8.

Decentralised open data publishing for the public transport route planning ecosystem

Julián Rojas, Bert Marcelis, Eveline Vlassenroot,
Mathias van Compernolle, Pieter Colpaert & Ruben Verborgh

Open data initiatives have created a revolution in the route planning ecosystem for the public transport sector. The creation of a large amount of route planning services like Google Maps, CityMapper or Navitia, has only been possible thanks to the availability of public transport data as open data. Ever since the disclosure of the London public transport data sources as open data (Hogge 2016) more public transport companies are following their lead around the world. The benefits obtained by disclosing public transport datasets as open data are diverse and influence the different actors present in the route planning ecosystem: public transport organisations in the role of data publishers for instance may increase their revenue streams as new and better information channels attract more travellers (UK Department for Transport et al. 2018). Also, new analysis and improvements to their operations become possible through feedback received from data reusers on areas where they do not collect data by themselves (e.g. crowdsourced data).

For common travellers the benefits are reflected on a more diverse service offer that covers a wider range of functionalities and facilitates ubiquitous access to public transport data through mobile applications. For example, the GoOV[1] application in the Netherlands provides support for anyone who has trouble travelling independently throughout the public transport network, like people with disabilities or seniors. The application relies on public transport open data to guide its users and provides a service to a more specific target group that was not offered before by the public transport operators in the Netherlands. The release of public transport datasets as open data has proven also to be a

---

1   http://www.go-ov.nl/

catalyser for innovation and an economy booster, as revealed by a study on the impact of opening up public transport datasets in London (Deloitte 2017). Over 13 000 registered developers or reusers have contributed to the creation of more than 600 applications that rely on the open data, reaching 42% of London's population and providing innovative commercial and non-commercial customer-face solutions that can tackle social and economic issues. This contributes to the digital economy of the city with an estimate of 500 direct and 230 indirect jobs and an estimated total gross value add from these companies, directly and across the supply chain and wider economy of £14 million per year (Deloitte 2017). Finally, open public transport data represent a valuable source of information for public authorities and NGOs who may use it during decision-making processes (e.g. urban planning) and for independent analysis and studies where public transport is relevant (Share-PSI 2016).

The existence of open data provides a continuum of value. The final parts of the value chain, which involve extracting meaning from data and applying it to address a particular matter, are as important as the earlier parts, which involve data collection, storage and publication (Van Schalkwyk et al. 2017). From a technical perspective in the public transport sector, the way open data is published directly influences the architectural design of route planning applications, which in turn affects the technical decisions that data reusers need to make when using open data.

On one hand, public transport operators may choose to share their data through Remote Procedure Call (RPC) APIs. In the public transport environment, an example of an RPC API is one that receives requests containing a set of parameters (e.g. origin, destination, departure time, etc.) from a remote client, such as a mobile or web route planning application, and uses them to calculate route alternatives over a transport network. Besides routes, RPC APIs could also allow reusers to access information about other related entities (e.g. stops, vehicles, departures, etc.) that can be integrated in their applications. However, with this approach, operators often impose querying limitations to reusers due to the associated computational costs that will increase as the amount of reusers grows. Such limitations go against the idea of open data, the proponents of which advocate for full and unlimited access to data. Furthermore, reusers are not able to influence the route planning algorithms to include new features (e.g. wheelchair accessibility, foldable bikes, shared cars, etc.) as these are perceived as black-boxes from the reusers' perspective.

On the other hand, operators can share their entire datasets using standard formats like General Transit Feed Specification (GTFS)[2] which third parties can integrate and reuse in their applications. Such a data dump approach fosters the creation of centralised data silos, as route planner developers need to process and host the entire dataset of every public transport network over which

2   https://developers.google.com/transit/gtfs/

they want to provide their service. Data silos are the result of data integration processes, where it is first necessary to align and reconcile data entity identifiers coming from different data providers, in order to enable route planning queries. For applications that ultimately want to provide a world-wide route planner, this means an immense investment on computational infrastructure.

Considering these approaches and their limitations, the Linked Connections[3] (LC) specification was introduced. LC aims at offering an in-between solution, that is between the RPC API supporting any type of query strategy and the data dump containing all data approach, that allows operators to share data in a cost-efficient way and that is optimised for performing route planning algorithms. By modelling transport networks as a list of vehicle departures and arrivals, sorting them in a timely fashion and publishing them as data fragments, reusers are able to request specific parts of a transport network dataset over which they can calculate a specific route on the fly. LC follows the Linked Data principles[4] by assigning unique identifiers to every element of a public transport network and relying on common semantic vocabularies to provide a description to each of them. This is intended to increase the interoperability of public transport datasets which reduces at the same time the adoption costs of data for open data reusers.

The approach of fragmenting datasets was taken from the linked data fragments concept (Verborgh et al. 2014) which allows for the definition of specific types of fragments of linked data datasets that can be generated with minimal effort by servers, while still enabling efficient client-side querying. This constitutes a decentralised solution as reusers can now directly request specific data fragments from different public transport operators that are distributed on the web and execute route planning algorithms just in time on the client side, therefore reducing data hosting and integration costs. Furthermore, LC allows clients to cache fetched data fragments in memory, enabling offline execution of new queries, which is not possible on RPC based solutions.

The LC framework also provides a solution for route planning that supports *privacy-by-design*. Since the route planning calculations can be performed on the client side, the users are not required to share the details of their queries with third party servers. Previous research (see Colpaert et al. 2017) has proven that from a scalability and cost-efficiency perspective for hosting the data, Linked Connections outperforms traditional RPC based route planning approaches. This is an advantage for data publishers as they are able to provide data to more reusers with lower operational costs. However, it is still unclear how an approach such as LC will impact other actors in the route planning ecosystem for public transport (e.g. reusers, common travellers).

Roy Fielding (2000) introduced Representational State Transfer (REST) in his PhD thesis while standardising the web's HTTP/1.1 protocol. REST is a

---

3   http://linkedconnections.org/
4   https://www.w3.org/DesignIssues/LinkedData.html

set of architectural constraints for large data architectures. Each constraint one decides to follow in a web architecture will return benefits such as scalability of the server, visibility, cost-efficiency, reliability and – also – *the user-perceived performance*. For instance, following the caching constraint on both client and server sides results in (i) more scalable servers by reducing the amount of requests that servers need to process, as the number of clients increases; and (ii) improved user-perceived performance by allowing clients to keep and reuse relevant data from memory instead of requesting it from the server every time it is needed, which is significantly slower. In this chapter we particularly zoom in on the *user-perceived performance* property and study it within the context of open data for route planning purposes.

The main research question we address in this work is: what is the impact on the actors that belong to the route planning ecosystem for public transport, of implementing an open data publishing approach as the Linked Connections framework? In this work we present and discuss an analysis of such effects. We also present a study that evaluated the technical performance of an LC based application that executes its route planning algorithm on the client side compared to a traditional application that relies on a RPC route planning API running on the server-side to determine what kind of considerations developers and data reusers must take into account when working with this approach.

Furthermore, we present the results of a user perceived performance study where 17 different regular public transport travellers tested both applications for different use cases and selected one or the other as their preferred choice based on perceived performance and provided features, in order to determine the effects of the LC approach on common public transport riders and their perception of it. For this we developed an isomorphic Android application that implemented both approaches and provided users with the same interface in both cases. In the next session we describe the open data route planning ecosystem for the public transport sector and the different actors that comprise it. Then we present a description of the methodology followed during the performed studies. The results obtained during the evaluations are presented afterwards. Finally, a discussion of the main findings is presented along with the correspondent conclusions.

The route planning ecosystem

The use of the ecosystem analogy in relation to business practices has become notably strong. Related literature defines digital ecosystems as cyclical, sustainable, demand-driven environments orientated around agents of a different nature who are mutually interdependent (Heimstädt et al. 2014). Scholars in information intensive environments have used the term to focus on the multiple and varying interrelationships between providers, users, data, infrastructure and institutions (Harrison et al. 2012). For open data route planning on the

public transport sector we devise an ecosystem as shown in Figure 1, where the different actors that benefit from and support open data, are represented.

Figure 1. Open data route planning ecosystem
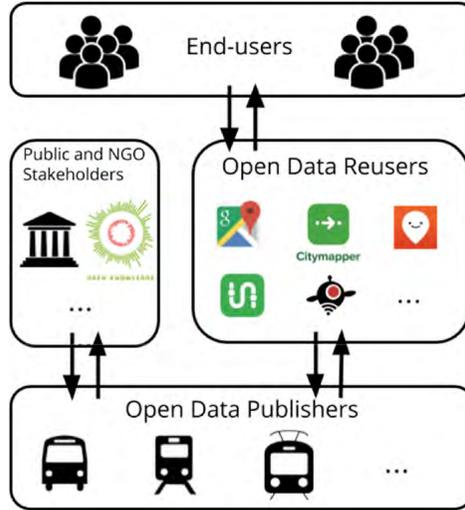
The first type of actors that can be identified in Figure 1 are the *open data publishers*. On a route planning ecosystem these correspond to the public transport companies which operate on a transport network infrastructure (e.g. bus, train, tram, metro, etc.) and produce related data (e.g. timetables, list of stations, live updates, etc.). They publish the data as open data in machine readable formats that allow its adoption and reuse by third parties. The GTFS specification, as the *de facto* standard, is commonly the selected format to publish and share the data.

The *open data reusers* reference every company and/or organisation that consumes and integrates open data for solving route planning queries over one or more public transport networks. Here we can find companies such as Google (Maps), CityMapper, Moovit, GoOV and many others of the sort, that collect and integrate public transport-related open data to offer route planning services on top of it. Public authorities also reuse public transport open data to offer route planning services as a mechanism to improve the mobility conditions of their regions and cities (Rode et al. 2015; Ahlers et al. 2018). Public authorities in cities such as Portland[5] (US), Antwerp[6] (Belgium) and London[7] (UK) can be considered as examples of open data reusers.

_____

5   https://trimet.org/#/planner
6   https://www.slimnaarantwerpen.be/en/home
7   https://tfl.gov.uk/plan-a-journey/

The *end-users* on a route planning ecosystem are the public transport travellers. These actors use the services offered by open data reusers through web or mobile applications to navigate the public transport networks and satisfy their mobility needs. Lastly, the *Public and NGO Stakeholders* normally do not have direct contact with the data flow through the route planning ecosystem (except when acting as data reusers) but can influence how the data is shared among its actors. Public authorities, for instance, define the legal framework and constraints of data sharing processes, and other types of organisations such as research institutes and non-profit agencies can contribute to the definition of standards and procedures that impact the way open data is shared and consumed.

This ecosystem constitutes the analytical framework of our work. Its definition has been made based on an empirical mapping of real-world route planning-related scenarios, by observing the relationships and interactions of the actively involved organisations. Determining how the different actors that comprise it are affected by the implementation of a decentralised open data publishing strategy represents a contribution to the open data community on the public transport sector by shedding light on the merits and also on the open challenges of the aforementioned approach.

Methodology

In this section we describe the methods used for assessing the impact of implementing a decentralised open data publishing strategy within the route planning ecosystem for public transport and its actors. Having identified the different actors that play a role in the ecosystem, in this work we focused specifically on assessing the impact on open data reusers and end-users.

Open data reusers

For conducting the impact evaluation on open data reuser tests we developed an isomorphic Android application[8] that implemented an LC based route planner and a client for an RPC based API hosted on a remote server (see Figure 2). The evaluation consisted of a series of performance tests that were conducted for both the LC and the RPC API based implementations. We used the same algorithm and datasets in both cases to allow a fair comparison. The chosen algorithm was the Connection Scan Algorithm (CSA) which was designed to operate over similar data structures as the one defined by the LC specification, making it a perfect fit. Moreover, previous research (see Dibbelt et al. 2018) has proven the CSA algorithm to be more efficient for public transport route planning than traditional route planning solutions based on variants of Dijkstra's algorithm (Dijkstra 1959).

---

8   Available at https://github.com/Bertware/masterthesis-LC-LC2Irail-android-client/releases/
     tag/RESEARCH40

Therefore, we propose the following hypothesis:

(H1): LC based implementations perform better with regard to response time when compared to traditional RPC API based implementations.

For the performance tests we defined a set of use cases to be tested on both implementations. These consisted of queries that requested routes (going from A to B at a given departure time), liveboards (the list of scheduled arrivals and departures at a given stop) and vehicles (the sequence of stops and scheduled times for a given vehicle). The dataset used for the evaluation is the public train transport network of Belgium operated by the NMBS company.[9] The set of queries used to test the different use cases were taken from real world requests made to the servers of the iRail API.[10] Both implementations were tested on two different smartphones with different hardware capabilities: the HTC One (Android 5.0) and the HTC 10 (Android 8.0) smartphones.

Figure 2. HyperRail: Isomorphic route planning mobile application
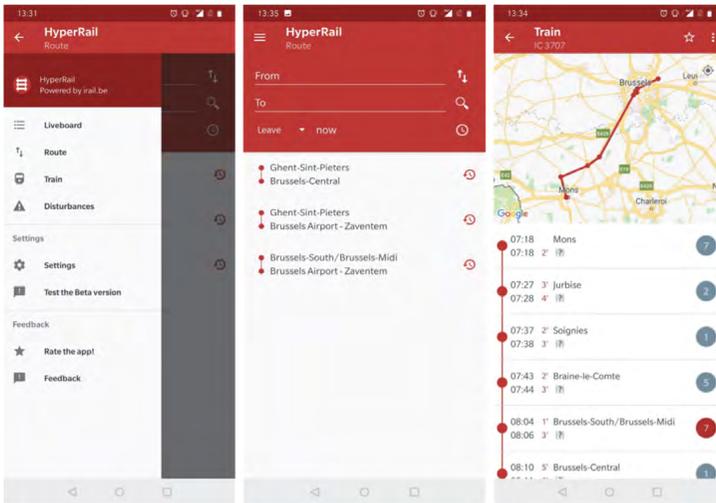
Open data end-users

The LC approach publishes raw public transport schedules as data fragments allowing route planning application clients to request only the required data to solve specific queries. Clients can cache these data fragments in memory

---

9   NMBS Linked Connections available at https://graph.irail.be/sncb/connections
10   The iRail project: https://hello.irail.be

and reuse them to solve future queries, accelerating the process of solving route planning queries. Considering this, we propose the following hypothesis:

> *(H2): The LC based route planning application has a higher user perceived performance when compared to traditional RPC API based implementations.*

The evaluation of the impact on end-users was carried out through a user perceived performance test (n=17) and a questionnaire (n=65). For the user perceived performance test we used the same isomorphic Android application used for the performance evaluation for open data reusers, as well as the same defined use cases and dataset. Each user was asked to execute a set of queries for each use case using both the RPC API based and the LC based approaches. Then for each use case, every user was asked to provide their opinion on which alternative they perceived to perform better. Furthermore, once the users completed testing each use case, they were asked to activate the airplane mode of the smartphone and to run the queries again for both approaches. This was done in order to show the users that the LC based approach could also solve queries while being offline.

Additionally, all the users had to decide on which approach they preferred the most, taking into account the performance they perceived during the different use cases and also the additional features, such as offline querying and privacy safeguarding. To conclude, the users received an explanation regarding the capabilities of the LC approach about privacy (where the details of their queries were not being sent to a remote server), speed (results can be shown quicker because LC can load and reuse information), offline querying and flexible route planning.

Findings

Open data reusers

> *(H1): LC based implementations perform better with regard to response time when compared to traditional RPC API based implementations.*

Figure 3 shows the results of the evaluation performed when querying for routes. The median response times are depicted for both approaches running on the HTC 10 and HTC One smartphones. In the HTC One, the LC approach performs 29% faster (~1s) than its counterpart in the HTC 10. However, for the HTC One the performance of LC deteriorates being 57% slower (~2s) than the RPC API. This behaviour can be explained by the fact that the HTC One smartphone has inferior hardware capabilities which impacts the execution of the route planning algorithm on the device. Also, as expected, the performance of the RPC approach is consistently similar on both devices as the algorithm is executed on the server-side.

Figure 3. Number of results in function of loading time (routes) – LC vs RPC API
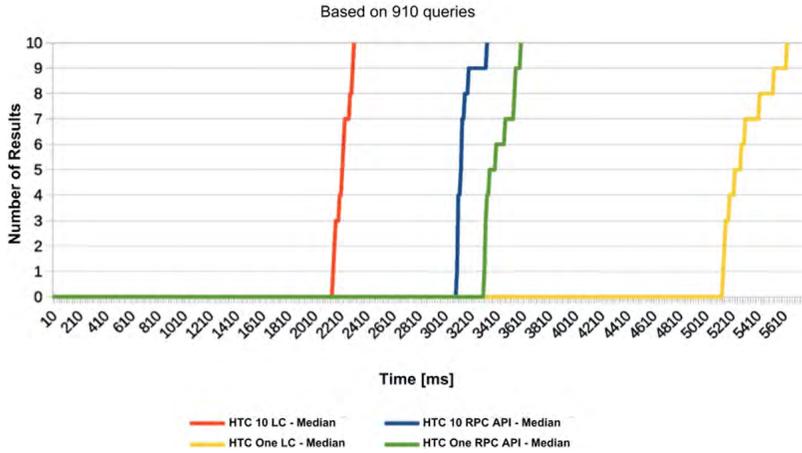
Figure 4 shows the results of the evaluation performed for *liveboards* queries. The median response times are depicted for both approaches running on the HTC One and HTC 10 smartphones. The incremental results make clear that there are differences between both architectures. Linked Connections is faster than the RPC API based approach on both devices for the first few results. However, when ten results are needed (about the number of results which fit on a large screen) the RPC API is faster on both devices. It is clear that the RPC API performs similarly on both devices, but the LC client does not. There is a gap between the time needed by the client-side LC implementation on both devices, which grows with the number of results needed.

Figure 4. Number of results in function of response time (liveboards) – LC vs. RPC API
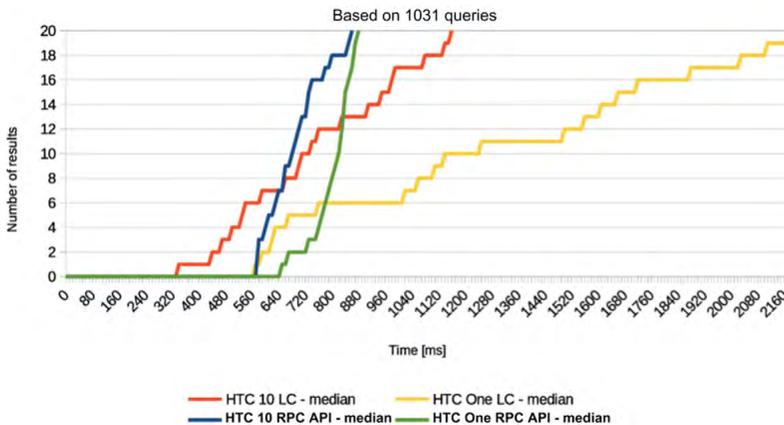
Figure 5 shows the distribution of response times for *vehicle* queries for both approaches and in the devices used for the tests. Every vehicle trip is considered to be one atomic result; therefore, incremental results are not supported for this data type. When looking at the distribution of the response times it becomes clear that vehicle data take a longer time to load using the LC implementation. The information about a single vehicle typically spans around 3 or 4 hours, which translates into a larger amount of data fragments that need to be retrieved and processed. The RPC API based approach, which has quicker access to the data, has an advantage here. It also performs consistently between devices, whereas the LC implementation needs two times as much time on the HTC One, compared to LC implementation on the HTC 10. Not only the data type and device affect the performance, but the exact query is of importance too. Calm stations, long routes, or vehicles with a long trip take longer to load compared to busy stations, short routes or vehicles with a short trip. The time to load a number of results is directly related to the timespan in which the results can be found. When a larger timespan needs to be evaluated, the results will take longer to load.

Figure 5. Technical performance distribution of vehicle queries on LC vs RPC API
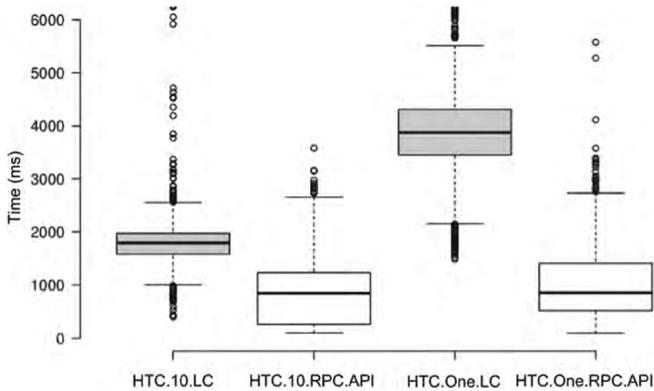
Open data end-users

> *(H2): The LC based route planning application has a higher user perceived performance when compared to traditional RPC API based implementations.*

Figure 6 depicts a summary of the user perceived performance tests for every defined use case and the overall choice made by the users between both approaches. Results show that the majority of users perceived the RPC approach as faster in every tested use case (*liveboards* – 47%, *routes* – 76% and *vehicles* – 65%). However, the overall choice shows that 59% of the users picked the LC

approach as the preferred choice. For most of them, this final choice was made mainly due to capacity of queries executed offline. Another reason for users to have given preference to the LC approach is privacy, where 85% expressed that they would be bothered if their location would be sent over the internet and 77% would be bothered if their journey itineraries would be known by third parties, which is the case for most route planning applications nowadays.

Figure 6. User perceived performance results for *liveboards*, *routes*, *vehicle* queries and overall choice between LC and RPC API
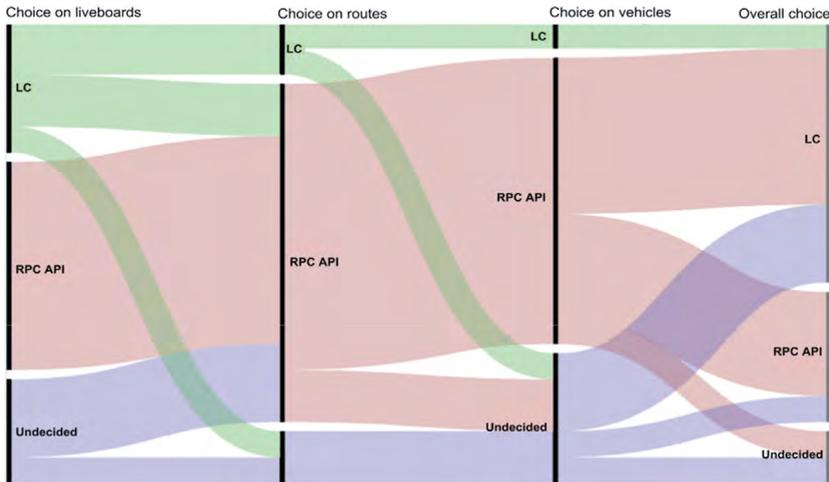
Discussion and conclusions

In order to determine the impact of implementing a decentralised publishing strategy of public transport open data for the route planning ecosystem, such as the Linked Connections framework, we conducted a series of evaluations that focused on the *open data reuser* and the *open data end-user* actors. However, even though we did not assess the impact of the LC approach on *open data publishers*, we can refer to previous work where the cost-efficiency of implementing the LC approach was measured (see Colpaert et al. 2017). Results showed that for data publishers, following the LC approach meant lower infrastructure associated costs as they can support a larger number of requests with less powerful servers, thus having better scalability. This has a positive and important effect for open data publishers, as one of the main goals of open data is to maximise data reuse and with this approach they can now support a larger number of clients with a lower investment.

Moreover, unrestricted access to data, which is one of the main challenges of open data in the route planning ecosystem, is also tackled by the LC approach.

Most traditional approaches use RPC API based architectures to expose route planning data and often require the imposition of access restrictions (e.g. in terms of number or requests per day) to their users to prevent overloading their servers. But with the higher efficiency achieved by the LC approach, open data publishers can give unrestricted access to the data mainly because the data becomes cacheable and the processing load of calculating routes is now moved on to the client side. But unrestricted access to data can be also interpreted from a query flexibility perspective. With an RPC API based approach, *open data reusers* are limited by the type of queries that the API has been built to support and cannot influence the type of data they obtain from each query. For route planning this means that *open data reusers* can request, for example, data about route alternatives to go from A to B from the route planning RPC API of the buses and trams operator, but cannot ask to include bike or car sharing options into the route calculation process. For an *open data reuser* to support new kinds of queries, this traditionally means creating a new route planning API from scratch and manually integrating the different datasets they want to include in their queries. The LC approach leverages this issue by simply publishing the raw data fragmented following a strategy optimised for route planning purposes. In this way, *open data reusers* can directly access the specific parts of a dataset that they need and combine them with any other external data sources, allowing them to support new types of queries. For example, an *open data reuser* could directly reuse the LC dataset from the bus and tram operator of a city and combine them with available bike or car sharing open datasets to support new types of queries and render new route alternatives, without being restricted by precalculated routes offered by their RPC API or the overhead of having to integrate the complete buses and trams dataset first.

We also did not focus on the *public and NGO stakeholder* actors. As mentioned before, these actors contribute to the ecosystem by providing the legal framework and the definition of mechanisms and standards through which the route planning ecosystem is supported. Therefore it can be argued that since they do not take a direct active role in the open data flow that takes place inside the route planning ecosystem, there would be no significant impact to this institution when implementing a decentralised open data publishing strategy.

When looking at the results obtained in the evaluation for *open data reusers*, we can observe that a route planning application implementing the LC specification and processing queries on the client side, performs acceptably well compared to its RPC API based counterpart, even obtaining a better performance for *routes* queries. We take into account that we only measured the app performance based on response time. Other performance benchmark methodologies could be used to get a more detailed insight into the performance, for example, bandwidth usage or battery consumption on end-user devices. Ensuring high performance of the applications is a main concern for reusers who seek to provide a high quality of service to their users and the results obtained during

these evaluations show that the LC approach provides a feasible alternative for the route planning ecosystem. However, there are still some types of queries where the LC approach does not perform as well as its counterpart, like for *vehicle* queries. Technically, this is due to *vehicle* queries needing access to data from larger timespans than other types of queries, which requires LC clients to request and process a higher amount of data fragments. This lays out a gap in the design of the LC specification that needs to be addressed by developers when implementing the specification, and since the LC framework is available as open source, reusers can keep optimising their implementations according to their needs. But without a doubt the greatest benefit that *open data reusers* get from following the LC framework is the full flexibility of data with low adoption costs. With LC, reusers can access the raw data from one or more public transport networks, which is not possible on traditional RPC API based approaches where the data flexibility is constrained by API implementations. Moreover, reusers can implement their own algorithms, integrating any kind of external data they could need to offer a specific service. Also, by publishing data as fragments and following the Linked Data principles, reusers can access the specific portions of data they need to solve any given query while using a unified model supported by semantic vocabularies. This lowers the data adoption costs for community reusers who do not need to incur data hosting and integration costs, allowing them to focus on the development of their core service (i.e. route planning algorithm). Considering this, it is possible to argue that a decentralised open data publishing strategy, such as the LC framework, may contribute to innovation and thus, the economic growth of the route planning ecosystem.

On the other hand, the results of the user perceived evaluation carried out for *open data end-users* shed light on the fact that the majority of the users in this study will value the additional features (such as offline querying and the safeguarding of their privacy) more than the performance. It is important to note that the empirical results reported in the research are subject to several limitations. First, there is the low number of participants. Only 17 respondents participated in the user perceived performance test and 81 in the questionnaire. Therefore, it is difficult to identify significant relationships in the data.

The second limitation concerns the internal validity of the user perceived performance test with regards to the offline and privacy features. In this case, the work of Nissenbaum (2011) is worth mentioning, as users are willing to give up privacy depending on the benefit they receive from a service. This means LC takes privacy more into account as the data, such as location or travelling preferences, are processed at the client side without the need of sending them to remote or third-party servers. Next to this we also note that the hardware capabilities of the end-user device have a major impact on the performance of client-side query evaluation. A less powerful device can reduce the user perceived performance of an LC client, as evidenced by most of the users selecting the RPC based approach as their preferred choice for performance.

This is an important aspect to be considered as not all users may have access to powerful devices, therefore it is an open issue for the LC framework to improve the performance of route planning use cases on devices of lower capabilities.

However, being capable of executing queries offline is a feature that the majority of the users regarded as more important than a better performance due to the fact that when travelling throughout public transport networks, mobile data connections are often lost. This can be noticed, for example, in rural zones with poor coverage or in subway tunnels which render RPC based route planning applications useless. In that case, an LC based client may already have pre-fetched the data or can use the data from a previous look up to keep answering queries. Pre-fetching is hardly possible in an RPC-style API, as this would require a request per every possible query. Also, being allowed to be in control of their own personal data (e.g. location and itineraries) is an important factor for end-users. With the recent breaking out of scandals about how personal data being collected through social media was being used to influence election results all over the world, users have become more aware of the importance of their data privacy.

Therefore, LC based applications provide them with a good alternative that takes this matter into account and protects the very sensitive data that is required in the route planning ecosystem. From a general perspective we could state that the impact for open data end-users of a decentralised approach would be reflected through a bigger, more varied and more personalised offer of services for the route planning ecosystem.

This work provides insights and an initial assessment of the potential effects of implementing a decentralised open data publishing strategy in the public transport route planning ecosystem. We have been able to observe that even though it still requires further work to improve some identified shortcomings, the potential benefits of such an approach are aligned with the ideals of open data of fostering innovation, boosting economic growth and providing solutions for more specific necessities (e.g. public transport accessibility for people with disabilities). Determining what key aspects end-users value the most when choosing an application and also which factors limit the performance of decentralised approaches are fundamental steps towards building a richer and sustainable route planning ecosystem that increases innovation and adoption of open data in the public transport sector.

However, it still remains as a research interest to determine how the decentralisation of open data publishing can be applied to other sectors. Also, how the actors in the ecosystem behave towards each other and how LC affects their current organisational and business models. At first sight, Linked Data and semantic technologies could provide the means to increase interoperability of datasets but further effort in creating comprehensive and common domain ontologies is still needed. Furthermore, exploring different strategies for fragmenting datasets that suit the needs of other (policy) domains and keep open data adoption costs low, is also an interesting research direction.

REFERENCES

Ahlers D, Akerkar R, Krogstie J, Opdahl AL, Tessem B & Zhang W (2018) Harnessing mobility data in cities: A case study from the Bergen region. *NOKOBIT* 1(26). http://hdl.handle.net/11250/2579413

Colpaert P, Verborgh R & Mannens E (2017) Public transit route planning through lightweight linked data interfaces. Paper presented at the 17th International Conference on Web Engineering, Lecture Notes in Computer Science: 403–411. https://doi.org/10.1007/978-3-319-60131-1_26

Deloitte (2017) Assessing the value of TfL's open data and digital partnerships. http://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf

Dibbelt J, Pajor T, Strasser B & Wagner D (2018) Connection scan algorithm. *The Journal of Experimental Algorithmics* 23: 1–56. http://doi.acm.org/10.1145/3274661

Dijkstra EW (1959) A note on two problems in connexion with graphs. *Numerische Mathematik* 1: 269–271. http://dx.doi.org/10.1007/BF01386390

Fielding R (2000) Architectural styles and the design of network-based software architectures. Doctoral dissertation, Information and Computer Science, University of California

Harrison T, Pardo T & Cook M (2012) Creating open government ecosystems: A research and development agenda. *Future Internet* 4(4): 900–928. https://www.mdpi.com/1999-5903/4/4/900

Heimstädt M, Saunderson F & Heath T (2014) Conceptualizing open data ecosystems: A timeline analysis of open data development in the UK. Paper presented at the International Conference for E-Democracy and Open Government. https://project.opendatamonitor.eu/wp-content/uploads/dissemination/OpenDataMonitor_Publication_Conceptualizing-Open-Data-Ecosystems.pdf

Hogge B (2016) *Open Data's Impact: Transport for London Get Set, Go!* Govlab & Omidyar Network. http://odimpact.org/files/case-studies-transport-for-london.pdf

Nissenbaum H (2011) A contextual approach to privacy online. *Daedalus* 140(4): 32–48. http://amacad.org/sites/default/files/academy/multimedia/pdfs/publications/daedalus/11_fall_nissenbaum.pdf

Rode P, Hoffmann C, Kandt J, Graff A & Smith D (2015) *Towards new urban mobility: The case of London and Berlin*. London: London School of Economics and Political Science. http://eprints.lse.ac.uk/id/eprint/68875

Share-PSI 2.0 (2016) Best Practice: Open up public transport data. https://www.w3.org/2013/share-psi/bp/ptd/

Van Schalkwyk F, Verhulst SG, Magalhães G, Pane J & Walker J (eds) (2017) *The Social Dynamics of Open Data*. Cape Town: African Minds. http://www.africanminds.co.za/wp-content/uploads/2017/06/9781928331568_txt.pdf

Verborgh R, Hartig O, De Meester B, Haesendonck G, De Vocht L, Vander Sande M, et al. (2014) Querying datasets on the web with high availability. Paper presented at the 13th International Semantic Web Conference, Trentino, 19–23 October

UK Department for Transport, Deloitte & Open Data Institute (2018) Bus Open Data – Collaboration to put the passenger first. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/722576/bus-open-data-case-for-change.pdf