

# VISUALISATION DE DONNÉES MASSIVES SNCF

Maguelonne Chandesris<sup>1</sup>, Anaïs Rémy<sup>1</sup>

<sup>1</sup> *SNCF Innovation & Recherche*

*40 avenue des terroirs de France, 75611 PARIS CEDEX 12*

*prenom.nom@sncf.fr*

**Résumé.** Innovation & Recherche SNCF mène depuis plusieurs années des travaux de visualisation de données tant sur la question spécifique de la représentation de données spatio-temporelles que sur la cartographie des données elles-mêmes. Les derniers développements pour visualiser la mobilité seront présentés.

**Mots-clés.** visualisation de données, spatio-temporel, transports

**Abstract.** Innovation & Research Department of SNCF has been investaged datavisualization for years. Some works have been done in visualizaing spatio-temporel data but also in mapping data. Last developments about how to visualize mobility will be shown.

**Keywords.** datavisualization, spatio-temporal, transportation

## 1 Visualisation de données SNCF

Depuis une dizaine d'années, nous nous sommes intéressés au sein de Innovation & Recherche SNCF à la visualisation de données et avons développé différents prototypes sur des jeux de données à dimension industrielle (cf. 2.2.1 p35 de (Chandesris, 2016)). En s'appuyant notamment sur les travaux de Bertin (1999) et Tufte (2001), différents travaux ont été réalisés : au départ très centrés sur le fait de rendre *visible* les résultats des travaux méthodologiques effectués par ailleurs, ils ont démontré également une capacité d'exploration au départ relativement inattendue.

### 1.1 Explorer la cartographie des données

Un premier cas d'intérêt original pour la visualisation de données est la cartographie des données elles-mêmes. SNCF, comme beaucoup d'entités génère et traite quotidiennement un grand nombre de données très variées. Un "annuaire de données", avec recherche par mot clé, comme celui actuellement proposé pour l'open data sur [data.sncf.com](http://data.sncf.com), est certes d'intérêt mais toujours un peu frustrant pour le statisticien qui cherche à faire des liens entre les données et pouvoir donc les explorer de cette manière. C'est précisément ce qui a été proposé dans les travaux de Pelloux (2016)<sup>1</sup> : une exploration des données sous forme de navigation dans un graphe. Une illustration extraite du prototype est reprise à la figure 1.

---

1. travaux visibles en ligne sur <http://visu.sciences-du-design.org/14.html>

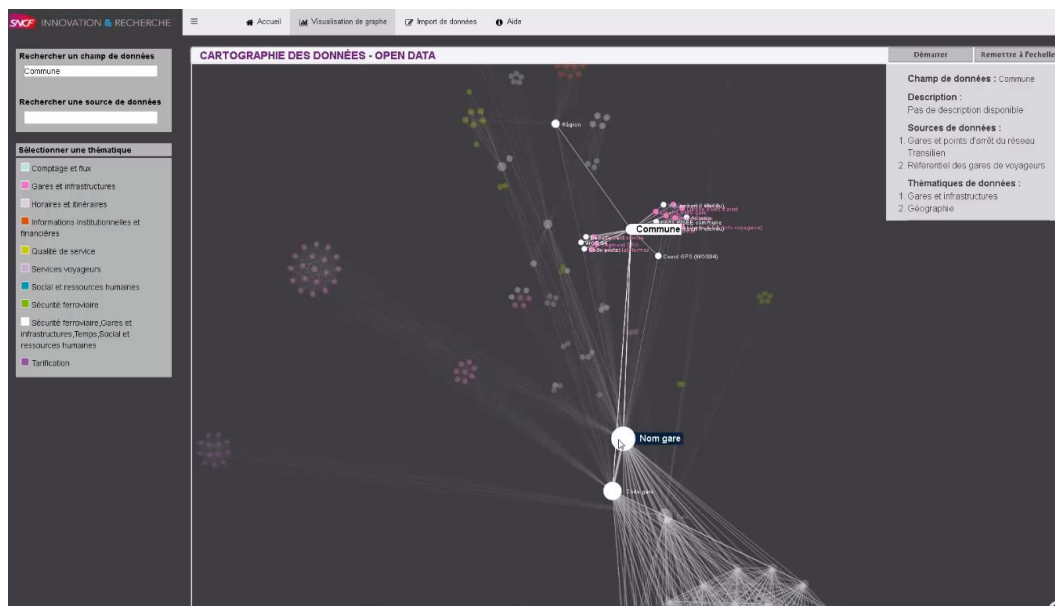


FIGURE 1 – Extrait d’une capture d’écran du prototype d’exploration des données de l’open data SNCF

## 1.2 Donner à voir des phénomènes spatio-temporels

Au delà de ce premier cas d’étude, la plupart des prototypes de visualisation de données réalisés ont permis de donner à voir des phénomènes spatio-temporels intrinsèques à différentes problématiques métiers : les pics et creux de trafic, la circulation TGV<sup>2</sup>, la régularité, les données du yield management, la consommation et le coût de l’énergie de traction<sup>3</sup>. Une synthèse de ces travaux peut être trouvée dans Chandesris et Rémy (2014).

Représenter des phénomènes spatio-temporels est toujours une question d’actualité comme nous allons le voir dans la suite de l’article qui vise à exposer nos dernières avancées dans le domaine.

## 2 Visualisation de données de mobilité

La visualisation de données de mobilité à grande échelle est toujours problématique car il existe des contraintes visuelles importantes. Typiquement, représenter directement chacun des déplacements recensés d’un point  $A$  à  $B$  sur une carte amène vite à l’effet ”spaghetti” : plus rien n’est visible, seul un enchevêtrement de segments rapidement

2. travaux également visibles en ligne <http://senseable.mit.edu/trainsofdata> du fait de la collaboration SNCF-MIT (2011)

3. énergie électrique utilisée pour faire avancer les trains

illisible. Pour autant, donner à voir la mobilité reste un enjeu essentiel pour mieux la comprendre et opérer des solutions pertinentes. De nouvelles approches et technologies permettent de mettre en œuvre des propositions innovantes. Nous souhaitons présenter ici quatre approches de visualisation de la mobilité qui ont été développées avec des données, des outils et des modes de représentations de natures variées : décrites dans la table 1, des visuels sont proposés à la figure 2.

	Type de visualisation	Données	Outils
1	standard	suivi anonyme d'opérateur téléphonique	Tableau Software
2	sur mesure	validations de titres de transports	D3.js
3	algorithme avancé	requêtes d'itinéraires	D3.js, Tulip <sup>4</sup>
4	cartographie avancée	requêtes d'itinéraires et données contextuelles (emploi, logement, événements, etc.)	D3.js, leaflet.js, pixi.js

TABLE 1 – Quatre propositions de visualisation de la mobilité

1. La première approche (illustrée en haut à gauche de la figure 2) permet d'explorer facilement grâce à un outil du marché des flux de mobilité observés sur le territoire suivant l'heure de la journée.
2. La deuxième approche présentée (en haut à droite) est en réalité une interface web réalisée sur mesure présentant différentes vues sur les données observées et estimées de trafic : point de vue géographique, tarifaire, flux entre gares, fréquentation dans les gares et ce aux différents moments de la journée.
3. La troisième vue (en bas à gauche) présente comment les flux des différents couples (origine, destination) en Ile-de-France (plus de 70 000 possibilités) peuvent être graphiquement agrégés par la technique du "bundling" (cf. Holten et van Wijk (2009)) pour faire apparaître des "flux principaux".
4. Enfin, la dernière réalisation (en bas à droite de la la figure 2) qui fait l'objet d'une communication spécifique Côme et Rémy (2017), permet d'explorer de manière spatio-temporelle, à différents niveaux d'agrégation, à la fois les requêtes d'itinéraires et le contexte local et événementiel correspondant.

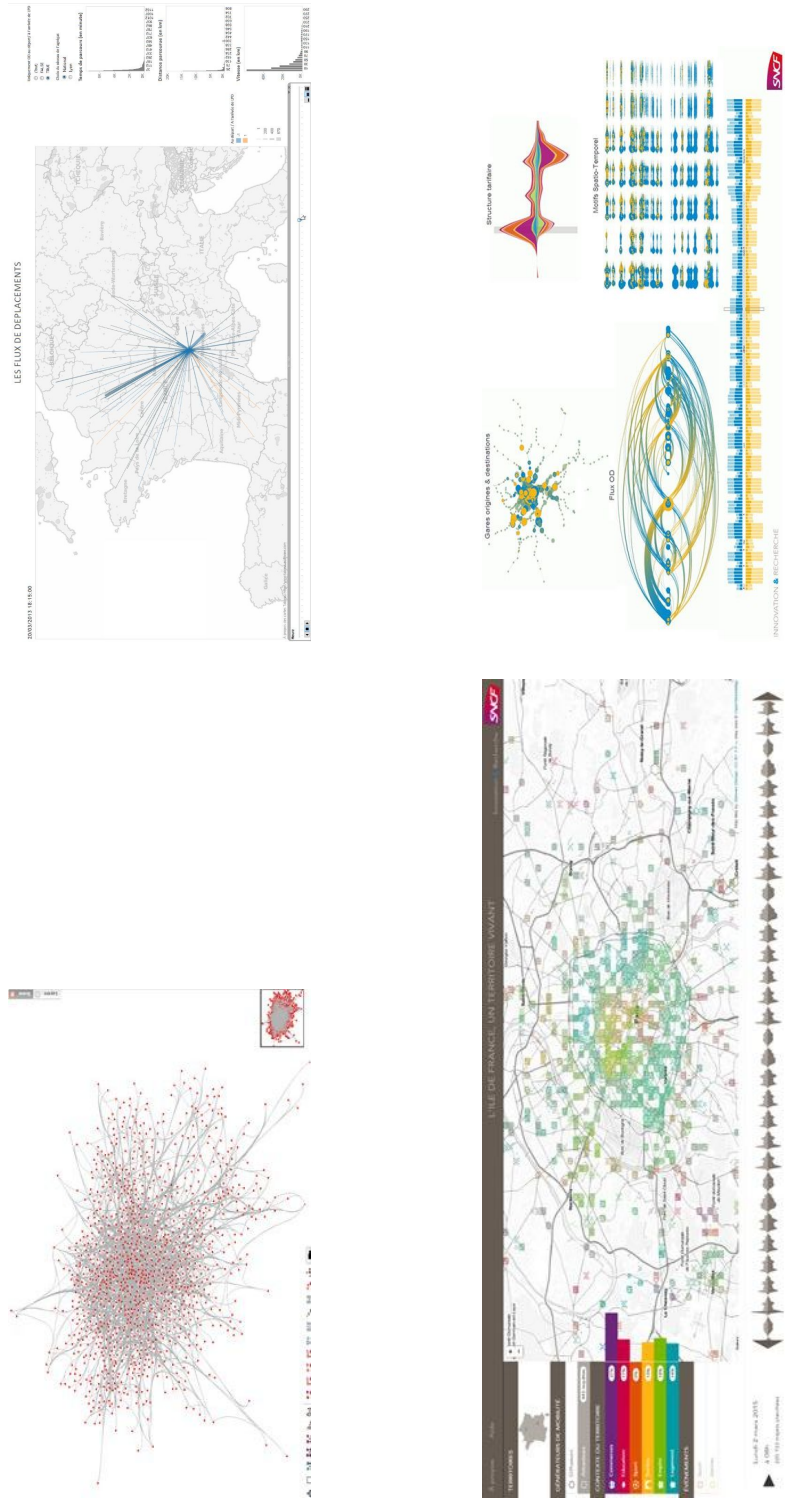


FIGURE 2 – Extraits de captures d'écrans des prototypes de visualisation de données de mobilité

En conclusion, nous souhaitons souligner ici l'intérêt d'un travail multidisciplinaire pour contribuer à réaliser des visualisations de données riches et porteuses de sens. Les derniers travaux (4) ont ainsi pu bénéficier d'une démarche associant statisticien, designer, expert en analyse de la mobilité, digital et développement informatique, cf. Guérin, Chandesris et Rémy (2016). Ce nouveau type d'approche de plus en plus nécessaire ne manquera pas d'interroger les compétences, savoirs-faire et être à développer lors de formation à la datavisualisation.

## Bibliographie

- [1] Bertin, J. (1999), *Sémiologie graphique : les diagrammes - les réseaux - les cartes*, Editions de l'Ecole des Hautes Etudes en Sciences, collection les réimpressions.
- [2] M. Chandesris (2016), *La fabrique de produits avancés de données*, mémoire du mastère spécialisé "Innovation by Design" de l'ENSCI - Les Ateliers, Paris
- [3] M. Chandesris et A. Rémy (2014), *Visualisations interactives et dynamiques de grands volumes de données spatio-temporelles. Exemples d'applications développées à la SNCF.*, Technique et Science Informatiques, vol. 33/5-6, pp439-463
- [4] (à venir) Côme, E. et Rémy, A. (2017), Multiscale Spatio-Temporal Data Aggregation and Mapping for Urban Data Exploitation, *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, Bruges, Belgique
- [5] Guérin, C., Chandesris, M. et Rémy, A. (2016), *Vers des algorithmes à dess(e)ins : intégration du design dans la conception d'une datavisualisation*, Sciences du Design, 2016/2, n°4, p30-50
- [6] Holten, D. et van Wijk, J. J. (2009), Force-directed edge bundling for graph visualization, *Eurographics/ IEEE-VGTC Symposium on Visualization*
- [7] Pelloux, S. (2016), *Cartographie des données : donner à voir l'open data SNCF*, Sciences du design, 2/2016 (n°4), p. 11-15
- [8] Tufte, E. (2001), *The visual display of quantitative information*, Grpahic Press, USA.