



Open or Closed? Open Licensing of Real-time Public Sector Transit Data

Teresa Scassa*, Alexandra Diebel**

*Faculty of Law, Common Law Section, University of Ottawa, 57 Louis Pasteur, Ottawa, ON K1N 6N5, Canada. Email/phone: teresa.scassa@uottawa.ca, phone: 613-562-5800 ext. 3872

** J.D., University of Ottawa. Email: alexandra.diebel@gmail.com

Abstract: This paper explores how real-time data are made available as “open data” using municipal transit data as a case study. Many transit authorities in North America and elsewhere have installed technology to gather GPS data in real-time from transit vehicles. These data are in high demand in app developer communities because of their use in communicating predicted, rather than scheduled, transit vehicle arrival times. While many municipalities have chosen to treat real-time GPS data as “open data”, the particular nature of real-time GPS data requires a different mode of access for developers than what is needed for static data files. This, in turn, has created a conflict between the “openness” of the underlying data and the sometimes restrictive terms of use which govern access to the real-time data through transit authority Application Program Interfaces (APIs). This paper explores the implications of these terms of use and considers whether real-time data require a separate standard for openness. While the focus is on the transit data context, the lessons from this area will have broader implications, particularly for open real-time data in the emerging ‘smart cities’ environment.

Keywords: open data, real-time data, transit data, intellectual property, licensing;

Acknowledgement: This work was supported by the Social Sciences and Humanities Research Council of Canada through the Geothink partnership grant. The support of the Canada Research Chairs Program is also gratefully acknowledged. Thanks also to Charles Sanders and to Wei Jiang for their helpful comments on the draft of the manuscript. An earlier version of this paper was workshopped with Geothink colleagues; their feedback is, as always, much appreciated.

1. Introduction

The open data movement has seen governments at all levels, in countries around the world, engage in the release of government data sets without charge and under licenses that impose few, if any, restrictions on data users. Open data are used by civil society groups for a range of purposes including promoting government transparency and accountability. They are also used by developers who draw on data from a variety of sources to make comparisons, perform analyses, or create innovative “mash-ups”. According to Cerrillo-i-Martinez (2012 at 217) the main objectives of the open data movement are “to make public sector information available to the public to increase transparency and participation as well as to achieve greater administrative efficiency and to create economic value from public data.” According to Kitchin (2014 at 48), the goal of open data is “to democratize the ability to produce information and knowledge rather than confining the power of data to its producers and those in a position to pay for access.” Open data principles and best practices are emerging to ensure broad reuse of government data. This has led to the development of, among other things, data standards to facilitate interoperability of data sets from diverse sources, and template data licenses.

There is no complete consensus on the essential characteristics of open data. However, in general, most definitions require that to be open, data must be: (1) made available in a reusable format; (2) prepared according to certain standards; and (3) available under an open license with minimal restrictions or conditions imposed on reuse. This paper focuses on the third element – open licensing. While much has already been written about open licensing in general and the licensing of open data in particular, this literature has concentrated on the licensing of static data. While static data sets are often downloaded through a portal in a one-time operation, real-time data must be accessed on an ongoing basis and often at fairly short intervals. This mode of access typically places greater demand on the resources of the data custodian and may also present security challenges.

In this paper, we consider whether real-time data requires a different approach to open licensing than that which has developed around static datasets. Our study of licensing practices focusses on real-time transit data. More specifically, we look at data that track the location of public transit vehicles. These data are sought by developers to create apps that predict the arrival time of transit vehicles at particular stops. While our inquiry is focused on real-time GPS public transit data, it is expected that similar issues will arise in the broad range of contexts where ‘open’ real-time data are at issue. We consider the extent to which additional licensing conditions are imposed upon users of real-time data, and the manner in which this is done. We also consider why real-time data might be subject to additional terms and conditions. We conclude that there is currently no standard for the licensing of real-time data and offer some suggestions for such a standard.

2. Methodology

Our study considers the open data licensing practices of six U.S. municipal transit authorities and four Canadian municipal transit authorities. The U.S.-based transit authorities are: Chicago Transit Authority, New York Metropolitan Transit Authority, Massachusetts Bay Transit Authority

(MBTA), Portland's Trimet, the Bay Area Rapid Transit Authority (BART), and the Washington Metropolitan Area Transit Authority (WMATA). These transit authorities were selected because each has publicly adopted an 'open' approach to sharing both static and real time transit data. Trimet and BART were among the earliest adopters of open data in the U.S., and all six agencies have relatively mature open data programs. The Canadian municipal transit authorities we considered have also made both static and real-time transit data available as open data. They are: OC Transpo (Ottawa), TransLink (Vancouver), the Agence Metropole de Transport (Montreal), and Winnipeg Transit. These were chosen based both on their size and their relative advancement in open data licensing and in the availability of real-time GPS data.

The selected transit authorities all indicate that both static and real-time data are available as open data. For our study, we downloaded the licenses (terms and conditions) that accompany both real time and static transit data for each transit authority. We also downloaded any separate Application Program Interface ("API") terms of use for developer access to the real-time GPS data. We examined the licenses for each transit authority and compared them across the group of municipal transit authorities.

3. The Ideal of Openness

Open data can mean many different things to different entities. Indeed, Kitchin (2014) observes that the concept of open data can include varying degrees of openness. Open data can range from data made available under a Public Domain Dedication Licence (with virtually no restrictions on reuse) to data that can be accessed and reused subject to onerous terms of use. In spite of this divergence in execution, advocates of open data have articulated definitions that set a kind of gold standard for open data. The 8 Principles of Open Data (2007) articulated by open government advocates are summarized by Cerrillo-i-Martinez (2012 at 216) as requiring that "data must be complete, primary, timely, accessible, machine-processable, non-discriminatory, non-proprietary and licence-free." The oft-cited definition from [Opendefinition.org](http://opendefinition.org) describes open data in these terms: "A piece of data or content is open if anyone is free to use, reuse, and redistribute it—subject only, at most, to the requirement to attribute and/or share-alike". To be open, according to this definition, data must be provided for free reuse subject to minimal or no conditions. This is the definition of openness we adopt as our baseline for assessment of the degree of openness of licences and API terms of use.

Among the 8 Principles of Open Data is the "non-discrimination" principle, which essentially provides that open data should be available to everyone and should not include a registration requirement. Similarly, the G8 Open Data Charter (2013) states that when open data are released, "it should be done without bureaucratic or administrative barriers, such as registration requirements, which can deter people from accessing the data." The absence of a registration requirement is considered important as registration can be seen as a barrier to access and reuse of the data.

Depending upon the degree of variance from the gold standard, questions may arise as to the true "openness" of a given open data project. Other scholarly work on open data has examined the ability of open data to meet the ideals articulated for it (Kitchen 2014) or again, of the veritable

impact of open data given the neo-liberal environment in which it is rolled out (Bates 2012). These discussions are interesting and important. In this paper we use the ideal of openness as a benchmark and consider the extent to which real-time data is actually made open. We focus on transit data and measure how the roll-out of this type of open data meets the posited benchmark. The multiplicity of public authorities making transit data “open” allows us to examine the diversity in approaches to openness.

4. Legal Interoperability

Ideally, open data licensing reflects a public policy decision to make data available for no-cost reuse under the fewest and least restrictive terms and conditions. Open licenses are also meant to be easy to understand so that users are not hindered by uncertainty over what is or is not permissible (Open Data Commons 2016). Open licensing often requires a ‘cultural’ shift for typically risk-averse governments because of the risks and related costs associated with liability – and as a minimum, the costs of defending oneself against liability claims (Kaufman 2012, Kitchin 2014). As Mewhort (2012 at 10) notes, decisions to open data “often come down to a choice between a small degree of risk for the licensor versus greater ‘openness’ and flexibility for the licensee.”

There is no single standard for government open data licenses. This has led to concerns over “legal interoperability” (Mewhort 2012). Legal interoperability refers to the ability to take data (or other material licensed under an open license) and combine it with other material made available under a different open license without having conflicting or overly different licensing conditions for the different elements of the combined product. Legal interoperability is particularly important where a developer is using data (or other materials) from multiple sources (Mewhort 2012). Dietrich (2012 at 14) notes that the lack of interoperability of licenses “raises transaction costs and creates barriers.” Indeed, without legal interoperability, the technical interoperability of the open data qua open data, which aims for a “commons” of data and free intermixing with other “open” data, is of limited value (Open Knowledge International 2016).

In spite of the importance of legal interoperability, there is no standard open government data license. In the United States, where federal works, including compilations of data, are in the public domain (Onsrud & Lopez 1998), licenses for government data are not required (although open licenses are required for proprietary state and municipal data sets). The de jure openness of government data at the federal level in the U.S. may explain a lack of coordinated federal leadership in the U.S. in developing a template open government license. The result is sometimes widely divergent open licenses at the state and municipal levels. The governments of Australia and New Zealand have adopted the Creative Commons (“CC BY”) 4.0 license as their default open licenses (AUSGOAL 2011; New Zealand Government 2014). As CC licenses are translated into many different languages, the use of such licenses can enhance interoperability. The UK has chosen to develop its own public sector open license called the Open Government License (“OGL”), which, unlike the CC BY, is specifically tailored to the government context (Open Government Licence, V3.0). Canada has developed its own open government license as well. Closely modeled on the OGL, it is adapted to the particular Canadian context (Walker 2013).

In federal states such as the U.S. and Canada the different regions (states or provinces) may also choose to adopt their own open data licenses. The Canadian government is working to increase legal interoperability in Canada by encouraging provincial and municipal governments to adopt their own versions of this license, with mixed success (Walker 2013). In the U.S., because there is no federal open government data license that can serve as a model, where there is open licensing of government data, there is considerable variance at the state level. Similarly, in both countries, at the municipal level, there can also be considerable variation from one open data license to another. This lack of standardization can pose challenges for legal interoperability.

5. Transit Data

Among the data sets most highly sought after by developers at the municipal level are those related to public transit. Early open data developments in this context involved the release by municipal transit authorities (“MTA”s) of static transit data (for example, route schedule information) under open licenses (Rojas 2012). Since being opened, municipal transit data has become widely used not only in the map services of major corporations such as Google and Microsoft, but also in public transit apps developed by small companies and solo developers (Dietrich, 2012, Scassa 2014). The evolution of public transit data from closed to open has been the subject of study and analysis (Scassa 2014; McHugh 2013; Rojas 2012; Thomas 2011).

The now widespread adoption of Automatic Vehicle Location systems (“AVL”) (also known as real-time GPS tracking) has allowed municipal transit authorities to gather real-time transit data that reveals the actual location and on-time performance of transit vehicles (Thomas 2011, Ferris et al 2010). Transit authorities have also used this information for safety, planning, emergency response, and other system monitoring purposes (Rojas 2012). This data has been of great interest to app developers, who have sought to use it in apps that provide arrival time information to transit users. As a result, there has been considerable pressure to make AVL data available as open data.

The shift from static data to real-time streaming data in the public transit context reflects developments elsewhere in smart cities. While at one point most urban data was relatively static in nature, the embedding of sensors throughout the smart cities environment gives rise to a vast body of dynamic data that records change, movement, consumption, travel, and progression (Townsend 2013). This data is valuable and interesting, and much of it will be sought after as open data.

The static or dynamic nature of particular transit data affects how developers interact with them. Data such as bus timetables are considered “static” because they are generally only updated at set intervals (for example, on a seasonal basis). As a result, they are typically made available in the form of a dataset that can be downloaded. Updated datasets are published by the transit authority as they become available. An app that relies on static data draws on a fixed set of data; it need only update that data at certain specified intervals that may be months apart. By contrast, AVL data are generated in real-time and on an ongoing basis. To use AVL data in an app that predicts, for example, the arrival times of specific buses at particular stops, developers must be given access to the data at frequent intervals. Thus, in contrast to static data, the developer does

not download the data it needs for its app; rather, the developer must design their app to make calls upon a server for data on a moment-to-moment basis.

Calls on a database to retrieve real-time data are made using an API. Typically, a municipality will develop and publish its own API, which it can modify as and when it needs to, with appropriate notice to developers. Where real-time transit data is made available as open data, there can be two layers of licensing. The data itself may be licensed under an open data license which may be the same one that is used for static data. Such a license predominantly addresses issues of ownership of the underlying data and permission to use them, as well as a limitation of liability for harm resulting from the use of the data. Typically, the downloading and use of the API is subject to a different set of terms and conditions. These API terms of use constitute a contract between the transit authority and the app developer. Since the underlying data cannot be accessed without the developer's agreement to the terms of use for the API, these terms add a layer of contractual obligation which may, in some cases, be significantly different from the terms of the open data license. The Canadian MTA's that we considered currently use such a dual-license approach. The U.S.-based MTAs have shifted towards an approach that involves a single "Developer's License Agreement" that merges the two agreements and covers both the rights to data and the terms of any API use. The result is a shift away from simple open data licensing and the creation of a new licensing approach for certain categories of data. Whether this is justified by the particular circumstances of real-time data, or whether it signals a more general shift away from open data principles is an interesting question, particularly as the collection of real-time data becomes commonplace and smart cities evolve.

The transit data context reveals another dimension of the particular problem of open real-time data - one that also has implications for other real-time smart cities data. Municipal transit authorities typically contract out to the private-sector for the hardware and software that enables the collection of AVL data. Contracts with private sector suppliers raise the possibility that ownership of the data may rest with the private sector company. Indeed, conflicts over ownership of AVL data in the context of MTA decisions to make this data open arose in the early days of open real-time transit data (Scassa 2014). Paradoxically, a municipality's willingness or capacity to contract for ownership of this data is a precondition for open real-time data. Even if municipal ownership is contracted for, a second issue is that of "trickle down" terms, or terms which the MTA, pursuant to the terms of its contract with the private sector supplier, must pass on to its users/developers. Terms that the MTA may be contractually obliged to pass on to developers under any license for reuse of the data might include an 'indemnify and hold harmless' clause, as discussed in the next section, as well as the obligation to obtain the consent of end-users, or the obligation to agree to the license agreement of the private sector company's own suppliers. To the extent that governments contract out for data-related services in the smart cities context, these contracts will pose ongoing challenges to any open data agenda. (Scassa 2016; Kitchin 2014).

6. Discussion

6.1. Registration Requirements

As set out above, generally speaking, registration requirements are considered antithetical to open data (Open Data Charter 2013). This is because they impose an additional barrier to access and may make some users (for example those who seek to use data for transparency purposes) uncomfortable about disclosing their identities to the government. Although from a government perspective, registration requirements might offer some advantages such as knowledge of the categories of stakeholders using the data, registration requirements are usually not imposed for open data.

While registration requirements are typically not imposed for access to static transit open data of the transit authorities considered, only one transit authority imposed a registration requirement on developers seeking to use their API to access real-time data (see Table I). Bay Area Rapid Transit Authority (“BART”) was the only transit authority that did not impose such a requirement, rather registration is optional. However, BART does offer a registration option and indicates that there are advantages to registration. For example, if access to the server must be limited due to excessive demand, these limits will be imposed first upon those who have not registered for the API (BART 2016). At the other extreme, WMATA has made all of its transit data, whether real-time or static, available via APIs. It imposes a single Developers’ License Agreement for both categories of data and requires registration for use of the APIs. This effectively imposes a registration requirement for “open” static data.

Registration requirements for real-time open data highlight a dilemma. While on the one hand, registration requirements are generally considered contrary to open data best practices, they may also decrease the likelihood that an MTA’s API will be misused. For example, because real-time data must be accessed multiple times and at frequent intervals, the APIs could be used in denial of service attacks. A registration requirement may reduce the chances of this happening.

Table I: Registration Requirements and API Terms of Use

Transit Authority	Registration Required for Access to Static Open Data?	Registration Required for API for Real-Time Data?
Chicago Transit Authority	No	Yes
New York Metropolitan Transit Authority (NYMTA)	No	Yes
Massachusetts Bay Transit Authority (MBTA)	No	Yes
Trimet (Portland OR)	No	Yes
Bay Area Rapid Transit Authority (BART)	No	Registration is optional Access by non-registered API users may be limited in periods of high demand
Washington Metropolitan Area Transit Authority (WMATA)	Yes	Yes

OC Transpo (Ottawa)	No	Yes
TransLink (Vancouver)	No	Yes
Agence Metropole de Transport (Montreal)	No	Yes
Winnipeg Transit	No	Yes

6.2. API Terms of Use or Developers' Agreement

In general, the MTAs considered in this study either impose terms and conditions for the use of the API for their real-time data in a license that is separate from and in addition to an open data license, or they address rights to access and use all transit data in a single agreement typically referred to as a 'Developer License Agreement' (see Table II). The Massachusetts Bay Transit Authority (MBTA) has both a relatively open data license combined with a set of Developers' Relationship Principles. Both documents apply to the use of data, whether static or real-time. Although registration is required for use of the API, there are no API terms of use that go beyond what are found in the open license or Developers Relationship Principles. BART has a single license for its data, whether static or real-time. It imposes no additional terms of use for its API, and, as noted earlier, does not even require registration in order to use the API.

Table II: Data Licenses and API Terms of Use

Transit Authority	Open Data License with Separate API Terms of Use	Single Developer Agreement for Static and Real-Time Data
Chicago Transit Authority		✓
New York Metropolitan Transit Authority (NYMTA)		✓
Massachusetts Bay Transit Authority (MBTA)		✓
Trimet		✓
Bay Area Rapid Transit Authority (BART)		✓
Washington Metropolitan Area Transit Authority (WMTA)		✓
OC Transpo	✓	
TransLink	✓	
Agence Metropole de Transport	✓	
Winnipeg Transit	✓	

If the imposition of additional terms and conditions for the use of real-time data can be justified, then the shift to a single Developers' License Agreement that addresses both static and real-time data may have the effect of imposing more onerous conditions on the use of static data than might otherwise be imposed under a more conventional open license. For example, Winnipeg Transit provides all of its data under a Public Domain Dedication License - a type of license that essentially renounces any claims to intellectual property rights in the data. Nevertheless, its API Terms of Use contain additional restrictions. Were Winnipeg to implement a single Developers'

Licence Agreement thereby fusing an open data licence with API terms of use, then this would be a step back from open licensing of the underlying data.

Many of the terms and conditions found in Developers’ License Agreements or in API Terms of Use are not consistent with more conventional open licensing and raise questions about the extent to which access to and use of this data is genuinely open (Perkins and Alpert 2010). Table III provides an overview of some of the more onerous terms and conditions found in the API Terms of Use or Developers’ License Agreements we examined. It should be noted that in general such documents were also lengthier and more legalistic than conventional open licenses.

Table III – Less “Open” Terms and Conditions in Developer Agreements or API Terms of Use

	CTA	NYMTA	MBTA	Trimet	BART	WMATA	OC Transpo	TransLink	AMT	Winnipeg Transit
Indemnify and Hold Harmless	✓			✓			✓	✓	✓	✓
Right to charge fees to developers of for profit apps reserved						✓		✓		
Restrictions on number of calls to database							✓	✓	✓	✓
Waive right to challenge transit authority’s IP rights in its data	✓									
Transit Authority retains right to alter or cancel agreement without notice		✓	✓		✓		✓		✓	✓
Limitation on purpose of use	✓									

Acceptable use restrictions relating to access to the host data server and other computing						✓	✓			✓
--	--	--	--	--	--	---	---	--	--	---

6.2.1. Indemnify and Hold-Harmless

As can be seen from Table III, all of the Canadian MTAs and a significant number of those in the U.S. impose on developers the obligation to indemnify and hold harmless the transit authority for any harm or liability arising from the developer’s use of the data. The contractual requirement to indemnify and hold harmless can impose a significant burden on developers. An example of such a clause, taken from the license of the Chicago Transit Authority is reproduced here:

You agree to indemnify, defend and hold the CTA Parties harmless from and against any and all fines, suits, proceedings, claims, causes of action, demands, or liabilities (including reasonable attorney’s fees) of any kind or of any nature arising out of or in connection with: (i) your use of CTA Data; (ii) a breach or alleged breach of your representations, warranties, agreements or obligations in this Agreement; (iii) any actual or alleged infringement, misappropriation, or violation of any third party’s intellectual property or other rights, including, without limitation, patents, copyrights, trademarks, service marks, or trade secrets by you or your application; and (iv) the development, marketing, distribution, licensing, sale and/or use of any applications, technology, products or services sold, licensed or otherwise distributed by you.

From the perspective of the licensing MTA, such a clause places the legal risk associated with the use of the data altogether on the developer and requires the developer not only to bear this risk, but also to compensate the municipality for any costs associated with its being joined in a lawsuit in relation to the developer’s use of the data. The obligation is imposed not just with respect to uses of the data that infringe the license terms – but to any and all uses of the data. Such a condition substantially reduces the ‘openness’ of the data by imposing the risk of a potentially significant and difficult to assess burden of liability on developers, who must be prepared not only to pay their own legal defense fees and any ordered compensation, but also to pay those of the MTA. Simply put, the developers bear all of the litigation risk associated with the use of the open data. A reasonable developer could decide that the only prudent way to manage these risks is to operate through a corporate entity whose veil would shield them from these risks. Although Kaufman (2012) recommends such clauses in the context of open transit data licenses in general, such terms are not typical of open data licenses. Their presence here raises the question whether there is anything particular to real-time data that justifies the imposition of this obligation. Mewhort (2012) notes that the unrestricted availability of federal data in the U.S., although in the public domain, has not yet lead to lawsuits. However, standard disclaimers in open licenses – to the effect that the data is provided “as is” on an “as available” basis with no warranty of any kind may be all that is needed to manage any risk. All the license agreements that we reviewed contained such disclaimers.

6.2.2. Ownership of Intellectual Property

Indemnify and hold harmless clauses are not the only clauses found in some of the Developers' License Agreements or API Terms of Use that call into question the openness of the real-time data. There is also the question of intellectual property. Generally speaking, the MTAs we reviewed include an intellectual property clause which asserts that the licensor retains title, ownership, rights and interests in the data. However, the Chicago Transit Authority also requires developers, as a condition of the license to agree to not challenge their rights to the underlying data:

You agree not to challenge, directly or indirectly, CTA's rights in the CTA Data or assert any right, title or interest in or to the CTA Data or assist any third party in doing so.

Given the potentially tenuous claim to intellectual property rights in some transit data (Thomas 2011; Scassa 2014), a clause of this kind is problematic because it appears to coerce those who seek to use ostensibly "open" data into accepting the existence and legitimacy of the legal rights that are asserted, but which do not necessarily exist in law.

6.2.3. Fees for Use

Two of the transit agencies considered in this study – TransLink in Canada and the WMATA in the U.S. – address the issue of fees for use of their data. In the case of TransLink, the license agreement provides that TransLink may charge fees and impose additional terms and conditions on commercial developers who "charge an end user a direct or indirect fee to use the Data." In the case of WMATA, the agreement contains a clause titled "License Fee" which states "WMATA does not currently charge a license fee to download the Transit Data." This latter does not create a two-tiered arrangement based on commercial or non-commercial use of the data. Rather, it merely alludes to the possibility that fees for use may be imposed in the future.

While it might be understandable that cash-strapped transit authorities would seek to recover some of the costs associated with the collection of their data, the charging of fees is irremediably inconsistent with the concept of open data. Because real-time data may have more significant commercial applications than static transit data, another question is whether the enhanced value of this public data and the increased costs associated with its collection and dissemination justify the imposition of fees (and thus a retreat from open data). If fees are charged for commercial uses of otherwise open government data, then it effectively creates a two-tiered system in which data is 'open' for non-commercial use, but subject to fees (and potentially other conditions) when used for commercial purposes. As a result, the incentive to use open data and to innovate with it may diminish.

6.2.4. Termination of Agreement

The majority of MTAs include clauses that reserve the MTA's right to alter or terminate the agreement at any time and without notice. This is in contrast to open licenses, which are typically irrevocable. This means that once permission is granted to use the licensed work, it cannot be withdrawn. The CC BY 4.0 license, for example, is an irrevocable license to use the licensed work so long as the licensee is in compliance with the terms of the agreement. Although the license also

provides that the licensor may choose any time to no longer distribute the licensed work, this clause does not alter the right of the licensee to continue to make use of their copy of the work under the terms of the license. Put otherwise, were the owner of the dataset to choose to no longer make it available to the public as open data, the licensee could continue to use the dataset that they had already downloaded – although there would be no future updates. The effect of the OGL is similar to that of the CC BY 4.0. The duration of the license is perpetual, although non-compliance with the terms will result in termination of the license. Under the OGL, the licensor specifically “does not guarantee the continued supply of the Information.”

Although real-time data is different in character from static data, the revocability of licenses for streamed data seems inconsistent with open licensing principles. Where what is licensed is access to a live stream of data on a regular and ongoing basis, the licensor might consider it necessary to retain a right to revoke the license for any number of reasons. These could include the decision to no longer gather the data or to make it available. It might also be because technological circumstances or security issues make it necessary either to limit access to the live streamed data or to change the nature or frequency of that access. While it is true that a streamed data license could be drafted as an irrevocable license with the licensor nonetheless retaining the right to cease supplying the data, the value of real-time data is in the real-time updates. There is likely little value in a provision that preserves the right of the licensee to continue to use the previously licensed data once access to the live data stream has been terminated. Nevertheless, the practice among the U.S. MTAs considered in this study of combining static and real-time data under a single Developers’ License Agreement means that the revocability of the single license diminishes the openness not only of the live data, but also of the static data.

6.2.5. Limits on Calls

One feature absent from the API Terms of Use for the U.S.-based transit authorities, but common among Canadian transit authorities was the imposition of limits on calls on the transit authority database using the API. Ottawa’s OC Transpo set a maximum of 10,000 connections per day; for Winnipeg it was 100 requests per IP address per minute; and for TransLink, the limit was 1,000 requests per day. AMT’s terms of use provided that access could be terminated if calls on the data were made at intervals of less than 30 seconds apart.

Limits on calls to the database might be imposed to avoid unmanageable demand on the server. If, for example, demand on the server results in considerable delays, the reliability of the arrival-time data for users is diminished. This in turn could have a negative impact on the reputation or users’ perception of the transit authority. What is unclear is why the Canadian MTAs impose such limits whereas their U.S. counterparts do not.

It is worth noting that for transit authorities there might also be potential concerns over too few calls on the server. For example, real-time predictions based on relatively infrequent samples of the data might be inaccurate, causing frustration and annoyance among riders who turn their dissatisfaction towards the transit system. While minimum intervals are not addressed in any of the terms of use considered here, the CTA Terms of Use specifically provide that the user of CTA transit data must “agree to maintain the same high standards of quality for any applications bearing the CTA Brand Assets.” This suggests that a Chicago-based real-time transit app that

makes too few calls on the data, thus providing inaccurate arrival time estimates, might violate this term of the license.

While limits might be necessary to manage server traffic, it is not clear whether these limits are set based on hypothetical risks to the stability of the system (for example, if there were wide-scale use of the API by a large number of developers) or on actual experienced usage of the database. Of the U.S. MTAs, only BART makes reference to a potential need to manage demand on the server. As noted above, it provides that developers who do not register for the API might have their access reduced if demand on the server became too high. One hypothesis to explain call limits is that MTAs with more limited data infrastructure may be more inclined to restrict access to their servers, so as not to generate excessive demand on the system, than MTAs with better infrastructure. This proposed explanation would highlight the relationship between the resources made available to support open data and the degree to which the data is truly “open”.

Table IV: Call Limits for Real-Time Data

Transit Authority	Limits on Calls to the Server for Real-Time Data
Chicago Transit Authority	No
New York Metropolitan Transit Authority (NYMTA)	No
Massachusetts Bay Transit Authority (MBTA)	No
Trimet	No
Bay Area Rapid Transit Authority (BART)	No - but non-registered API users may face limits if demand is high
Washington Metropolitan Area Transit Authority (WMATA)	No
OC Transpo	10,000 connections/day
TransLink	1000 requests/day
Agence Metropole de Transport	Requests no more frequent than every 30 seconds
Winnipeg Transit	100 requests per IP address per minute

6.2.6. Limitations on Use

The Chicago Transit Authority developers’ agreement contains another interesting clause that diverges from the open data licensing principle. Specifically, it limits use of the data to the sole purpose of assisting public transportation riders or furthering public transportation. The imposition of a purpose limitation on the use of open data is inconsistent with the principles of open data (8 Principles of Open Data 2007). After all, open data can be used for innovation purposes, for example, to develop apps or other services. It can also be used for a broad range of other purposes like transparency, criticism, or research. A license term that operates in the sole interest of the licensor by permitting it to shut down a potential use that they do not see as furthering their own agenda is both inconsistent with the concept of open data and highly problematic.

6.2.7. Acceptable Use

A few MTAs also included terms in their Developers' License Agreements or API Terms of Use relating to "acceptable use". These clauses barred uses of the data or API in ways that might be harmful to the MTA or to other parties. For example, the prohibitions in the OC Transpo API license prohibits uses that interfere with or disrupt services or servers owned by the City of Ottawa, as well as uses that would circumvent the API keys. The WMATA developers' agreement similarly prohibits such activities, as well as activities that are a misuse of the access provided either because illegal or offensive to public order. Prohibited activities include those that involve impersonation, defamation, the appropriation of ownership, the infringement of intellectual property rights and the promotion of illegal activities or the dissemination of illegal content.

Interestingly, while access to real-time data can create security risks for MTAs, only a minority of those licenses considered in this study specifically address conduct that poses such security risks. Besides OC Transpo and WMATA, Winnipeg Transit reserves the right to terminate access to its servers for breach of the terms of the agreement, or for any activity which "is otherwise unlawful or harmful to others", while the NYMTA provides that users may not use the data in a "false or unlawful" way.

Acceptable use limitations are potentially problematic. While a case could certainly be made that government institutions that provide open data are entitled to insist that such data be used for legal purposes, open-ended purpose limitations can pose problems. Winnipeg Transit's limitation on uses that are "harmful to others" is a good illustration. Some uses of data might be 'harmful' but not illegal; the term 'harmful' is capable of a very broad interpretation. For example, a use of open data that results in the reallocation of resources from one area of a city to another might be considered 'harmful' to those who are deprived of the resource. Clauses that allow for licence termination in open-ended or ambiguous terms should be avoided.

6.3. Summary

Our study shows that in the municipal transit context, many MTAs that identify themselves as providing "open live-streamed data" do so under terms and conditions that are inconsistent with open data licensing best practices. While some of these conditions may be linked to the different realities of providing public access to live-streamed data (registration requirements and perhaps limits on calls to the database), others are more difficult to justify (revocability, indemnification and hold harmless clauses), and some are simply antithetical to any reasonable vision of open data (restrictions on purpose of use, and the right to charge fees for commercial uses).

Beyond this, the high degree of variation from one developers' license agreement to another is a significant barrier to legal interoperability. This could be a problem if real-time transit data is incorporated into apps that also build in other municipal data, or if a developer attempts to integrate data from multiple adjacent transit authorities, for example. Developers who seek to migrate a successful app to new jurisdictions will also face a multiplicity of different legal terms and conditions, creating potentially unnecessary barriers to innovation.

Our study also revealed differences in licensing practices between Canadian and U.S. MTAs. The two sharpest differences are the tendency among Canadian MTAs to impose limits on developers' calls on the host database for live-streamed data, as well as the tendency in Canada to have an open license for the data itself and a separate agreement governing the use of the API.

7. Conclusions

Our study reveals considerable variance in open licensing practices for real-time municipal transit data in Canada and the U.S. In the U.S., the trend appears to be towards a single "developers' license agreement" that covers both real-time and static transit data. In Canada, the MTAs considered currently use a combination of an open data license for the underlying data and a developers' agreement that sets out separate terms and conditions for access to real-time data via the MTA's API.

Few of the API licenses or developers' agreements considered in this study would meet the Open Definition of "open" or the 8 Principles of Open Data. According to these standards, open data licenses are meant to be simple, straightforward, and unburdensome. They are intended to facilitate a widespread use of government data by a broad range of users that may include legally unsophisticated users. Complex terms, terms that impose uncertain liability or that require legal advice to assess, are not compatible with open licensing. Almost universally the licenses considered in this study contain terms that are inconsistent with conventional open government data licenses, including the OGL (UK), the Canadian Open Data License, and the CC BY 4.0 license. While a case could be made that some of these conditions are necessary in the context of real-time data, others appear to be both unnecessary and difficult to justify in an open data context. The individual licenses may each be said to be problematic from an open data perspective. Moreover, collectively the number of differences and the wide variation in terms and conditions also raise issues of legal interoperability.

It is not immediately clear why the use of real-time transit data is subject to terms and conditions that are inconsistent with open data principles. Indeed, the example of BART demonstrates that it is possible to make real-time transit data available under an open data license without additional restrictions relating to the API. As set out above, access to BART's real-time data does not even include a registration requirement for the API (although BART does offer certain advantages to those who register). The conclusion is ineluctable: Access to real-time data can be provided under relatively open terms. Thus, a key question for those government agencies that make real-time data available as open data is whether there are any differences in the nature of real-time data that justify the inclusion of additional terms of use in the license.

Security issues may be a factor for transit authorities. After all, access to real-time transit data involves frequent, ongoing interaction between the transit authority's server and application developers' software. However, it is not clear what the full range of security concerns are (beyond, for example, the risk that the API might be used to carry out a denial of service attack), or whether these are even resolved by the additional terms and conditions. The ability to maintain a certain level or quality of service may also be an issue for some municipalities who would not want excessive demand on their servers to compromise the quality or availability of arrival time

predictions. Nonetheless, although the potential burden on the server from multiple requests from a large number of developers appears to be a concern for some municipalities (particularly those in Canada), the variation in call limits imposed in the different licenses are not obviously linked to actual demand. The justification for such clauses may therefore be questioned.

It is possible that well-developed open licensing communities, and the proliferation of template open licenses, including licenses specific to open data may have both encouraged and supported the development of open data licensing for municipal transit data, and informed approaches to the licensing of that data. However, what the transit context reveals is that when it comes to terms of use for APIs for real-time data, there are no generally-accepted templates. The result is a high degree of divergence, and the imposition of certain terms and conditions that are varyingly inconsistent with and even antithetical to the underlying open data principles. With a general increase in real-time municipal data across a range of service areas, it is clear that attention needs to be paid to the development of standards for access to real-time data in a municipal open data environment.

We propose that open licensing for real-time data should follow principles that are similar to those for open data. The goal should be to support innovation and to further transparency by providing free access to data for relatively unrestricted reuse. While real-time data does raise some issues that are different from static data, it is difficult to accept that the conditions need be substantially different from those for access to open data. We therefore recommend the following:

- 1) Terms of Use for access to real-time data that are being made available as open data should be as compliant as possible with both the Opendefinition.org definition and with the underlying principles of open data. This means that restrictions on access to the data and to its reuse should be minimal. In particular, neither fees, nor limitations on the legal uses to which the data can be put should be imposed on developers. The need to make the data available as open data should be part of any negotiations in contracting with private sector companies for hardware and software associated with real-time data.
- 2) Given the potential security issues involved with providing access to real-time data, it is not inherently unreasonable to impose a registration requirement as a condition of using an API to access real-time data.
- 3) Similarly, it is not unreasonable to provide that the license can be terminated for conduct that is malicious or harmful to the licensor's systems or servers, or to the public more generally, or where conduct or uses to which the data are put are illegal. However, any such termination clause should not be overly broad. It should be as specific as possible, and should focus on harms that require a precise and immediate response.
- 4) A clause providing that the licensor does not guarantee a continued supply of real-time data is also acceptable. Most template open licenses already contain terms that warn the user that the licensor is not obliged to continue to provide the licensed datasets or works. Such a term should be framed as a caveat that continued supply is not guaranteed. It should not be expressed as a basis on which the license can be revoked. In situations where the open data license and the API terms of use are combined into a single agreement, there is a risk that revocability clauses will have the effect of terminating rights to use already downloaded data.

- 5) Finally, the trend towards fusing the open data license with the API terms of use is a matter for some concern, particularly where this introduces more onerous licensing terms that then apply to both static and real-time data. There is a difference between a license to use open data and terms of use for an API that provides access to open real-time data. The trend to combine both in a single agreement creates the risk that the openness of all forms of data (whether static or real-time) is affected by additional terms required in the case of real-time data.

References

- 8 Principles of Open Data. 2007. https://public.resource.org/8_principles.html Accessed 06/06/16.
- AUSGOAL. 2011. AUSGOAL Suite of Licences. <http://www.ausgoal.gov.au/the-ausgoal-licence-suite> Accessed 06/06/16.
- Bates, J. (2012). This is what modern deregulation looks like: Co-optation and contestation in the shaping of the UK's Open Government Data Initiative. *Journal of Community Informatics*, 8:2, <http://www.ci-journal.net/index.php/ciej/article/view/845/916> Accessed 06/06/16.
- Cerrillo-I-Martinez, A. (2012). Fundamental interests and open data for re-use. *International Journal of Law and Information Technology*, 20:3, 203-222.
- Dietrich, D. (2012). State of Play: Re-use of Transport Data. European Public Sector Information Platform Topic Report No. 2012/01. http://www.epsplatform.eu/sites/default/files/TopicReport_TransportData.pdf. Accessed 06/06/14.
- Ferris, B., Watkins, K., & Borning, A. (2010). OneBusAway: Results from Providing Real-Time Arrival Information for Public Transit. CHI 2010, April 10-15, 2010, Atlanta, GA, USA.
- Kaufman, S. M. (2012). Getting Started with Open Data: A Guide for Transportation Agencies. 5/1/2012, NY: Rudin Center for Transportation Policy and Management, <http://wagner.nyu.edu/files/rudincenter/opendata.pdf>. Accessed 06/06/16.
- Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and their Consequences*. London: Sage.
- McHugh, B. (2013). Pioneering Open Data Standards: The GTFS Story. In B. Goldstein & L. Dyson (Eds.), *Beyond Transparency: Open Data and the Future of Civic Innovation* (pp. 125-135). San Francisco: Code for America Press.
- Mewhort, K. (2012). Creative Commons Licences: Options for Canadian Open Data Providers. Samuelson-Glushko Canadian Internet Policy and Public Interest Clinic, June 1, 2012. <https://www.cippic.ca/sites/default/files/Creative%20Commons%20Licenses%20-%20Options%20for%20Canadian%20Open%20Data%20Providers.pdf>. Accessed 06/06/16.

New Zealand Government. (2014). New Zealand Government Open Access Licensing Framework (NZGOAL), v. 2. <https://www.ict.govt.nz/assets/Uploads/NZGOAL-Version-2.pdf>. Accessed 06/06/16.

Onsrud, H.J. and Lopez, X. (1998). Intellectual Property Rights in Disseminating Digital Geographic Data, Products, and Services: Conflicts and Commonalities among European Union and United States Approaches (pp. 153-167). In I. Masser & F. Salge (Eds.), *European Geographic Information Infrastructures: Opportunities and Pitfalls*. London: Taylor and Francis.

Open Data Charter. (2013). <http://opendatacharter.net/>. Accessed 06/06/16.

Open Data Commons. Making Your Data Open: a Guide. <http://opendatacommons.org/guide/>. Accessed 06/06/16.

Open Government Licence, V3.0. <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>. Accessed 06/06/16.

Open Knowledge International. Open Data Handbook. <http://opendatahandbook.org/guide/en/>. Accessed 06/06/16.

Perkins, M. & Alpert, D. (2010). Fairfax County reluctant to release open transit data. Greater Greater Washington. <http://greatergreaterwashington.org/post/6197/fairfax-county-reluctant-to-release-open-transit-data/>. Accessed 06/06/16.

Pessoa, L., Reed, L. Tzegaegbe, J., Wong J. & Yan, B. (2011). *Enabling Transit Solutions: A Case for Open Data*. Georgia Institute of Technology, Fall 2011.

Rojas, F.M. (2012). Transit Transparency: Effective Disclosure Through Open Data. Ash Center for Democratic Governance and Innovation. http://www.transparencypolicy.net/assets/FINAL_UTC_TransitTransparency_8%2028%202012.pdf. Accessed 06/06/16.

Scassa, T. (2016). Police Service Crime Mapping as Civic Technology: A Critical Assessment. In press: *International Journal of e-Planning Research*

Scassa, T. (2014). Public Transit Data Through an Intellectual Property Lens: Lessons About Open Data. *Fordham Urban Law Journal* 41, 1759-1810.

Thomas, L.W. (2011). *Legal Research Digest 37: Legal Arrangements for Use and Control of Real-Time Data*, Washington, D.C.: Transportation Research Board of the National Academies. <http://www.trb.org/Publications/Blurbs/165626.aspx>. Accessed 06/06/16.

Townsend, A. M. (2013). *Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia*. New York: W.W. Norton & Co. Inc.

Walker, S.B. (2013). Inspiring Story - Open Government Licence. Open Government Partnership. <http://www.opengovpartnership.org/country/canada/case-study>. Accessed 06/06/16.

Developer License Agreements, Open Data Licenses and Terms of Use

Agence Métropolitaine de Transport:

Licence d'Utilisation des Données Ouvertes de L'Agence Métropolitaine de Transport.

<https://www.amt.qc.ca/Media/Default/pdf/section8/amt-licence-donnees-ouvertes.pdf>. Accessed 06/06/16.

Bay Area Rapid Transit Authority:

Developer License Agreement. <http://www.bart.gov/schedules/developers/developer-license-agreement>. Accessed 06/06/16.

The Real BART API (API Terms of Use). <http://www.bart.gov/schedules/developers/api>. Accessed 06/06/16.

Chicago Transit Authority:

Developer License Agreement and Terms of Use. <http://www.transitchicago.com/developers/terms.aspx>. Accessed 06/06/16.

Massachusetts Bay Transit Authority:

Massachusetts Department of Transportation, Developers' License Agreement.

https://www.massdot.state.ma.us/Portals/0/docs/developers/develop_license_agree.pdf. Accessed 06/06/16.

New York Metropolitan Transportation Authority:

Agreement for Access to Metropolitan Transportation Authority ("MTA") Data Feeds.

<http://web.mta.info/developers/developer-data-terms.html>. Accessed 06/06/16.

OC Transpo:

City of Ottawa, Open Data Terms of Use. <http://ottawa.ca/en/mobile-apps-and-open-data/open-data-terms-use>. Accessed 06/06/16.

OC Transpo Live Next Bus Arrival Data Feed API Account Terms of Service.

http://www.octranspo1.com/developers/terms_of_service. Accessed 06/06/16.

TransLink

Open API Terms of Use. <https://developer.translink.ca/Home/TermsOfUse>. Accessed 06/06/16.

GTFS Terms and Conditions. <https://developer.translink.ca/ServicesGtfs/GtfsData>. Accessed 06/06/16.

Trimet:

Terms of Use. https://developer.trimet.org/terms_of_use.shtml. Accessed 06/06/16.

Washington Metropolitan Area Transit Authority:

WMATA Developer License Agreement. <https://developer.wmata.com/license>. Accessed 06/06/16.

Winnipeg Transit:

Terms of Use for Web Service and API. <http://api.winnipegtransit.com/home/terms>. Accessed 06/06/16.

Public Domain Dedication Licence. <http://api.winnipegtransit.com/home/licence>. Accessed 06/06/16.

About the Authors

Teresa Scassa

Teresa Scassa is the Canada Research Chair in Information Law at the University of Ottawa, Faculty of Law. She is the author or co-author of several books on intellectual property and technology law subjects. She is a past member of the External Advisory Committee of the Office of the Privacy Commissioner of Canada, and of the Canadian Government Advisory Committee on Open Government. She is a member of the GEOTHINK research partnership, and has written widely in the areas of intellectual property law, law and technology, privacy, and open government.

Alexandra Diebel

Alexandra Diebel was a student in the bilingual bijural law program given by the Faculty of Law of the University of Ottawa from 2012 to 2015. During this time she had the pleasure of working part-time as a research assistant with Professor Scassa on the SSHRC funded Geothink project. She is now employed by the Federal Government (Canada) in the area of taxation.