# DESIGNING A DATA VISUALISATION AND ANALYSIS TOOL FOR SUPPORTING DECISION-MAKING WITH PUBLIC TRANSPORTATION NETWORK

**Vallet, Flore (1,2);**
**Khouadjia, Mostepha (2);**
**Amrani, Ahmed (2);**
**Pouzet, Juliette (3)**

1: Université Paris-Saclay, CentraleSupélec, Laboratoire Genie Industriel, France;
2: IRT SystemX, Paris-Saclay, France;
3: SNCF Innovation & Research, France

## ABSTRACT

Massive data are surrounding us in our daily lives. Urban mobility generates a very high number of complex data reflecting the mobility of people, vehicles and objects. Transport operators are primary users who strive to discover the meaning of phenomena behind traffic data, aiming at regulation and transport planning. This paper tackles the question "How to design a supportive tool for visual exploration of digital mobility data to help a transport analyst in decision making?" The objective is to support an analyst to conduct an ex post analysis of train circulation and passenger flows, notably in disrupted situations. We propose a problem-solution process combined with data visualisation. It relies on the observation of operational agents, creativity sessions and the development of user scenarios. The process is illustrated for a case study on one of the commuter line of the Paris metropolitan area. Results encompass three different layers and multiple interlinked views to explore spatial patterns, spatio-temporal clusters and passenger flows. We join several transport network indicators whether are measured, forecasted, or estimated. A user scenario is developed to investigate disrupted situations in public transport.

**Keywords**: Data visualisation, User centred design, Creativity, Digital / Digitised engineering value chains, mobility data

**Contact**:
Vallet, Flore
IRT SystemX
Territoires Intelligents
France
flore.vallet@irt-systemx.fr

**Cite this article:** Vallet, F., Khouadjia, M., Amrani, A., Pouzet, J. (2021) 'Designing a Data Visualisation and Analysis Tool for Supporting Decision-Making with Public Transportation Network', in *Proceedings of the International Conference on Engineering Design (ICED21),* Gothenburg, Sweden, 16-20 August 2021. DOI:10.1017/pds.2021.109

# 1 INTRODUCTION

Nowadays, we are constantly surrounded with multiple streaming data sources (Internet Of Things, sensors…) that support our daily lives and activities. From that observation, design engineering also becomes more influenced by data-driven inspirations in several areas of interest, for instance: data-driven construction of personas (Stevenson and Mattson, 2019) (Salminen et al., 2020); creation of an undergraduate curriculum on data materialisation, i.e. learning how to interact with data (Beghelli, Huerta-Canepa and Segal, 2019); exploration of the design space in Design by shopping (Abi-Akle, Minel and Yannou, 2017); elaboration of a data-driven concept generation and evaluation approaches for supporting designers in the early phases of the design process (Han et al., 2020).

Design for mobility systems becomes an increasing challenging topic for the design community since mobility is a major societal issue regarding its impact on global sustainability, liveability of cities and also well-being of citizens. A public transport system ensures collective movement of passengers by using different transport modes. Management of such networks is a quite challenging task particularly in metropolitan cities where crowded and jammed traffic is daily recorded. In addition, operators have to deal with different types of disturbances that affect the quality of service.

Usually, the organization of transport network is carried either in planning or in regulation. The planning phase is done beforehand, the timetables are determined and communicated in advance to all public transport stakeholders. The timetables are established to meet the transport demand according to traffic conditions (Gkiotsalitis et al., 2019). However, in practice, theoretical timetables may deviate from real-time ones due to disturbances (Daraio et al., 2016) caused by different factors (sport or cultural events, weather conditions, incidents related to equipment or personnel, accidents, etc.). Due to their stochastic nature, these variations cannot be handled in the planning phase (Gkiotsalitis et al., 2019) but rather in the real-time regulation phase which insures acceptable levels of quality of service.

In order to improve the efficiency of transport systems by the masses of collected data, new exploitation and decision support tools are needed. These tools will help operators in the management of the transportation network and passenger flows by providing them useful information for decision making according to the underling context.

Urban mobility generates a very high number of complex data reflecting the mobility of people, vehicles and objects. It becomes a major stake for stakeholders like cities and transport operators to express the meaning of urban phenomena hidden behind data, with the aim to improve the performance of transport systems and propose new services (Chen et al., 2015) (Andrienko et al., 2017). This paper is part of the IVA (Augmented Travel Information) project, which aims to enhance passenger information by providing digital decision-making tools for passengers as well as for transport operators to improve respectively their traveling experience and their monitoring on transportation network.

The agenda of the paper is the following. Section 2 defines digital mobility data and reviews research works in visualisation of public transport mobility data. This leads to the research question "*How to design a supportive tool for visual exploration of digital mobility data to help a transport analyst in decision making ?*". In section 3, we introduce a problem-solution design approach combined to data visualisation process to create a support tool for a transport analyst. This approach is illustrated through a case study of a commuter line of the Paris metropolitan area. Section 4 presents a common thread starting from the primary functionalities to one of the scenario usages. Finally section 5 discusses the limits of the research approach, the outcomes for transport operators and potential extensions to other design application fields, as well as perspectives of development.

# 2 DEALING WITH DIGITAL MOBILITY DATA

This section first defines traffic data and presents the challenges for transport operators in using massive data. A second part is devoted to contributions in the field of data visualisation for public transport data.

## 2.1 Challenges in using digital mobility data

The increase in the amount of data collected in the transport domain and the development of data streaming technologies can greatly benefit mobility studies and create high value-added mobility information for passengers, data analysts, and transport operators.

Traffic data are defined as "*datasets generated and collected on moving vehicles and objects*" (Chen et al., 2015). Traffic data obtained from equipment and sensors fall into three categories: trajectories (i.e. positions recorded at each time step; incident logs (why, where and when the incidents occur, for instance on a highway, metro or a tunnel) and other type of data (for instance velocity, direction, acceleration of objects). Mobility-related data encompass diverse data related to people dynamics and passenger flows, their habits and interaction with the surrounding environment (survey, air pollution, social network …) (Sobral et al., 2019). In the domain of public transport, experts may be overloaded by the large quantity of operational data to be statistically analysed (Dimanche et al., 2017). In such a context, data visualisation approaches seem to be very relevant (Chen et al., 2015) (Andrienko et al., 2017).

However, these multiple data sources show a high heterogeneity due to their variability in type, format and scale (space and time). Indeed, they could be a mix between structured and unstructured format (CSV, Json…), and target different temporal (timestamp, time slice, daily…) and spatial scales (station, hub, line, network…). This require an additional effort in order to setting them into the same scales.

The availability of such digital mobility footprint provides an opportunity to develop innovative decision making tools for urban stakeholders (operators and authorities) that could allow them to better understand and predict passenger flows in a large city and to improve levels of service and the scheduling of the transport. This is the purpose of this work, i.e. propose an analyst tool for decision making for transportation co-regulation both for traffic and passenger flows.

### 2.2 Data visualisation for digital mobility data

The following review is more specifically addressing the issue of visualisation for public transport data. Visualizing traffic data can support four different types of tasks (Chen et al., 2005):

- Monitoring traffic situations like real time congestion phenomena,
- Discovering patterns and clustering individual trajectories,
- Exploring situations and prediction,
- Planning routes and making recommendations for traffic regulation.

Chen et al. (2015) proposed a conceptual pipeline of traffic data visualisation in four steps. It starts with raw data collection, pre-processing, followed by the choice of visual symbols (diagram types) and graphic visualisation (animation, colour). There are multiple possibilities for representing massive transport data, including the offer (vehicles or trains) and the demand (passengers).

Zeng et al. (2013) paid a special attention to represent, in an interactive manner, passenger flows connecting in Singapore metro, which is composed of 4 lines and 89 stations. By choosing a chord representation (coined 'circos diagram'), the authors emphasized the repartition of passenger flows across the day on the four lines. The busiest station at peak hour clearly appears on the graphs. Several scales are represented: station, metro network and regional area.

Barry and Card (2014) used GTFS (General Transit Feed Specification) data from Boston metro lines, this metropolitan system being one of the busiest in the USA. They produced six types of representations to better understand the train circulation, how passengers use trains and which interactions occur between trains and passengers. Three time scales are available: day, week and month.

(Nagel et al., 2014) created an application aimed to ordinary citizens and transport experts, to enable the exploration of transit data at Singapore, either to improve personal traveling or for planning and monitoring in real-time in the case of experts. They provided three types of tempo-spatial views (maps, time-series and arc-views). The multi-touch support was tested and evaluated with 27 participants at an exhibition, showing different levels of insights and interactions. Finally the design process included visualisation experiments and discussions with experts and test users to feed the system.

Itoh et al. (2016) seeked to explore causes and consequences of unusual phenomena (e.g. earthquakes, storms, accidents, large public events) in Tokyo metro thanks to a 3D exploration system. The system provided views of passenger flows (based on smart card data) and tweets to account for the effects in the passengers' voice. Finally Zhao et al. (2020) created a visualisation system to explore patterns of crime or detect missing passengers in Beijing metro.

### 2.3 Synthesis

Table 1 proposes an overview of the visualisation focus of each contribution examined in 2.1. We emphasize that user-centred approaches for exploring massive transportation data are scarce. On the one


https://doi.org/10.1017/pds.2021.109 Published online by Cambridge University Press

hand the study from (Nagel et al., 2014) is interested in the design and interactions with transit data views for different user groups (i.e. the user experience). On the other hand Zhao et al. (2020) tested the usability of their system along two experiments and evaluated the intuitiveness, interactiveness and usability of the system. From the above survey, we find out that there are few works that put the operator analyst as a central user of the underlying system. This constitutes for us an opportunity to instil driving principles of design into the development of data exploration approaches in the field of mobility and transport. We hence investigate the following research question "*How to design a supportive tool for visual exploration of digital mobility data to help a transport analyst in decision making ?*".

*Table 1: Synthesis of data visualisation studies for public transport*

| Reference | Context | Trains | Passen-gers | User experience | Focus of visualisation |
|---|---|---|---|---|---|
| Zeng et al. (2013) | Singapore metro | | x | NO | Passenger flows and interchange patterns |
| Barry and Card (2014) | Boston metro | x | x | NO | Interactions trains-passengers |
| (Nagel et al., 2014) | Singapore metro and bus | x | x | YES | Interactive access to data for experts and non -experts |
| Itoh et al., (2016) | Tokyo metro | | x | NO | Connecting passengers and tweets |
| Zhao et al., 2020 | Beijing metro | | x | YES | Patterns of crimes and missing commuters |

# 3 BUILDING A DESIGN APPROACH FOR THE VISUAL EXPLORATION OF PUBLIC TRANSPORT DATA BY AN ANALYST

## 3.1 Research approach

For this research, we combined a traditional problem-solution design process (Cross, 2008) with the pipeline of traffic visualisation data proposed by Chen et al. (2005) (Figure 1). This enables to design a support for a visual exploration of data, from the setting of requirements to the realization of a functional mock-up. The user-centred approach started by the observation of on-site operators. The backbone of the work is composed of four creative sessions facilitated by the first author of the paper. The first workshop was especially important to surpass cognitive blocking and fixation thanks to visual stimulation as defined in engineering design, see (Vasconcelos and Crilly, 2015). To this end, participants were stimulated by a mosaic of most widespread representations in data visualisation. Lastly for building a just-necessary usage scenario, we adopted a user journey map representation.

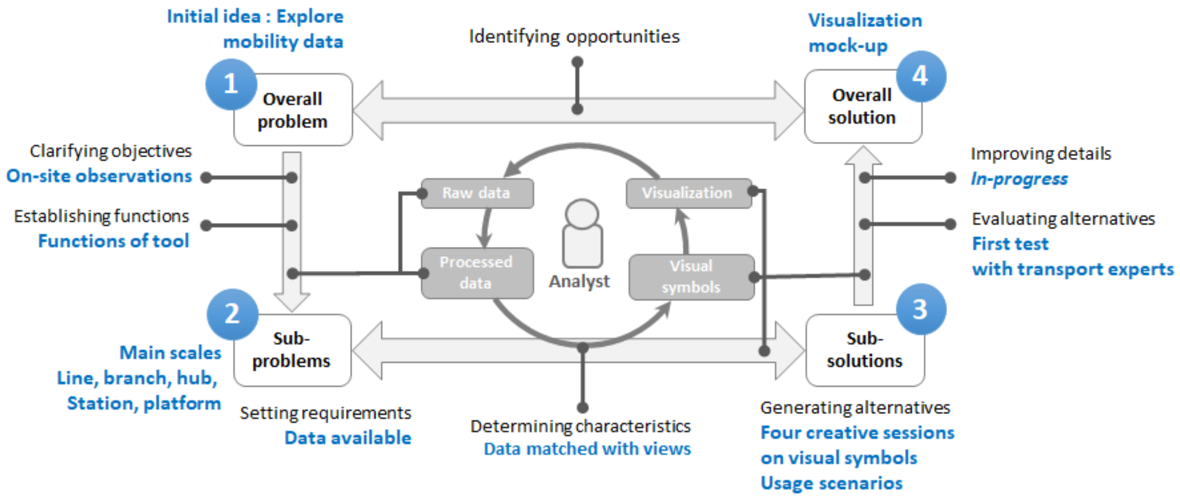


*Figure 1: Design approach for a visual exploration of data based on Cross (2008) and Chen et al. (2015)*



https://doi.org/10.1017/pds.2021.109 Published online by Cambridge University Press

## 3.2 Case study: Visual exploration tool for public transport data analysis

In this section we focus on the design of front-end of the supportive tool, which in close interaction with back-end, i.e. the data based services which are involved into feed the designed views. They aim to help a transport analyst to understand phenomena in public transport, causes and consequences at different scales.

Concerning the scope of the case study, the data deal with railway line Transilien H which serves the northern area of Paris suburbs and supports the commute of more than 235,000 passengers per day. It serves 46 stations situated along five branches (H1 to H5) where 478 trains are committed on this line per day. A brief description of data sources is exposed hereafter, structured into: (1) Observed and measured data; (2) Predicted and estimated data.

### 3.2.1 *Observed and measured data*

The first category of data are those collected in situ and for which we made some transformations and processing in order to compute measured KPI and indicators. From this collection, we find ticketing data corresponding to validations of smart cards at the AFC (Automatic Fare Collection) systems. These cards are known in Paris metropolitan area as "Navigo pass".

The AFC allows to collect the ticketing logs for bus, train and tram. The dataset consists of aggregation of ticketing logs per 15 minutes. As mentioned above, we only deal with the stations and stops served by the line Transilien H that connects Paris with its northern suburb. The dataset has been provided by the transport organization authority (Ile-de-France Mobilités) that controls and coordinates the different transport companies operating in the Paris metropolitan area.

The second data source is the passenger counting at the boarding, alighting and on board the trains of the studied line. This data is collected by Automatic Passenger Counting (PAC) sensors positioned at the landing doors of the train. The counting is transmitted for each train at the moment of its passage at the station. The third source is the transportation scheduling; i.e. data related to the proposed service that regroups: routes, trips and timetables. For that purpose, we have at our disposal theoretical timetables, as well as realised ones.

The last data set is related to the incidents information regrouping events, planned works, real-time incidents that occurred on the transport network.

### 3.2.2 *Estimated and predicted data*

The second category of indicators are those provided by applying the developed transportation models related to the flow passenger assignment based on (Yao et al, 2017), and forecasting of load on board the trains (Pasini et al, 2019), and the ridership at the station (Toqué et al., 2017). The passenger flow assignment model is the core to estimate the load on the public transport network. To characterize this model, two factors are of the most importance: origin/destination demand matrix and route choice behaviour. The main function is then to assign the passenger flow to the transport network (Yao et al., 2017). The Origin-Destination (OD) matrix describes travel demands from an origin to a destination node. In our case, the first step in Origin Destination (OD) assignment process is to generate all the possible routes for each OD pair, then based on utility function, the utility of each route is computed. Thanks to the assignment process, it is possible to estimate different passenger load indicators such as load in the train, on the platform and in the station. It is also possible to follow the route of one or several groups of passengers from a given station. Finally we perform the prediction of two indicators concerning the passenger flows (Amrani et al., 2020) thanks to machine learning algorithms: the passenger load on trains (Pasini et al., 2019) and the attendance at the station (Toqué et al, 2017), with a long term and a short term prediction. The long term prediction aims to predict one year ahead of time, whereas the short term provides this indicator for the next time interval for station attendance or the next train passage for train load. Furthermore, we encounter these models in the feeding of a predictive journey planner (Amrani et al., 2020).

# 4 RESULTS

## 4.1 Establishing functions and architecture of the visual exploration tool

This section summarizes the main outcomes of the design process between stage 1 (Overall problem) and stage 2 (Sub-problems). The process started with a half-day on-site observation of transport agents

involved in the regulation of the train line under study. Important spatial scales of exploration were emphasized: (1) The line scale encompassing the 46 stations; (2) The branch scale composed of 5 branches H1 to H5, some parts being common to several branches; (3) The hub scale, where hubs are the nine main stations connecting or terminating branches, and also connecting to other modes (rapid train, tram or bus); (4) The station scale; (5) The platform scale, allowing to check the density of passengers.
Besides, the functions reflect the expectations of an analyst to be able to identify an unusual situation (whether planned or unplanned) and also check for alerts. This allowed to formalize a first set of functional requirements which were refined after several iterations (Table 2). Each function is connected to a set of indicators based on observed, estimated or predicted data (see 3.2).

*Table 2: Functions for the visual exploration tool*

| Function | Expression |
|---|---|
| F1 | Represent indicators for trains and stations at line scale |
| F2 | Represent spatio-temporal indicators at line scale |
| F3 | Represent indicators for passengers and trains at branch scale |
| F4 | Represent multimodal indicators at hub scale |
| F5 | Select the train direction |
| F6 | Inform analyst on alerts for passengers and trains |
| F7 | Represent indicators for passenger routes |
| F8 | Inform analyst of an unusual situation compared to a reference |

## 4.2 Generating alternatives: Creative outcomes and conceptual elements

Four main creative sessions were conducted with project members to create (1) a set of views making the most of the available data while integrating the habits and practices of transport analysts; (2) a rationale to navigate between views in an original and structured way. The first session consisted in a broad exploration of modal shift and potential offered by 87 miscellaneous data visualisation graphs (see for instance datavizproject.com), which were triggers to explore the space of visual representations for transport data (Figure 2). During the three following sessions, main visual choices and structure of the tool were defined as given in Table 3: multi-scale view and navigation between scales (session 2); focus on passenger route and modal shift during a disruption (session 3); initial work on anomaly indicators (session 4). Conceptual elements were derived from the four sessions:

- Three structuring panels: Line (showing trains and stations), Branches and Hubs (showing interchanges with other modes); the traditional Scheduling graph displaying trains in time and space for each branch of the line.
- Two types of synoptic maps for the spatial positioning on the left hand side of each view, either extensive or simplified (i.e. a logic grid with hubs as nodes).
- Three types of visual indicators for observed and predicted data: Binary alert, continuous scale for anomaly [-1; +1] and 4-level indicator relatively to a reference. The alert and the anomalies values are computed based on anomaly detection techniques which aim to qualify abnormal behaviour on time series (Chandola et al., 2009). The level of load indicator is defined in regard to the maximum capacity (Amrani et al., 2020) while its reference value is set according to the long-term prediction. This is done using a model based on machine learning that is not influenced by real time information.

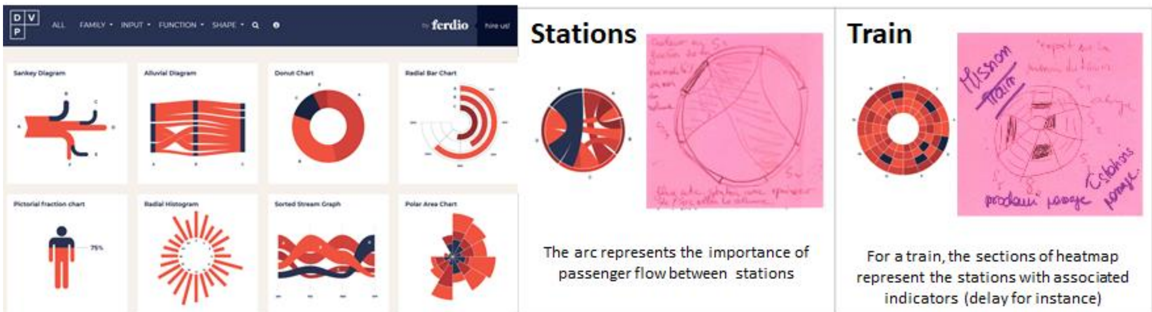


*Figure 2: Examples of outcomes for the first creative session (left: excerpts of visual prompts; right: excerpts from sketches)*


https://doi.org/10.1017/pds.2021.109 Published online by Cambridge University Press

Table 3: Synthesis of outcomes of creative sessions

| Participants | Objective | Outcomes |
|---|---|---|
| 1 facilitator<br>1 web developer<br>1 PhD student<br>4 data engineers | Disrupted situation and modal shift | Proposition of 7 main visual graphs: chord, heatmap, radar, progression bar, proportional areas… |
| 1 facilitator<br>1 web developer<br>2 data engineers | Views on stations, trains, branches | Proposition of 5 additional graphs<br>Three overall panels and normality indicators characterized |
| 1 facilitator<br>1 web developer<br>2 data engineers<br>2 transport engineers | Passenger routes and modal shift et report modal | Creation of a disrupted scenario on branch H2<br>Graphs for passenger flows : Sankey, Sankey alluvial and sunburst; Simplified synoptic map |
| 1 facilitator<br>2 data engineers | Network anomaly | Proposition of anomaly heatmap (load, delay) |

## 4.3 Generating alternatives: User scenario for a transport analyst

The generation of alternative views for the tool was driven by the usage expectations of the analyst, which was captured by means of a user scenario (Figure 3). This exploration timeline allowed to introduce the tool to 15 transport engineers to gain first feedbacks.

The main exploration mode is the following: Analyse data from past periods through the exploration to discover trends and explain phenomena while screening the multiple spatio-temporal scales. To build the user scenario, the incident logs was explored over the time period 2016-2018 (data availability) and highlighted several major incidents impacting the train line. We illustrated the scenario on day 7-12-16: eight incidents were recorded, among which one major incident at Paris-Nord station (PNB) impacting more than 200 trains until the end of the day (cancelled trains, major delays and modified routes).

Figure 3 illustrates a scenario where the analyst wishes to explore a specific disrupted day in a targeted manner. It also points out how Fi functions (see Table 2) are tackled. After selecting the target day on the calendar and the hour of the day, the analyst opens the Scheduling panel to check the approximate number of trains and stations impacted by this major incident. To have a more detailed perception of delayed trains, he/she opens the Line panel, choses Paris-Nord hub on the synoptic graph and screens the train indicators from 13:00 until the end of the day. Moving to the sunburst diagram, he/she is able to dynamically visualize the flow of passengers at each hub node (in trains and on platforms) and also those leaving the train line. By this unique graph, it is possible to follow the flow on each branch. Finally he/she can move to the modal shift view for each node hub to examine histograms related to the connecting modes as detailed hereafter.

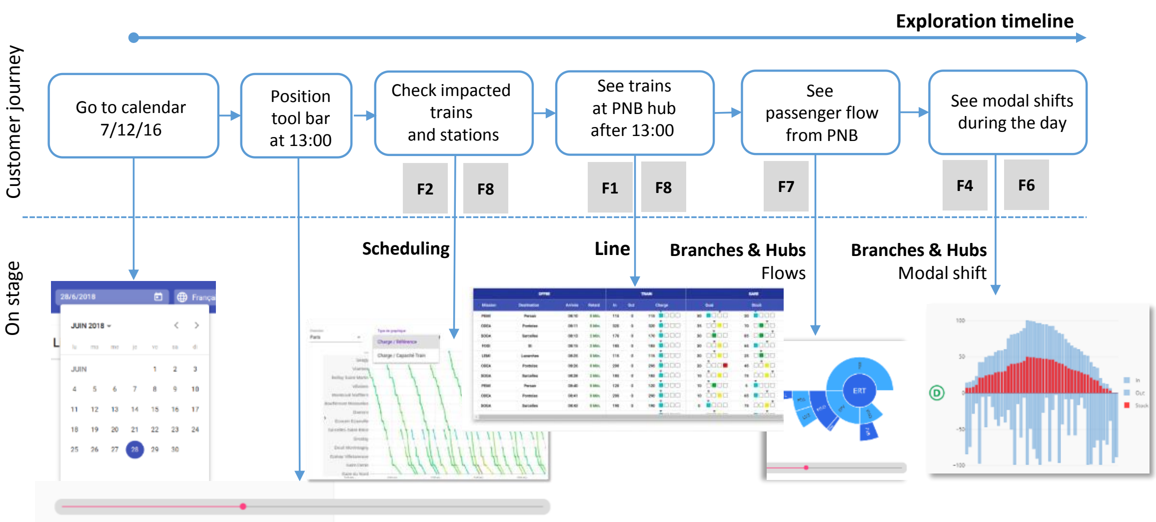


Figure 3: User scenario for exploring a disrupted situation (PNB: Paris-Nord hub)



https://doi.org/10.1017/pds.2021.109 Published online by Cambridge University Press

Figure 4 provides an example of a screen view from the Branches and Hubs panel, reflecting modal shifts at each hub, here Epinay Villetaneuse (EPV) located on the simplified synoptic. This hub is connected to train line D, bus and tramway line T. The graph type is a histogram on the right hand side displayed for the day (time stamp one hour, zoom in 15 minutes). Indicators are estimated flows of passengers entering (respectively exiting) line H through D, bus and T (labelled In, respectively Out). Additionally estimated stock at the hub is superimposed to in-out flows.
In the shoes of the analyst, this screen can help answering two sorts of practical questions:

- Knowing that a train was cancelled or that planned works occurred, I would like to know how the passenger flows shifted to the connecting rapid trains, the tramway lines or buses, and if problems of overload could be noted at hubs.
- After an incident, I would like to appraise the return to normalcy along the day for a given hub through the examination of flows entering and exiting line H.

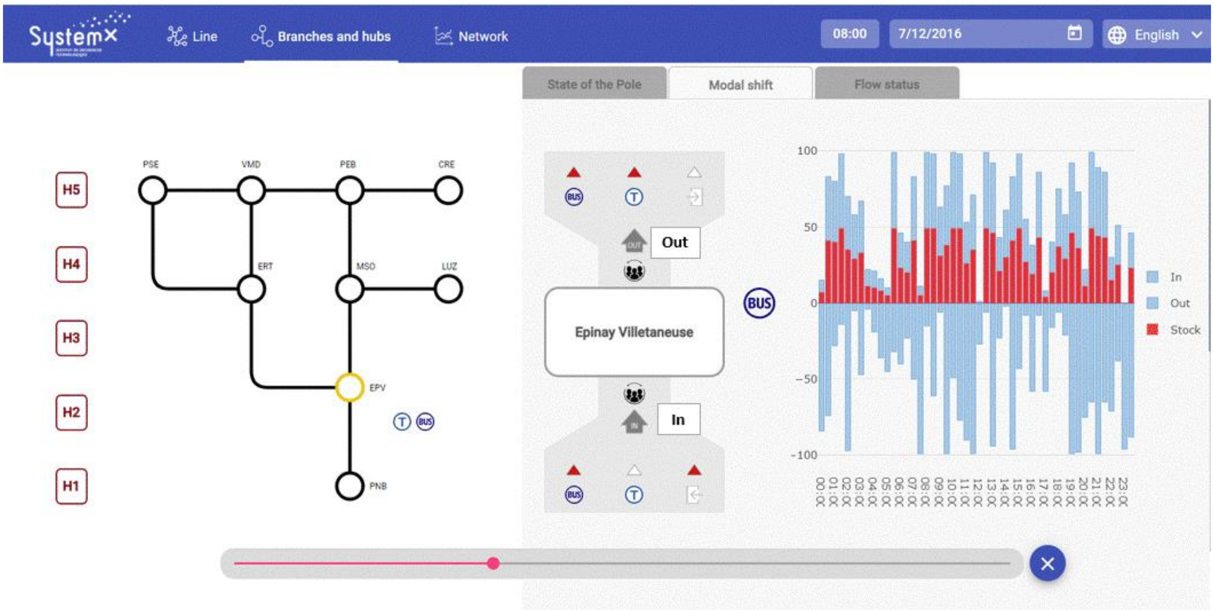


*Figure 4: Screenshot from branches & hubs panel / modal shift*

## 5 CONCLUSION AND FUTURE WORK

The aim of the paper was to investigate "*How to design a supportive tool for visual exploration of digital mobility data to help a transport analyst in decision making ?*" . The literature review showed that data visualisation studies usually miss to reflect on their design process clarifying the provided user experience.
Building upon a classic design process, we proposed to integrate a data visualisation pipeline to develop an exploration tool for a transport analyst. The design process was fed by four extensive creative sessions with multidisciplinary teams, prompted by 87 data visualisations graphic representations. The originality of the approach is to make explicit the potential expectations of a transport analyst to explore past disrupted situations on the network. The exploration jointly occurs for trains and passengers, systematically articulating spatial and temporal scales (from network to branch, and stations). The introduction of an intermediate scale of branches and hubs is also new to our knowledge.
Some limits of this work concern experiments with transport experts. If the tool mock-up was already introduced to a panel of experts, a full experiment is a remaining task in the short term, as in protocols delivered in (Nagel et al., 2014) or Abi-Akle et al. (2016). A deployment has to be organized with experts to enable interactive tests. Further exploration of the characterization of anomaly scores are an additional undergoing work. The usage perspective of the tool are multiple, for instance helping to handle more complex situations of regulation or assist transport planning. By exploring previous sets of data, the transport analyst could infer best choices for future transport plans if there are disturbances to come (for instance strikes or planned works).


https://doi.org/10.1017/pds.2021.109 Published online by Cambridge University Press

The exploitation of data and visualization techniques through the proposed tool allows the analyst to understand and analyse complex situations, especially in case of an incident. This tool has an important potential in the context of decision support by coupling it with a transport simulator in order to evaluate several recovery strategies starting from real situations.
Other fields of application of this data design approach may be envisaged. It can be generalized to road or air traffic data but also to traffic in general, for instance detecting abnormal network traffic activities (Ji et al., 2020). For air traffic, multiple data are available form airports such as ticketing data, take-off and landing time, delay, cancellation etc. The analysis of data is expected to help decision making for controlling the airline company fleets as well as for better management of the airports for the scheduling and rotation of the their different working teams.
Finally the proposed methodology is compliant with the data-driven smart factory to give domain experts the possibility to analyse and optimise the production chain taking into account production facilities and related performances (Tao et al., 2018).

## ACKNOWLEDGMENTS

This research is a part of the IVA Project, which aims to enhance traveler information and is carried out under the leadership of the Technological Research Institute SystemX, with the partnership and support of the transport organization authority Ile-De-France Mobilités (IDFM), SNCF, and public funds under the scope of the French Program “Investissements d’Avenir”.

## REFERENCES

Abi Akle, A., Minel, S. & Yannou, B (2017), “Information visualization for selection in Design by Shopping”, *Res Eng Design*, Vol. 28, pp. 99–117. https://doi.org/10.1007/s00163-016-0235-2
Amrani, A., Pasini, K., Khouadjia M. (2020), “Enhance Journey Planner with Predictive Travel Information for Smart City Routing Services”, *Forum on Integrated and Sustainable Transportation Systems* 2020, Delft.
Andrienko, G., Andrienko, N., Chen, W., Maciejewski, R. and Zhao, Y. (2017), “Visual Analytics of Mobility and Transportation: State of the Art and Further Research Directions”, *IEEE Transactions on Intelligent Transportation Systems*, pp.99, https://dx.doi.org/10.1109/TITS.2017.2683539
Barry, M, Card, B. (2014), “Visualizing MBTA Data - An interactive exploration of Boston’s subway system”, Available at http://mbtaviz.github.io/#trains (Accessed 01/09/19).
Beghelli, A., Huerta-Cánepa, G. and Segal, R. (2019), “Data Materialisation: A New Undergraduate Course for a Data Driven Society”, Proceedings of the Design Society: International Conference on Engineering Design. Cambridge University Press, Vol. 1(1), pp. 2061–2070. https://dx.doi.org/10.1017/dsi.2019.212.
Chandola V, Banerjee A, Kumar V (2009) Anomaly detection: A survey. ACM computing surveys (CSUR) 41(3), pp.1-58.
Chen W., Guo F., Wang F.Y. (2015), “A Survey of Traffic Data Visualization”, *IEEE Trans. Intell. Transp. Syst.*, Vol.16, pp. 2970–2984. https://dx.doi.org/10.1109/TITS.2015.2436897
Cross, N. (2008), Engineering design methods: strategies for product design. Chichester, John Wiley.
Daraio, C., Diana, M., Di Costa, F., Leporelli, C., Matteucci, G. and Nastasi, A. (2016), “Efficiency and effectiveness in the urban public transport sector: A critical review with directions for future research”, *Eur. J. Oper. Res.*, Vol. 248, No 1, pp. 1-20, https://dx.doi.org/10.1016/j.ejor.2015.05.059.
Dimanche, V., Goupil, A., Philippot, A., Riera, B., Urban, A., Gabriel, G. (2017), “Massive Railway Operating Data Visualization; a Tool for RATP Operating Expert”, *IFAC-PapersOnLine*, Vol.50, Issue 1, pp. 15841-15846.
Gkiotsalitis, K., Wu, Z. and Cats, O. (2019), “A cost-minimization model for bus fleet allocation featuring the tactical generation of short-turning and interlining options”, *Transp. Res. Part C Emerg. Technol.*, Vol. 98, pp. 14-36, https://dx.doi.org/10.1016/j.trc.2018.11.007.
Han, J., Forbes, H., Shi, F., Hao, J. and Schaefer, D. (2020), “A data-driven approach for creative concept generation and evaluation”, Proceedings of the Design Society: *DESIGN Conference*. Cambridge University Press, Vol.1, pp. 167–176. https://dx.doi.org/10.1017/dsd.2020.5.
Itoh, M., Yokoyama, D., Toyoda, M., Tomita, Y., Kawamura, S., Kitsuregawa, M. ( 2016), “Visual Exploration of Changes in Passenger Flows and Tweets on Mega-City Metro Network,” *IEEE Transactions on Big Data*, Vol. 2, No. 1, pp. 85-99, https://dx.doi.org/10.1109/TBDATA.2016.2546301.
Ji, S.Y., Jeong, B.K., Jeong, D.H. (2020). “Evaluating visualization approaches to detect abnormal activities in network traffic data”. International Journal of Information Security. https://doi.org/10.1007/s10207-020-00504-9
Nagel, T., Maitan, M., Duval, E., Vande Moere, A., Klerkx, J., Kloeckl, K., and Ratti, C. (2014), “Touching transport - a case study on visualizing metropolitan public transit on interactive tabletops”. *Proceedings of*



https://doi.org/10.1017/pds.2021.109 Published online by Cambridge University Press

*the 2014 International Working Conference on Advanced Visual Interfaces* (AVI ‘14). Association for Computing Machinery, New York, NY, USA, pp. 281–288. DOI:https://doi.org/10.1145/2598153.2598180

Pasini K., Khouadjia M., Samé A., Ganansia F., Oukhellou L. (2019), “LSTM Encoder-Predictor for Short-Term Train Load Forecasting”. In: Brefeld U., Fromont E., Hotho A., Knobbe A., Maathuis M., Robardet C. (eds) *Machine Learning and Knowledge Discovery in Databases*. ECML PKDD 2019. Lecture Notes in Computer Science, Vol. 11908. Springer, Cham. https://doi.org/10.1007/978-3-030-46133-1_32

Salminen, J., Jung, S.G., Kamel, A. M., Santos, J. M., Kwak, H., An, J., and Jansen, B. J. (2020) Using Artificially Generated Pictures in Customer-facing Systems: An Evaluation Study with Data-Driven Personas. Behaviour & Information Technology. https://dx.doi.org/10.1080/0144929X.2020.1838610

Stevenson, P. D. and Mattson, C. A. (2019), “The Personification of Big Data,” Proceedings of the Design Society: *International Conference on Engineering Design*. Cambridge University Press, Vol. 1(1), pp. 4019–4028. https://dx.doi.org/10.1017/dsi.2019.409.

Sobral, T., Galvão, T., & Borges, J. (2019), “Visualization of Urban Mobility Data from Intelligent Transportation Systems”. Sensors (Basel, Switzerland), Vol. 19(2), 332. https://doi.org/10.3390/s19020332

Tao, F., Qi, Q., Liu, A., Kusiak, A. (2018). “Data-driven smart manufacturing”. Journal of Manufacturing Systems, Vol. 48, pp.157-169. https://dx.doi.org/10.1016/j.jmsy.2018.01.006

Toqué F., Khouadjia M., Come E., Trepanier M., Oukhellou L. (2017). “Short long term forecasting of multimodal transport passenger flows with machine learning methods”, *IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 560-566, https://dx.doi.org/10.1109/ITSC.2017.8317939.

Vasconcelos, L., & Crilly, N. (2016), “Inspiration and fixation: Questions, methods, findings, and challenges”, *Design Studies*, Vol. 42, pp.1-32. https://doi.org/10.1016/j.destud.2015.11.001.

Yao, X., Han, B., Yu, D., & Ren, H. (2017). Simulation-Based Dynamic Passenger Flow Assignment Modelling for a Schedule-Based Transit Network. Discrete Dynamics in Nature and Society, 2017, 1-15.

Zhao X. et al. (2020), “Interactive Visual Exploration of Human Mobility Correlation Based on Smart Card Data,” *IEEE Transactions on Intelligent Transportation Systems*, https://dx.doi.org/10.1109/TITS.2020.2983853.

Zeng, W., Fu, C.-W., Arisona, S.M. and Qu, H. (2013), “Visualizing Interchange Patterns in Massive Movement Data”, *Computer Graphics Forum*, Vol. 32, pp. 271-280. https://doi.org/10.1111/cgf.12114