dans cet article, nous proposons un modèle conceptuel de narration de données. Il décrit tout le processus de narration : la définition d'un objectif d'analyse, la collecte et l'exploration de données, la découverte de faits marquants, l'extraction des principaux messages, leur structuration pour construire l'intrigue et la présentation finale par des moyens visuels.

1 Introduction

La narration est considérée comme l'une des plus anciennes professions et un pilier de la communication d'informations comme moyen éducatif. Elle concerne l'usage de techniques pour communiquer une histoire à une audience. Plus récemment, la narration de données (*i.e.* narrer une histoire avec des visualisations de données) (Hullman et al., 2013) reçoit un intérêt croissant dans plusieurs communautés telles que le journalisme, les affaires, le e-gouvernement et la science des données. Elle est décrite comme l'activité de production de narrations, étayée par des faits, extraits de l'analyse de données, à l'aide de visualisations potentiellement interactives (Carpendale et al., 2016). Plus précisément, ces narrations peuvent être vues comme des séquences ordonnées d'étapes, chacune pouvant contenir des mots, images, visualisations, audio, vidéo, etc., basés sur des données (Kosara et Mackinlay, 2013).

En dehors de ces considérations générales, et à notre connaissance, il n'existe pas de définition consensuelle ni de modèle formel de la narration de données. Bien que la narration des données ait essentiellement attiré l'attention de la communauté visualisation (Carpendale et al., 2016; Kosara et Mackinlay, 2013), nous affirmons : (i) qu'une approche plus globale est nécessaire dans des domaines tels que la visualisation, la gestion des données, l'exploration de données et l'apprentissage automatique, et (ii) que la modélisation conceptuelle du domaine devrait orienter les recherches futures, afin d'aider à comprendre les narrations de données. Une telle modélisation des aspects et des choix de conception dans le domaine de la narration des données, fournie aux concepteurs de systèmes et d'algorithmes, permettra de faciliter l'exploration

rapide, d'automatiser des tâches et de soutenir les workflows itératifs et collaboratifs (Wongsuphasawat et al., 2019).

Dans (Outa et al., 2020b), nous proposons un modèle conceptuel décrivant les concepts clés du domaine de la narration de données, leurs relations et leurs usages. L'objectif de ce modèle est de guider l'auteur tout au long du processus de narration : collecter des données, découvrir des faits marquants, structurer des messages pour construire la narration et la restituer avec des moyens visuels.

Dans cet article, nous présentons un résumé de ces contributions, organisé ainsi : la section 2 décrit les principaux concepts, synthétise les travaux connexes et introduit notre définition de narration de données, la section 3 décrit le modèle de narration de données et la section 4 présente nos conclusions et perspectives.

2 De narration à narration de données

Les théoriciens de la narration s'accordent à dire qu'il y a au moins deux niveaux dans toute narration : certains événements se produisent (ce qui est raconté) et ces événements sont présentés et transmis à un public d'une certaine manière (comment c'est raconté). Dans la terminologie structuraliste la plus répandue, la réponse au "quoi" s'appelle une **histoire** et la réponse au "comment" s'appelle un **discours** (Akleman et al., 2015). En distinguant ces éléments, (Chatman, 1980) définit une narration comme un couple histoire (contenu de la narration) et discours (expression de celle-ci). L'histoire a une forme, qui est l'ensemble des objets et des événements possibles, et une substance, qui est une **composition** des éléments de l'histoire (c'est-à-dire les événements, les décors, les comportements, les personnages) telle que pré-traités par le code culturel de l'auteur. Le discours a une forme d'expression, qui est une traduction du contenu de l'histoire en une **combinaison structurée** des éléments de l'histoire et une substance, qui comprend l'ensemble de tous les **médias** utilisés pour montrer des éléments structurés (texte, images, tableaux ou graphiques).

La communauté de visualisation de données s'est récemment intéressée à la narration visuelle des données (en anglais, visual data narration, ou visual storytelling) (Carpendale et al., 2016). En particulier, Chen et al. (2018) distinguent (a) l'analyse visuelle, qui nécessite de voir tous les aspects des données complexes, d'explorer leurs interrelations, et est soutenue par de multiples points de vue coordonnés et des techniques d'interaction sophistiquées, de (b) la narration, qui est censée ne transmettre que des informations intéressantes et/ou importantes, extraites à travers l'analyse et présentées de manière simple et facilement compréhensible. Les deux processus diffèrent par leurs objectifs, les utilisateurs cibles, le type d'informations traitées et les méthodes de présentation des informations. Une étape intermédiaire entre l'analyse et la narration est nécessaire pour soutenir la narration des résultats de l'analyse visuelle, dans laquelle **l'analyste assemble et organise les informations à communiquer**.

Inspirés par (Chatman, 1980) et (Chen et al., 2018), nous proposons la définition suivante de narration de données : *Une narration de données est une composition structurée de messages qui (a) transmettent des trouvailles ou découvertes¹ dans les*

1. Appelées *findings* dans notre modèle. Par la suite nous donnerons la correspondance en anglais des termes du modèle dans les notes de bas de page.

données (b) généralement par des moyens visuels afin de faciliter leur réception par un public cible.

Nous empruntons la terminologie de Chatman et étendons sa structure narrative en considérant que la narration de données doit décrire comment le contenu de l'histoire est dérivé à partir des données. Cela se fait en distinguant 4 couches dans notre modèle de narration des données, comme l'illustre la figure 1 : les deux premières couches représentent l'*histoire* et les deux dernières représentent le *discours*. Dans l'histoire, une couche *factuelle* représente sa forme tandis qu'une couche *intentionnelle* représente sa substance. Dans le discours, une couche *structurelle* représente sa forme et une couche *présentationnelle* représente sa substance.

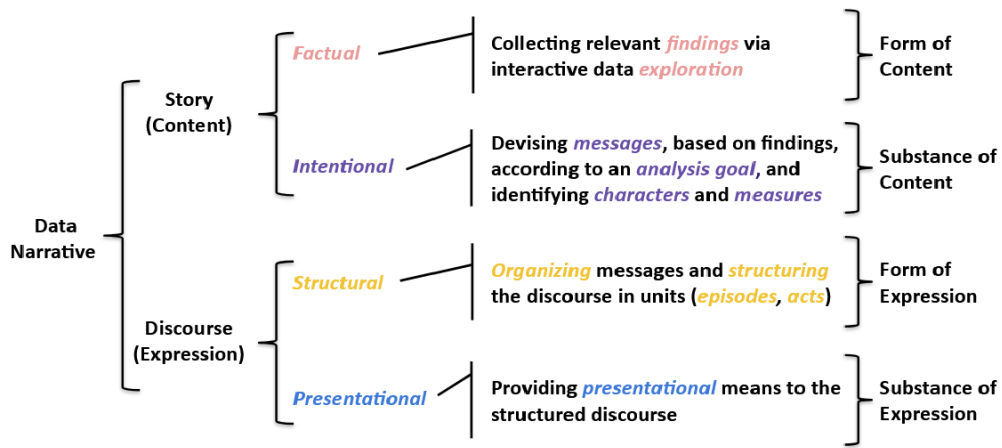


FIG. 1 – Couches du modèle de narration de données, inspirés de (Chatman, 1980)

3 Modèle de narration de données

Le modèle proposé est présenté dans la figure 2. Ses 4 couches reflètent la transition des données brutes jusqu'au rendu visuel. Plus précisément, en cheminant vers l'audience, les données traversent les couches suivantes :

- **Couche factuelle** : Cette couche concerne l'*exploration* des sources de données à travers un ensemble de *collecteurs* qui permettent de manipuler les données avec des outils variés. Les *trouvailles* issues des données explorées sont des candidates pour participer à l'histoire.
- **Couche intentionnelle** : Cette couche modélise la substance de l'histoire, en identifiant les *messages*, *caractères* et *mesures* que l'auteur a l'intention de communiquer. Ces messages répondent à des *questions analytiques* que se pose l'auteur, selon son *objectif d'analyse*². Les questions analytiques sont implémentées par les collecteurs, leurs trouvailles permettant de produire des messages.

2. analysis goal

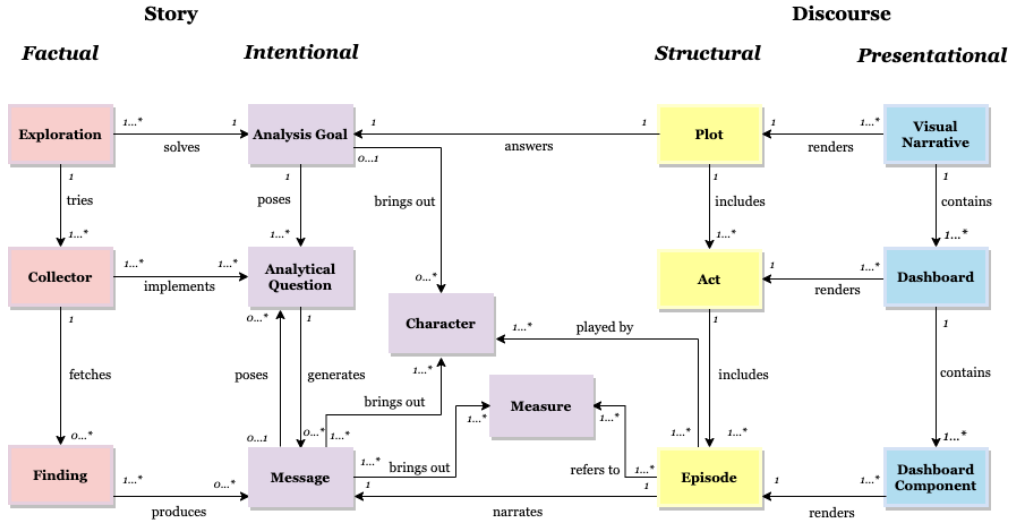


FIG. 2 – Le modèle de narration des données, organisé en 4 couches (Outa et al., 2020b)

- **Couche structurelle** : La couche structurelle concerne la structuration du discours. Son but est d’organiser les différents messages de l’*intrigue*³ en termes d’*actes* et *épisodes*. Les caractères et mesures définis précédemment pourront être mis en avant dans les épisodes.
- **Couche présentationnelle** : Cette couche modélise le rendu de la narration de données, c’est-à-dire la *narration visuelle* qui est communiquée à l’audience via des artefacts visuels : *tableaux de bord*⁴ et *composants*⁵.

Pour comprendre l’organisation du modèle, il faut noter que le concept de *message* est la pierre angulaire du modèle, ce qui est clairement attesté par la façon dont nous avons relié le message aux autres concepts. Essentiellement, un message spécifique est ancré dans les données analysées, véhiculant des trouvailles essentielles dans les données, qui répondent à une des questions analytiques et peuvent déclencher des nouvelles questions. Ce message spécifique est alors l’élément structurel du discours : les épisodes racontent des messages spécifiques, et les actes, construits comme des ensembles d’épisodes, racontent un message plus large. Un message est également connecté indirectement à la couche de présentation : un message global est véhiculé visuellement par un tableau de bord, chaque composant du tableau de bord illustrant un message spécifique.

Pour prouver la faisabilité de notre modèle conceptuel, nous avons développé un prototype d’une application web basée sur le modèle conceptuel proposé (Outa et al., 2020a). Il est implémenté en Java utilisant Spring, d3.js, JFreeChart et Apache PDF-

3. plot
4. dashboards
5. dashboard components

Box. Il guide un narrateur dans la création d’une narration de données de A à Z : il peut chercher et explorer les données, abstraire les messages importants basés sur un objectif intentionnel, structurer le contenu de l’histoire de données et le présenter d’une manière visuelle. Ce prototype s’appuie sur le modèle pour aider l’auteur à structurer une narration de données tout en explorant une base de données de manière interactive. Le code, les captures d’écran et un PDF de narration de données générée avec le prototype sont disponibles sur Github⁶.

4 Conclusion

Cet article a introduit un modèle conceptuel pour la narration des données, en étendant un modèle classique de narration (Chatman, 1980) pour refléter la transition des informations, dérivées de l’analyse des données brutes, jusqu’à leur rendu visuel. Ce modèle s’appuie sur les concepts classiques de la narration pour recouvrir la collecte de données, l’extraction des trouvailles et des messages correspondants, la structuration de ces messages et finalement leur présentation à l’audience via un ensemble de tableaux de bord. Pour mettre en valeur le modèle, comme preuve de concept, nous avons implémenté une application web qui aide un auteur à structurer une narration de données tout en explorant de manière interactive une base de données.

Parmi les améliorations possibles du modèle, nous avons identifié les éléments suivants : l’ajout de transitions entre les épisodes/actes supportant différentes structurations sémantiques (par exemple : linéaire, causalité (Kosara et Mackinlay, 2013)), l’ajout d’un support pour la plausibilité et la cohérence des messages et des arguments, la distinction entre acteurs principaux et acteurs secondaires, et la prise en compte de l’interactivité (Sarıkaya et al., 2019). Nous travaillons également à la définition d’une méthodologie complète pour guider le narrateur de données dans l’utilisation de notre modèle conceptuel.

Références

- Akleman, E., S. Franchi, D. Kaleci, L. Mandell, T. Yamauchi, et D. Akleman (2015). A theoretical framework to represent narrative structures for visual storytelling. In *Bridges*, pp. 129–136.
- Carpendale, S., N. Diakopoulos, N. H. Riche, et C. Hurter (2016). Data-driven storytelling (dagstuhl seminar 16061). *Dagstuhl Reports* 6(2), 1–27.
- Chatman, S. (1980). *Story and Discourse : Narrative Structure in Fiction and Film*. Cornell paperbacks. Cornell University Press.
- Chen, S., J. Li, G. Andrienko, N. Andrienko, Y. Wang, P. H. Nguyen, et C. Turkay (2018). Supporting story synthesis : Bridging the gap between visual analytics and storytelling. *TVCG*, 1–1.

6. <https://github.com/OLAP3/pocdatastorytelling>

- Hullman, J., S. M. Drucker, N. H. Riche, B. Lee, D. Fisher, et E. Adar (2013). A deeper understanding of sequence in narrative visualization. *IEEE TVCG* 19(12), 2406–2415.
- Kosara, R. et J. Mackinlay (2013). Storytelling : The next step for visualization. *IEEE Computer* 46.
- Outa, F. E., M. Francia, P. Marcel, V. Peralta, et P. Vassiliadis (2020a). Supporting the generation of data narratives. In J. Michael et V. Torres (Eds.), *ER Forum, Demo and Posters 2020 co-located with 39th International Conference on Conceptual Modeling (ER 2020), Vienna, Austria, November 3-6, 2020*, Volume 2716 of *CEUR Workshop Proceedings*, pp. 168–172. CEUR-WS.org.
- Outa, F. E., M. Francia, P. Marcel, V. Peralta, et P. Vassiliadis (2020b). Towards a conceptual model for data narratives. In G. Dobbie, U. Frank, G. Kappel, S. W. Liddle, et H. C. Mayr (Eds.), *Conceptual Modeling - 39th International Conference, ER 2020, Vienna, Austria, November 3-6, 2020, Proceedings*, Volume 12400 of *Lecture Notes in Computer Science*, pp. 261–270. Springer.
- Sarikaya, A., M. Correll, L. Bartram, M. Tory, et D. Fisher (2019). What do we talk about when we talk about dashboards? *IEEE Trans. Vis. Comput. Graph.* 25(1), 682–692.
- Wongsuphasawat, K., Y. Liu, et J. Heer (2019). Goals, process, and challenges of exploratory data analysis : An interview study. *CoRR abs/1911.00568*.

Summary

Data narration is the activity of producing narrative supported by facts extracted from data analysis, possibly using interactive visualizations. In spite of the increasing interest in data narration in several communities, there is no consensual definition of data narrative, let alone a formal model of it. In this paper, we propose a conceptual model of data narrative. It aims to support the entire lifecycle of building a data narrative: define an intentional goal: fetch and explore data, bring out highlights, derive important messages, structure the plot of the data narrative, and render it in a visual manner.