

# Open Data Standards: Vertical Industry Standards to Unlock Digital Ecosystems

Daniel Rudmark  
RISE Research Institutes of Sweden  
[daniel.rudmark@ri.se](mailto:daniel.rudmark@ri.se)

## Abstract

*Standards are considered an essential means to facilitate value creation from open data. Despite this importance, we find that empirical studies of open data standards have not been conducted in proportion to its importance. In particular, the literature has insofar been silent about why specific standards are chosen and how these standards are implemented. To this end, we report from an action research project with the Swedish public transport industry, where open data standards were both chosen and implemented. Consistent with the literature, we find standards were selected based on expected increased attractiveness for re-users. Also, and more surprisingly, we found that open data standards were chosen as a means to harness resources in adjacent digital ecosystems. Finally, our findings convey that implementing open data standards may hamper the possibility to publish datasets, with its original qualities.*

## 1. Introduction

For some 15 years, governmental agencies around the globe have published its internal datasets publicly with little or no re-use restrictions as open data. Such datasets cover a wide variety of sectors including expenditures and tenders, air quality sensors, weather forecasts, and public transport networks.

The societal value that these open datasets may yield is contingent on re-use [1, 2], and much research has hence been devoted to understanding how this value may be realized. One fundamental technical requirement for open data to enable value creation is interoperability [3, 4], a behavior typically achieved through open data standards. For this reason, the current broad consensus on the importance of open data adhering to standards is perhaps not surprising.

Despite this fundamental importance, the topic of open standards has received surprisingly little attention in the literature. First, there is currently a lack of precision about what is meant by open data standards, as standards may apply to both format, metadata, and semantic level. Moreover, there is currently a dearth of in-depth research on how these indisputably relevant data standards are chosen and implemented.

Previous research has stressed that the prime rationale for open data publishers to publish data according to specific data standards is typically connected to usability [5-8] and interoperability [9-11] *vis-à-vis* the re-use community [12]. E.g., if a particular category of open data were to be published in the same standard across agencies, the threshold for re-users would be substantially lowered.

While we find this position plausible, we in this paper examine whether open data standards can be chosen for other reasons. For instance, adoption of non-proprietary data standards, may not only be beneficial to external re-users [13]. Besides, such standards can also be used as a means to create compatibility with valuable IT resources in the organizational ecology [14]. As digital innovation becomes more distributed and relying on loosely coupled actors, more such service innovation opportunities are offered by digital ecosystems, where data standards play a pivotal role [15].

In this research, we thus seek to complement the current literature on open data standards by getting more in-depth insights on how such standards are chosen. Also, data standards are rarely neutral descriptions of reality but embedded in socio-technical contexts, ripe with political tensions [16]. For this reason, there is also a need to understand what challenges may arise when agencies implement open data standards. To develop new knowledge on these topics, we have thus explored the following research question:

*Why and how do open data providers choose and implement open data standards?*

The remainder of this paper is organized as follows: we first review the related literature on open data, related standards and how standards can be used to redeem resources in nascent ecosystems. Next, based on data from an action research venture, we detail how Swedish public transport organizations chose and implemented several open data standards. Building on these findings, we conclude this paper by offering a discussion on why and how open data providers choose and implement open data standards.

## 2. Open Data

In this paper, we refer to open data as internal data passed beyond an organizational border [17], where intellectual property rights have been outright relinquished or reduced to a minimum [18]. Often such open data is made available through open Application Programming Interfaces in machine-readable formats [19] to facilitate integration. Open data in this form is typically intended to be used by external developers, enabling these innovators to create value by accessing data through the platform's boundary resources [20].

As re-use is at the core of any open data initiative, an ecosystem view of open data has been advocated [21]. At the core of such ecosystems are open data providers [21], sharing its digital resources with the public. Open data providers can, in turn, be split into decisions makers (that provide legitimacy and resources to the ecosystem) and administrative agencies (that ensure data publication). Value creation in such ecosystems is typically contingent on both public and private organizations where each organization contributes to the value creation. [22]. However, as argued by Sieber and Johnson [23], governments as data providers must ensure that value is indeed created for citizens and government, and not exclusively for corporations. One key to such value creation lies in standardization of open data, a topic which is expanded upon below.

### 2.1. Open Data Standards

In the open data literature, there is a broad consensus that standards are a necessary part of an open data program. The literature stresses that standards are necessary to enable interoperability between datasets [9-11] and increase data usability [5-8], and eventually provide the infrastructure for a vibrant open data ecosystem. However, within open data, standards operate on several levels, and to

answer the research question in this paper, it is necessary to clarify these levels.

On the most general level, standards in the open data literature may refer to general-purpose and domain-agnostic standards, seeking to ease processability and decouple data from proprietary formats [24-26]. Through these types of standards, data becomes digestible regardless of data processing software, and examples of such standards include XML, JSON, and CSV. We refer to these high-level standards as *format standards*.

Second, a large portion of the literature deals with metadata standards [9, 10, 27]. Metadata is defined as "data that describes and gives information about other data" [28]. Such information may include license information, encoding schemes, content declaration, and quality attributes. There are several metadata standards used for open data. In the EU, for example, DCAT-AP is used as a standard for open data portals, and the INSPIRE standard is used to describe geodata. Consequently, we refer to these types of standard as *metadata standards*,

The last type of open data standard found in the literature is typically domain-specific and mandates how the detailed semantics of a particular domain should be expressed in the open datasets. For instance, many authorities on a city level want its citizens to report non-emergency issues (such as potholes and fallen trees). While such service traditionally in US cities have been reachable by dialing 311, the Open311 standard allows for machine-readable postings of such issues [29]. As a result, the Open311 API standard is supported by several cities, as well as third-party software. Other examples of vertical industry standards for open data includes GTFS for public transport [30] and the IATI standard for international aid [31]. We refer to this type of standard as *vertical industry standards* [32].

Since vertical industry standards prescribe both what data that shall be published and how this data should be structured, this demarcates vertical industry standards from both format standards and metadata standards. In what follows, we thus expand on the implications of using vertical industry standards to publish open data.

### 2.2. Vertical Industry Standards and Open Data

Open data as a resource is typically portrayed as a sort of digital spillover, in that the data already exists within agencies yet the data's "stickiness" prohibit further value-adding activities [33]. When such data instead is released and published openly, it can be harnessed by external actors for transparency,

efficiency, or innovation purposes. In practice, however, any such data have been stored for quite particular application purposes and is, in addition, subjected to various forms of processing and transformations during its lifecycle [34].

For instance, an essential part of the empirical setting in this research concerns e.g., real-time data from public transport operations. At its most basic level, such data is generated by signal processing units in public transport vehicles and corresponding infrastructure, generating verbose data streams about the current state of a particular vehicle. Such data is subsequently refined and used by back-office fleet management applications before it is transformed to more upstream-oriented departure time estimations targeting travelers. As this example shows, real-time data is not given but rather inherently context-dependent. A central tenet of open data is that data released openly should be complete. The rationale for such a principle is that no data should be lost during the publication process, e.g., through aggregation. Consequently, this inherent plasticity of data [35] poses a delicate dilemma for data publishers: how can data be published as a whole yet be digestible for the re-users?

Vertical industry standards with sufficiently high industry penetration offer a solution to this dilemma. Since such standards “address business problems unique to particular industries” [32, p. 81] the standard stipulates what data that should be published to represent essential entities in the particular domain. For data publishers, this means they need to convert their internal datasets to conform with the particular industry standard, and thereby achieve compatibility and data usability. Such compatibility is however not unique to open data resources. In addition, vertical industry standards can be used to harness digital ecosystems, a topic which is expanded upon below.

### **3. Vertical Industry Standards to Harness Digital Ecosystems**

A digital ecosystem is a particular type of organizational form, and can be described as “a distributed adaptive open socio-technical system with properties of self-organization, scalability, and sustainability” [36, p. 18]. A digital ecosystem can thus be seen as mimicking the characteristics of a natural ecosystem [37]. A thriving natural ecosystem is contingent on symbiosis since different organisms can sustain habitat survival only as a result of their relative diversity towards – and interactions with – other organisms in the ecosystem. A growing body of literature has hence used ecosystems as a metaphor to

convey the dynamics of simultaneously cooperating and competing actors seeking to propel a particular shared interest mutually (e.g., a new technology) [14, 38-40].

Most studies on digital ecosystems have primarily inquired into focal firms [39, 41-44]. However, as argued by Selander, Henfridsson and Svahn [14], a mere few can expect to act from such a central and commanding position [45] – albeit still benefit from ecosystem participation. The large body of non-focal actors instead needs to make strategic decisions about what ecosystems to participate in or redraw from, rather than maintain a position as the keystone actor. By actively searching for and redeeming innovation capabilities in nascent ecosystems [14], successful non-focal organizations may, in a cost-efficient manner, draw on these resources offered by digital ecosystems. Vertical industry standards can play a pivotal role for such ecosystem participation.

Given the arms-length relationships and necessity for digital ecosystems to scale quickly, there is increasing awareness that open vertical industry standards can help leverage the growth of digital ecosystems [13]. When such protocols and semantics have been established and adopted by data publishers, it creates the necessary preconditions for a vibrant re-use ecosystem.

Given this theoretical background, we next dive into our empirical research, investigating how the Swedish public transport industry as open data providers have selected and implemented open data standards.

## **4. Method**

The findings presented in this paper stem from an on-going canonical action research (CAR) venture [46] between Samtrafiken AB and the authors of this paper. Samtrafiken collects, develops, and maintains traffic data, industry standards, and combined ticketing and mobility solutions for Sweden's national public transport network. Samtrafiken is co-owned by all public transport authorities as well as most commercial, public transport companies in Sweden. A core mission for Samtrafiken is to provide open data for the entire public transport industry in Sweden, and the overarching purpose of the collaborative research project was to increase re-use of open data published by the Swedish public transport industry.

The CAR project explored the idea of public transport, assuming a more peripheral position towards digital ecosystems [14] as a novel way of increasing diffusion of open public transport data. To achieve this, it was necessary for the industry to develop required participation-enabling capabilities –

work that had been assessed to be performed as an action research project. During this work, it became apparent that open data standards played a vital role to unlock adjacent digital ecosystem – yet the role of open data standards had insofar been underexplored, which led to findings presented in this paper [47, 48].

#### 4.1. Data collection

The empirical data presented in this paper was collected from the diagnosis, action planning, and action taking phase [46]. During the diagnosis of the public transport industry in Sweden, we collected both the final report and the detailed notes from six workshops leading up to the ratification of five strategic objectives for open public transport data (see ch 5 below). This way, a more thorough and coherent understanding of the public transport industry's strategic challenges and plans was made possible. During the diagnosis phase, we also interviewed several representatives from the public transport industry involved with open data, having positions both on a technical and strategical level (N=11). These interviews lasted between 60 and 90 minutes, were conducted face-to-face or via video conferencing software. All these interviews were recorded and transcribed. To sum up the diagnosis phase, the authors of this paper conducted a workshop with 8 representatives from Samtrafiken and 1 from the Swedish Transport Administration (all engaged in open data), where the preliminary findings were discussed and elaborated. The workshop lasted for 3 hours and was recorded and transcribed.

The data from the action planning phase consisted of 4 coordination meetings, where possible actions were discussed and assessed. These meetings were captured through field notes. In the action taking phase, we conducted an additional 3 meetings with members from Samtrafiken. In addition, we interviewed key personnel (N=7), involved in the open data standards work, in order to get a deeper understanding of the issues at hand. These interviews were conducted face-to-face, recorded and transcribed.

#### 4.2. Data analysis

Our approach to qualitative data analysis followed the guidelines by Miles and Huberman [49] and has in this research been based on iterative and concurrent data reduction, data display, and conclusion. Data reduction in this paper meant revisiting the full dataset from diagnosing, action planning, and action taking through the lens of open data standards. Since this

original dataset contained interesting yet too limited empirical findings on Samtrafiken's work on implementing the NeTEx standard, we thus collected the additional interviews described above. For data display, we used the network view in our analysis tool, Atlas.ti, to visualize our emerging codes, and to draw informed conclusions from our empirical material. This approach allowed us to craft an in-depth account of how open data standards have been selected and implemented within Samtrafiken.

### 5. Results

In 2016, the Government Offices of Sweden, through "Forum for Transport Innovation," ignited a redesign of open public transport data in Sweden. The primary reason for this initiative was to create a more comprehensive and harmonized open data delivery from the public transport industry. For instance, real-time data were only available in a few regions, and the datasets from different regions were difficult to combine. From a policy perspective, more comprehensive data from the public transport industry was a necessity to enable new mobility services, as emphasized by the project's funder, a program manager at Sweden's innovation agency:

"If developers of mobility service get access to high-quality real-time data, we are convinced that these developers can convert such information into proactivity in the service towards the customer. For instance, your phone can notify you that 'please leave the train at the next station, because there is trouble ahead, and you can instead use a carpool to reach your final destination.' To succeed with this, you need access to data!"

Samtrafiken led this project, and as a result of this 9-month work (consisting of interviews, workshops with public transport experts and mobility services developers, and management decision meetings), five strategic objectives were eventually formulated and accepted by the public industry as a whole. One of these objectives prescribed a new systems architecture for handling open public transport data in Sweden. In this architecture, Samtrafiken would collect data from all public transport agencies in Sweden and provide it as open data.

A core principle of this architecture concerned open data standards. During the study, it was evident that the standards used by the public transport actors and external re-users were not the same. For this reason, the new architecture stipulated that public transport agencies would send data to Samtrafiken using standards that their back-office systems relied on, NOPTIS [50]. Samtrafiken would then convert and

publish the open data in standards with highest re-use demands.

The first pinpointed open data standard with high reuse demand was GTFS [30]. Effectively, this standard prescribes how network and timetable data is exported as a text-based relational database. While this standard initially was designed by Google for public transport actors to export their public transport data into Google maps, many public transport organizations have also published these feeds openly. As a result, GTFS has become the de-facto global standard for open public transport data. Samtrafiken had been publishing GTFS feeds since 2013, and together with its travel planning API, GTFS was the most popular open data feed. Most notably, for major international services such as Moovit, CityMapper, and Trafi, the GTFS feed from Samtrafiken had served as a key to unlock the Swedish market.

The second open data standard that Samtrafiken would export data on was NOPTIS. This standard was both developed and used by many public transport actors in Sweden (and, as mentioned above, also used as an input data standard for Samtrafiken). The reason for supporting NOPTIS as an open data standard was requirements from the public transport industry itself to use the open data for intra-industry purposes. Currently, many agencies shared data on a case-by-case basis, and by scrapping existing integrations and start exchanging data through the open data portal, much cost savings was foreseen. However, data sharing through NOPTIS relied on database replication, and thus, limited use beyond the industry actors was foreseen.

The third and final category concerned the related standards NeTeX [51].

## 5.1. The NeTeX profile

The reason NeTeX was chosen as an open data standard was an upcoming EU-regulation [52], stipulating that several data categories from the public transport industry should be released openly and be compliant with the public transport standard NeTeX. Although this regulation inferred much work for the public transport sector, it was endorsed by Samtrafiken, as described by its chief system architect:

“There's a lot of things happening in Europe, since EU are trying to standardize traffic data, and the CEN standards NeTeX and SIRI<sup>1</sup> have been chosen as European standards. So, it's very logical for us

to support these standards. In our work towards simpler data sharing between actors, we really should agree on the language we should be talking.”

However, NeTeX is a broad standard, and the documentation comprises more than 3000 pages. As such, NeTeX allows the standard's user to represent the same real-world entities through different NeTeX constructs. For this reason, the user to must decide on the core principles for how central objects, such as timetables and bus stops, shall be represented. This more focused interpretation of the NeTeX standard is called a *profile* and is a text document prescribing what parts of the NeTeX standards that are used to define central public transport constructs. Given this situation, the actual implementation of NeTeX differs across countries, as commented by the chief architect:

“Since the scope of NeTeX is so large, there will be a need to have a profile. Currently, the German profile differs from the European, and the Norwegian profile differs from the German profile and so on. In the long term, however, I believe the de facto usage of NeTeX will converge across countries.”

Designing a coherent profile from the whole NeTeX standard is a challenging task, and Samtrafiken thus considered several ways of doing this. The first way relied on using the NOPTIS standard. Since both NOPTIS and NeTeX were based on the same object model, Transmodel [54], this would have been a moderately simple task, as commented by the chief architect at Samtrafiken:

“We did consider designing a Swedish NeTeX profile based on NOPTIS since that by far would have been the easiest way forward. Since our current systems build on NOPTIS, we would have gotten a straightforward systems solution and a very clear data model. It had great benefits doing the profile this way.”

However, despite these apparent benefits from basing the profile on NOPTIS, Samtrafiken instead chose to implement the Norwegian NeTeX profile.

In Norway, Entur AS plays a similar role as Samtrafiken does in Sweden. They collect data from all public transport actors across Norway and publish both open data and provides travel planning services. In the last few years, Entur had made a significant redesign of its systems infrastructure and moved from

---

<sup>1</sup> NeTeX covers static data (such as bus stops, time tables, and line geometries), but the regulation also encourages member states to publish corresponding real-time data (such as vehicle positions and arrival time estimations) in the SIRI [53] format.

procured solutions from public transport software houses towards in-house development and open source software. In this transformation, NeTeX and the associated Norwegian NeTeX profile had played a pivotal role as a vertical industry standard, both between public transport actors and Entur, and as an open data standard targeting developers and other re-users. In what follows, we expand on why the Norwegian NeTeX profile was chosen.

## 5.2. Benefits of NeTeX profile

One perhaps less surprising reason for choosing the Norwegian NeTeX profile concerned that Entur had been an international spearhead in using NeTeX and the Norwegian profile had thus been tested in production for a significant amount of time. Given the importance of this data, having a verified profile was considered a significant advantage, as commented by the object owner of open data at Samtrafiken:

“Entur is both importing and exporting data from most public transport actors in Norway through the Norwegian profile, and quite a few public transport actors actually re-use this NeTeX data in their daily operations. So, the profile works. Also, if one wants to build a service with national coverage, this just works too.”

Not only did this mean that integration was possible, but also that system providers already had support for the Norwegian NeTeX profile, as commented by the chief architect at Samtrafiken:

“There are several system suppliers that created export modules for the Norwegian profile, and this makes it so much easier for our partners when they are integrating their data towards us.”

While the most fundamental requirement on an open public transport data standard is the capability to exchange data between actors, another, almost as important, is the ability to convert data into actionable travel options for travelers. A trip involving public transport may potentially include several public transport lines, as well as connecting modes of transport (such as walking, park-and-ride, hail-riding or rental bikes). Also, high-quality map data is necessary to be able to produce walking links to and between stops. Finally, travel searches must be fast and thus be able to extract the best trip from a vast array of travel options – all in fractions of a second. Hence, the algorithms necessary to produce such travel options is therefore complex and requires substantial resources to develop.

Traditionally, these algorithms have been part of commercial software packages, developed and

marketed by software houses. In the last few years, however, more public transport agencies had started to use the open source package OpenTripPlanner [55]. This package was initially developed by non-profit organization OpenPlans and is currently used and maintained by several private software companies and public transport agencies such as TriMet, Oregon, USA, Plannerstack, the Netherlands and HSL, Helsinki, Finland. In the last few years, Entur had made OpenTripPlanner a central part of their systems architecture and entirely relied on the framework for travel planning purposes. In this work, they were not merely users of the algorithm but had also become active contributors to the OpenTripPlanner codebase. One rationale for this closer collaboration from Entur with the OpenTripPlanner community was to include support for NeTeX. Before this engagement by Entur, OpenTripPlanner only supported the GTFS standard. However, GTFS can not afford the kind of details that NeTeX does. The chief architect at Samtrafiken elaborated on these capabilities:

“NeTeX has so much higher data resolution and allows for structures on several levels. You can have a site, and in this site, you can have three stations, and in these stations, you can have substations with stops and platforms. You can just describe the infrastructure so much better with NeTeX. Say that you have a train platform, you can describe how long this platform is, where it is located, I believe you can even define an area for the platform. And then you can define whether the trains stop at the beginning or the end of this platform, you can provide so much more details! And while GTFS is very straightforward and easy to manage, it does not provide any of these details. Everything is just a stop with coordinates.”

As a result, using NeTeX for travel planning would provide the users with more detailed travel planning options than the ones GTFS could afford. The flexibility of NeTeX and the fact that Entur had implemented support for the Norwegian NeTeX profile was thus a prime reason that Samtrafiken chose to implement the Norwegian NeTeX profile, as commented by a data architect at Samtrafiken:

“It certainly was. One of the core reasons we chose to scrutinize and eventually use the Norwegian NeTeX profile was the native support for OpenTripPlanner. We do think that OpenTripPlanner will be our main track when it’s time to implement a new travel planner.”

Being able to get compatibility with OpenTripPlanner “out of the box” was hence a core rationale for Samtrafiken to use the Norwegian NeTeX

profile, as commented by the chief architect at Samtrafiken:

"It would require substantial development effort and a couple of developers for perhaps six months, and this we don't have to do by using the Norwegian NeTEx profile. Also, we can draw on knowledge that Entur has developed by working with OpenTripPlanner for four, five years."

More specifically, Samtrafiken wanted to investigate whether OpenTripPlanner could serve, not just for presenting travelers with different travel options, but also as a basis for selling tickets. A project manager at Samtrafiken expanded on this strategy:

"We currently have three strategic projects: we have open data and another one building on our ticket sales standard, BoB. And then we have this third strategic project we call combined ticket sales of the future. All these three are connected since we both need high-quality data and a secure standard for ticket sales to achieve combined ticketing. And over these three projects hovers the magic travel planner."

"This has been one of the driving forces behind starting early with NeTEx, to deliver high-quality data to a travel planner, which in turn will be used to sell tickets. So, this has been a chain of dependencies where NeTEx has been the first link."

However, OpenTripPlanner was not the only reason that the Norwegian NeTEx profile was chosen. The compatibility with the open source software package Chouette was another one:

"Entur are using a tool called Chouette, a web interface for traffic data. This is a solution we'd like use for our partners that don't have systems of their own, that they just can log in to our systems and create the timetable and other information that is needed. Today they are emailing Excel files and what not, and this consumes too many resources on our side. And Chouette can handle the Norwegian NeTEx profile out of the box."

While Samtrafiken identified several benefits from using the Norwegian NeTEx profile, this choice also inferred several challenges, in need of being addressed. These challenges are described below.

### **5.3. Challenges when Implementing the NeTEx Profile**

On a technical level, the core challenge for Samtrafiken was to make the conversion between its

NOPTIS-based data model and the data model used by the Norwegian NeTEx profile, as commented by a systems developer:

"Some entities are described in more detail in NOPTIS than in the Norwegian NeTEx profile, and vice versa, so you always have to map different concepts between these two standards when you are importing or exporting."

In this mapping process, developers at Samtrafiken needed to handle different discrepancies between how Swedish public transport actors' models and export its data.

"Take accessibility data; for instance, this is required by the Norwegian NeTEx profile. This data is not sent to us, although some of the regional public transport authorities store it. Sometimes you could find ways just to omit this data in the NeTEx export, but when it is required, we just export 'unknown'".

Other issues brought up by the developers doing the actual conversion was train number (used by Swedish train companies rather than line numbers) and how to define passing times in timetables:

"A bus *en route* will pass many stops. If only a few passengers board the bus, it may actually go faster than what is estimated in the timetable, and what you're typically seeing in the timetable for your little local bus stop is an estimate, not a promise. But then you have these larger stops that have controlled times. And when you have a controlled time, the bus will stand still on the bus stop, waiting for the time entered in the timetable. In Sweden, most regional actors use this differentiation, but there is currently no easy way to implement it in the Norwegian NeTEx profile "

These discrepancies forced the developers to, in the short term, resort to various workarounds. For instance, the public transport actors that were to send in data in the NeTEx format to Samtrafiken needed to export stops twice, depending on whether the stop was subjected to estimated or controlled time.

Although these workarounds provided an immediate solution to the differences in the data models, Samtrafiken anticipated that some changes to the profile, such as these issues with timetable passages, would be necessary. In order to be able to exert some influence of the Norwegian NeTEx profile, a mutual change control board was formed together with Entur and representatives from Denmark and Finland. In this process, the profile was also renamed to "the Nordic NeTEx profile." The idea with this board was to create a forum that allowed to adjust the

profile according to new requirements from the actors that were actively using the profile (such as Samtrafiken).

The governance document that contained the guiding principles for the work by the change control board stated that proposed changes that broke compatibility with earlier versions of the profile needed support from every member of the change control board. Hence, such changes were unlikely to be passed, and the changes proposed by the board's member should thus strive for maintaining compatibility with the earlier versions of the profile. For this reason, a system developer at Samtrafiken sought to design a change request that resolved the issues with timetable passing times, yet maintained backward compatibility:

"I'm quite certain that this distinction could be made on departures, rather than networks as it is today. Then the profile would allow both the original and Swedish way to handle passage time and still maintain compatibility with previous versions of the NeTEx profile."

## 6. Discussion

In this paper, we have investigated the following research question: *Why and how do open data providers choose and implement open data standards?* We embarked on this endeavor based on the paradoxical finding that open data standards are seen as a fundamental component to a successful open data program, yet the selection and implementation of such standards remain opaque in the literature. In our clinical setting, we found that our client organization chose and implemented not one, but three different vertical industry standards for their open data.

First, GTFS was chosen based on its massive re-use community. Although the standard is limited in terms of what the standard can express, it is well established among developers. This rationale is consistent with previous research that asserts that standards are used to increase uptake by developers for data that adheres to standards.

Second, NOPTIS was chosen as a standard primarily to facilitate data sharing among public transport agencies. As such, this standard had little anticipated relevance to extra-industry actors, such as civic technologists. While this reason has not been explicitly noted in the literature, it could be inferred since intra-agency data sharing is a well-known use case, and standards are thought to facilitate such data sharing.

Third, the Norwegian/Nordic NeTEx profile was chosen for two reasons. First, there were policy

pressures from the EU to release data under NeTEx. Second, and more surprisingly, the choice to choose the Norwegian profile as the actual implementation of NeTEx did not relate to external re-use, nor ease of implementation. Instead, the rationale was rooted in the notion of harvesting resources in adjacent digital ecosystems, to be used in our client organization's internal systems architecture. More specifically, the most valuable resource was the journey planning algorithm OpenTripPlanner. Since the Norwegian profile was fully compatible with this framework, it was a prime rationale to choose this profile, rather than creating an own profile, based on the data models currently used by Samtrafiken.

Moreover, we found that choosing to release open data through vertical industry standards was not without challenges. In particular, since vertical industry standards prescribe how core business entities should be represented, this may force the open data publisher to process the data in a way that it loses some of its original qualities. This side-effect of using vertical industry standards can also be considered to interfere with open data policies stating that open data shall be published as-is. We found that Samtrafiken used multiple actions to overcome these challenges. First, they resorted to technical workarounds, more specifically by producing duplicate representations of the same bus stops. Second, they sought to influence the standard on a more longer-term level, by engaging in the standards governance and focusing on changes that were likely to be ratified by other board members.

In this research, we have shown that choosing and implementing open data standards contains more dimensions than what is currently reported in the literature. More specifically, we have shown that as open data matures, it is a resource that also can be used for internal benefits. These benefits can be realized by drawing on resources from digital ecosystems, and an essential key to unlocking these resources are open data standards. In fact, we speculate that such actions will become more common as policymakers increasingly mandate the publication of open data. For instance, an updated PSI directive was recently ratified by the European Union. The directive sets out the mandatory publication of several high-value datasets from authorities and publicly owned organizations, following industry standards. As such standards are likely to afford a certain degree of interpretative flexibility (like NeTEx), such policies may steer these organizations towards adopting standards enabling the redeeming of digital ecosystem resources.

We see several research opportunities, building on the findings in this paper. First, as our findings rest on data from a single organization, we see a need for more research on additional organizations, preferably

outside the public transport industry, to get more insight on rationales and implementations of open data standards. Second, we see a dire need to understand the implications of the selection of different vertical industry standards. As our dataset highlights, choosing a more constrained yet scalable standard like GTFS infers a quite a different re-use trajectory than a more comprehensive and flexible standard like NeTeX. Finally, we would encourage researchers to investigate whether, or to what extent, policy-driven open data standardization efforts lead to *de facto* harmonization for re-users.

## 7. References

- [1] F. Ahmadi Zeleti, A. Ojo, E. Curry, "Exploring the economic value of open government data", *Government Information Quarterly* 33(3), 2016, pp. 535-551.
- [2] J. Manyika, M. Chui, P. Groves, D. Farrell, S. Van Kuiken, E. Almasi Doshi, *Open data: Unlocking innovation and performance with liquid information*, McKinsey Global Institute, 2013.
- [3] B. Ahlgren, M. Hidell, E. Ngai, "Internet of Things for Smart Cities: Interoperability and Open Data", *IEEE Internet Computing* 20(6), 2016, pp. 52-56.
- [4] M. Janssen, E. Estevez, T. Janowski, "Interoperability in Big, Open, and Linked Data--Organizational Maturity, Capabilities, and Data Portfolios", *Computer* 47(10), 2014, pp. 44-49.
- [5] M. Haklay, A. Singleton, C. Parker, "Web Mapping 2.0: The Neogeography of the GeoWeb", *Geography Compass* 2(6), 2008, pp. 2011-2039.
- [6] A. Zuiderwijk, M. Janssen, C. Davis, "Innovation with open data: Essential elements of open data ecosystems", *Information Polity* 19(1,2), 2014, pp. 17-33.
- [7] P. Johnson, R. Sieber, T. Scassa, M. Stephens, P. Robinson, "The Cost(s) of Geospatial Open Data", *Transactions in GIS* 21(3), 2017, pp. 434-445.
- [8] H. Krambeck, L. Qu, "Toward an Open Transit Service Data Standard in Developing Asian Countries", *Transportation Research Record* 2538(1), 2015, pp. 30-36.
- [9] J. Bertot, U. Gorham, P. Jaeger, L. Sarin, H. Choi, "Big data, open government and e-government: Issues, policies and recommendations", *Information Polity* 19(1/2), 2014, pp. 5-16.
- [10] T. Harrison, T. Pardo, M. Cook, "Creating Open Government Ecosystems: A Research and Development Agenda", *Future Internet* 4(4), 2012, pp. 900-928.
- [11] P. Parycek, J. Höchtl, M. Ginner, "Open Government Data Implementation Evaluation", *Journal of Theoretical and Applied Electronic Commerce Research* 9(2), 2014, pp. 13-14.
- [12] T. Jetzek, "Managing complexity across multiple dimensions of liquid open data: The case of the Danish Basic Data Program", *Government Information Quarterly* 33(1), 2016, pp. 89-104.
- [13] M. Markus, C. Loebbecke, "Commoditized digital processes and business community platforms: new opportunities and challenges for digital business strategies", *MIS Quarterly* 37(2), 2013, pp. 649-654.
- [14] L. Selander, O. Henfridsson, F. Svahn, "Capability search and redeem across digital ecosystems", *Journal of Information Technology* 28(3), 2013, pp. 183-197.
- [15] Y. Yoo, R. Boland, K. Lyytinen, A. Majchrzak, "Organizing for Innovation in the Digitized World", *Organization Science* 23(2), 2012, pp. 1398-1408.
- [16] S. Goëta, T. Davies, "The daily shaping of state transparency: Standards, machine-readability and the configuration of open government data policies", *Science & Technology Studies* 29(4), 2016, pp. 10-30.
- [17] A. Marton, M. Avital, T. Jensen, *Reframing Open Big Data*, ECIS 2013, 2013, p. 146.
- [18] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, Z. Ives, *DBpedia: A Nucleus for a Web of Open Data*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, pp. 722-735.
- [19] A. Latif, A.U. Saeed, P. Hoefler, A. Stocker, C. Wagner, *The Linked Data Value Chain: A Lightweight Model for Business Engineers*, Proceedings of the 5th International Conference on Semantic Systems (I-SEMANTICS), 2009.
- [20] C. Bonina, B. Eaton, S. Henningson, *Governing Open Data Platforms to Cultivate Innovation Ecosystems: The Case of the Government of Buenos Aires*, Proceedings of ICIS2018, 2018.
- [21] S. Dawes, L. Vidasova, O. Parkhimovich, "Planning and designing open government data programs: An ecosystem approach", *Government Information Quarterly* 33(1), 2016, pp. 15-27.
- [22] M. Janssen, A. Zuiderwijk, "Infomediary Business Models for Connecting Open Data Providers and Users", *Social Science Computer Review* 32(5), 2014, pp. 694-711.
- [23] R. Sieber, P. Johnson, "Civic open data at a crossroads: Dominant models and current challenges", *Government Information Quarterly* 32(3), 2015, pp. 308-315.
- [24] E. Kalampokis, E. Tambouris, K. Tarabanis, "A classification scheme for open government data: towards linking decentralised data", *International Journal of Web Engineering and Technology* 6(3), 2011, pp. 266.
- [25] N. Veljković, S. Bogdanović-Dinić, L. Stoimenov, "Benchmarking open government: An open data perspective", *Government Information Quarterly* 31(2), 2014, pp. 278-290.

- [26] A. Zuiderwijk, M. Janssen, "Open data policies, their implementation and impact: A framework for comparison", *Government Information Quarterly* 31(1), 2014, pp. 17-29.
- [27] A. Zuiderwijk, M. Janssen, S. Choenni, R. Meijer, R.S. Alibaks, "Socio-technical Impediments of Open Data.", *Electronic Journal of e-Government* 10(2), 2012, pp. 156-172.
- [28] Oxford English Dictionary, ""metadata, n."" , 2019. <https://www.oed.com/view/Entry/117150?rskkey=DKUCv3&result=4#eid37413841>. (Accessed 2019-05-01).
- [29] D. Offenhuber, "Infrastructure legibility—a comparative analysis of open311-based citizen feedback systems", *Cambridge Journal of Regions, Economy and Society* 8(1), 2015, pp. 93-112.
- [30] Google Inc., "General Transit Feed Specification Reference", 2019. <https://developers.google.com/transit/gtfs/reference/>. (Accessed 2019-04-24).
- [31] International Aid Transparency Initiative, "IATI Standard", 2019. <https://iatistandard.org/en/iati-standard/>. (Accessed 2019-04-24).
- [32] M. Markus, C. Steinfield, R. Wigand, The evolution of vertical IS standards: Electronic interchange standards in the US home mortgage industry, *International Conference on IS Special Workshop on Standard Making* 2003.
- [33] E. von Hippel, "Economics of Product Development by Users: The Impact of "Sticky" Local Information", *Management Science* 44(5), 1998, pp. 629-644.
- [34] T. Davies, M. Frank, 'There's no such thing as raw data'. Exploring the socio-technical life of a government dataset, *Proceedings of the 5th Annual ACM Web Science Conference, WebSci'13*, 2013, pp. 75-78.
- [35] J. Kallinikos, A. Aaltonen, A. Marton, "The ambivalent ontology of digital artifacts", *MIS Quarterly* 37(2), 2013, pp. 357-370.
- [36] S. Jansen, M. Cusumano, Defining Software Ecosystems: A Survey of Software Platforms and Business Network Governance, in: S. Jansen, M. Cusumano, S. Brinkkemper (Eds.), *Software Ecosystems: Analyzing and Managing Business Networks in the Software Industry*, Edward Elgar Publishing, 2013, pp. 13-29.
- [37] J. Moore, "Predators and prey: a new ecology of competition", *Harvard Business Review* 71(3), 1993, pp. 75-83.
- [38] R.C. Basole, "Visualization of interfirm relations in a converging mobile ecosystem", *Journal Information Technology* 24(2), 2009, pp. 144-159.
- [39] J. Wareham, P. Fox, J. Cano Giner, "Technology Ecosystem Governance", *Organization Science* 25(4), 2014, pp. 1195-1215.
- [40] M. Iansiti, R. Levien, The keystone advantage: what the new dynamics of business ecosystems mean for strategy, innovation, and sustainability, *Harvard Business School Press*, Boston, Mass, 2004.
- [41] J. West, "How open is open enough? Melding proprietary and open source platform strategies", *Research Policy* 32(7), 2003, pp. 1259-1285.
- [42] A. Gawer, R. Henderson, "Platform Owner Entry and Innovation in Complementary Markets: Evidence from Intel", *Journal of Economics & Management Strategy* 16(1), 2007, pp. 1-34.
- [43] B. Eaton, S. Elaluf-Calderwood, C. Sørensen, Y. Yoo, "Distributed Tuning of Boundary Resources - The Case of Apple's iOS Service System", *MIS Quarterly* 39(1), 2015, pp. 217-244.
- [44] A. Ghazawneh, O. Henfridsson, "Balancing platform control and external contribution in third-party development: the boundary resources model", *Information Systems Journal* 23(2), 2013, pp. 173-192.
- [45] M.A. Schilling, "Technology Success and Failure in Winner-Take-All Markets: The Impact of Learning Orientation, Timing, and Network Externalities", *Academy of Management Journal* 45(2), 2002, pp. 387-398.
- [46] G.I. Susman, R.D. Evered, "An assessment of the scientific merits of action research", *Administrative science quarterly* 23(4), 1978, pp. 582-603.
- [47] L. Mathiassen, M. Chiasson, M. Germonprez, "Style Composition in Action Research Publication", *MIS Quarterly* 36(2), 2012, pp. 347-363.
- [48] J. McKay, P. Marshall, "The dual imperatives of action research", *Information Technology & People* 14(1), 2001, pp. 46-59.
- [49] M. Miles, A. Huberman, *Qualitative data analysis: an expanded sourcebook*, 2nd ed., Sage Publications, Thousand Oaks, 1994.
- [50] NOPTIS, "Nordic Public Transport Interface Standard", 2019. <http://www.noptis.org/>. (Accessed 2019-04-24).
- [51] CEN TC278 Working Group 3 Sub Group 9, "Network Timetable Exchange", 2019. <http://netex-cen.eu/>. (Accessed 2019-04-24).
- [52] European Union, Commission Delegated Regulation (EU) 2017/1926 in: *European Commission (Ed.) Official Journal of the European Union*, 2017.
- [53] CEN TC278 Working Group 3 Sub Group 7, "Service Interface for Real-time Information", 2019. <http://www.transmodel-cen.eu/standards/siri/>. (Accessed 2019-04-24).
- [54] European Committee for Standardization (CEN), "Transmodel", 2019. <http://www.transmodel-cen.eu/model-visualisation-html/>. (Accessed 2019-04-24).
- [55] OpenTripPlanner PLC, "OpenTripPlanner", 2019. <https://www.opentripplanner.org/>. (Accessed 2019-04-24).