UNIVERSITY OF MICHIGAN

COVINGTON

Contracting Tools for Transportation Data

SUZANNE BELL
Partner, Co-Chair of Technology and IP Transactions Group
Covington

OLIVIA DWORKIN
Associate, Technology, Digital Health & CAV Regulatory Group
Covington

Special thanks to Sachi Spalding, a Summer Associate at Covington & Burling, LLP for her research and preparation of the first draft.

Contents

This white paper was produced in partnership with Mcity Leadership Circle member:

COVINGTON

I. INTRODUCTION

Troves of transportation data can be, and are, produced by smart infrastructure. Municipalities collect various kinds of transportation data, including traffic information such as accidents[1], flows, and volumes[2]; bicycle information such as bike counts[3]; pedestrian information such as pedestrian counts[4]; smart bus stop information;[5] street mapping information;[6] location information for traffic signals;[7] mapping details such as the miles of city streets;[8] and information on roadwork and infrastructure planning[9] such as construction or road closures expected to affect traffic.

Governments, educational institutions, non-profit enterprises, and businesses find transportation data useful for purposes such as improving infrastructure, reducing traffic congestion, improving vehicle and pedestrian safety, providing public security and emergency services, making transportation services more accessible, improving civic planning and design, research and development of new mobility products and services (including machine learning), and researching other potential uses for the data. Wider availability and sharing of transportation data would help to facilitate the development, testing, and adoption of smart infrastructure and connected and automated modes of transportation (collectively, "**smart mobility products and services**").

However, there are barriers to the accessibility of transportation data for these purposes. One is that there is a lack of standardization and clarity in the permissions granted when transportation data is[10] made available, and another is that privacy and other concerns prevent much of the data that could be useful from being made available; an example of the latter is the discontinuation of a smart streetlights project in San Diego due to concerns about the potential use of transportation data for surveillance purposes.[11]

This paper explores license provisions and contracting tools that data providers can consider using when making transportation data publicly available. Part II describes the kinds of provisions that data providers typically include in the licenses or other terms and conditions that they apply to transportation data. Part III examines the agreements under which specific municipalities in four states (Michigan, California, Pennsylvania, and Arizona) make transportation data publicly available, including pursuant to template agreements.[12] Part IV identifies additional template agreements that are available for use by data providers when making data publicly available. Finally, Part V sets out key considerations for data providers in choosing the terms under which they make their transportation data available.[13]

II. DATA TERMS GENERALLY

In this paper, a "**data provider**" means any entity, whether public or private, that makes transportation data publicly available, and a "**data user**" means any user of such data. "**Open data**" refers to transportation data that is made publicly available, and "**data terms**" means the terms and conditions applied to such data. Data terms may be characterized as a license, as an agreement, or in some other fashion.[14]

Because the term "license" implies that the data or database is protected by copyright, it is important to understand that may not be the case under U.S. law for data and databases. There is no copyright protection for mere data such as facts and figures,[15] but if the data is an original work of authorship such as an image, then copyright may subsist in that data. A database itself may be protectable under U.S. copyright law as a compilation[16] if it meets the minimum standard for originality in the selection, coordination, or arrangement of the data.[17] Copyright protection of a database does not extend to the data itself. Even if the data or database is subject to copyright protection, certain uses may be permissible under the doctrine of fair use.[18] It can be difficult to

determine in some instances whether copyright applies to particular data and databases, and whether fair use applies to particular uses.

Data terms may be:

- standard, meaning that the data provider has adopted a license agreement that has been published as a template, usually by a non-profit organization, for anyone who wishes to apply it to their content ("**standard terms**"), or
- custom, meaning that the data terms are not verbatim copies of standard terms, and could either be based on standard terms or could be very different entirely ("**custom terms**").

The following types of provisions are commonly included in data terms for publicly available data (though not all of them are present in all data terms):

- permitted uses of the data ("**use rights**"),
- restrictions on use of the data ("**use restrictions**"),
- obligations to apply the same terms to new works based on the data, to make such new works publicly available or to include specific provisions in terms for such new works, or other prescribed terms ("**sharing requirements**"),
- privacy protections, such as an obligation not to attempt to re-identify individuals whose personal data may be included in de-identified form in the published data ("**privacy protections**"),
- obligations to attribute the source of the data in works based on the data ("**attribution requirements**"),
- disclaimers of warranties, statements that the data provider does not guarantee the accuracy, completeness, or timeliness of the data, and the like ("**disclaimers**"),
- limitations of liability, which generally state that the data provider will not be held liable for any claims or damages related to use of the data ("**limitations of liability**"), and
- provisions for which jurisdiction's law governs the data terms, and for how (e.g., litigation in court or arbitration outside of court) or where (e.g. which courts or sites of arbitration) disputes will be resolved ("**governing law/dispute resolution provisions**").

Of these provisions, data users tend to be most concerned with use rights, use restrictions, sharing requirements and, to the extent applicable, privacy protections. Data users need to be sure that they have the rights they need to use the data for their intended purposes, and to clearly understand any associated restrictions. Data users need

to understand whether and to what extent they have an obligation to share the results of research they may conduct using open data, or to share data, information, products or services that they may develop based on the open data. Finally, to the extent that the published open data was based on personal data, a data user will want to ensure that the individuals' identities have been protected.

The other common provisions noted above may be more or less desirable to a particular data user, but are likely not to be material to a decision of whether or not to use a particular open data set. For example, an obligation to attribute the fact that data from a particular source was used can generally be accomplished in a number of ways and therefore may not be unduly burdensome. Disclaimers and limitations of liability are present in all manner of commercial transactions, and will be expected by data users, particularly when the data is made available at no charge. Finally, the governing law and dispute resolution provisions may not be optimal to the data user, but are unlikely to be the deciding factor in whether a data user chooses to use a particular open data set.

III. DATA TERMS USED BY MUNICIPALITIES

An analysis of data terms available for transportation data sets in Michigan, California, Pennsylvania, and Arizona revealed that municipalities use both standard and custom terms, and that the data terms vary quite a lot.[19] Details are set forth in the Appendix.

A: Custom Terms. Municipalities that use custom terms take a variety of approaches, ranging from simple terms that are quite permissive, to more restrictive terms.

1. Mere Disclaimers/Limitations of Liability. One type of custom terms merely sets forth disclaimers and limitations of liability, but does not include any specific use rights. Without a specific use right, business enterprises may be reluctant to use the data for commercial purposes because, as noted above, it can be difficult to determine whether copyright subsists in the data or database and whether fair use applies, and commercial enterprises may not want to risk a claim of copyright infringement.
2. Permissive Use Rights. Some custom terms are very permissive, expressly allowing commercial and other uses. These broad rights are likely acceptable to a wide range of data users.
3. Vague Use Rights. Some custom terms grant use rights that are stated in vague terms, such as "informational purposes only" or "for public use." Providers of

commercial products and services may not find this approach to be clear enough to ensure that they may use the data to develop their products and services. A statement saying that the data is "open data" is another example of vague use rights, but does not seem to be ambiguous in its intent that the transportation data is not subject to any restrictions on use and therefore is likely to be acceptable to commercial users.

4. Ambiguous Copyright-Related Terms. Some custom terms state that, while broad rights are generally granted, certain of the data may be protected by intellectual property rights and that no licenses are granted. This type of statement leaves it to the data user to determine what data might be copyright-protected, and when it's not obvious one way or the other, a data user may be reluctant to use the data, especially for commercial purposes.
5. Obligations to Share Derivatives. Some custom terms require the data user to share certain results of use of the data, whether with the data provider or publicly. Sometimes the obligation is clear, such as an obligation to share a database that builds on the original database but adds more data; sometimes it is not clear, such as an obligation to share "derivatives" or "derivative works." "Derivative work" is a copyright term and not all data or databases are copyright-protected so what a "derivative" is meant to convey is unclear. For example, an obligation to share an enhanced database on the same data terms as the original database may be acceptable to many data users, while a term that requires a commercial data user to disclose details of an artificial intelligence ("AI") model built using the open data or the software that implements the model is likely not to be acceptable to commercial users. This type of requirement can be a real barrier to use of open data in smart mobility products and services since the reason that many commercial entities are interested in using publicly available data is for machine learning.
6. Other Obligations. Some custom terms apply other obligations on the data user, such as an obligation to attribute the data set or data provider in any published work.

B: Standard Terms. Municipalities have chosen a number of different standard terms to apply to their data.

1. PDDL. The Open Data Commons Public Domain Dedication and License v1.0 (**"PDDL"**) license[20] is used by some of the municipalities studied. The PDDL license can be applied to a database or to the data within the database, or both, and places the database and its contents "in or as close as possible within the public domain." As such, there are no restrictions on, or requirements arising from, use of the data. The license includes a waiver of any copyright protections and database rights that

can be waived, and a license to such rights that cannot be waived. The license expressly allows commercial use. Because it was designed specifically for data and databases, the PDDL is well-suited to transportation data and therefore is likely to be acceptable to data users who wish to use the licensed data.

2. Creative Commons Licenses. The Creative Commons 0 1.0 Universal (**"CC0"**) license[21] is similar to the PDDL; all copyright and database rights that can be waived are waived and a backup license is included if the waiver fails. In addition, the data provider (referred to as the "Affirmer" in CC0) affirms that if the data terms are judged legally invalid or ineffective, they will not exercise any copyright or database rights in the published work. Though not specifically designed for data and databases, this license is also likely to be acceptable to data users.

a. The CC0 license is just one of many "Creative Commons" licenses. Given the prevalence of Creative Commons licenses, it is worth going into some detail about their applicability to databases and data.

b. By their terms, Creative Commons licenses only govern activity that is protected by copyright (or the European Union sui generis database right).[22, 23] Accordingly, in the U.S., if a particular activity is not within the scope of the exclusive rights conferred by copyright (reproduction, preparation of derivative works, distribution, public performance, and public display),[24] or if the activity is covered by an exception to copyright such as the fair use doctrine, then Creative Commons licenses do not purport to restrict that activity.

c. According to the Creative Commons website, "even where database contents are subject to copyright and published under a CC license, use of the facts and ideas embedded within the contents will not require attribution (or compliance with other applicable license conditions), unless doing so implicates copyright in the database structure as explained above. This important limitation of all CC licenses is highlighted on the license deeds in the Notice section, where we emphasize that compliance with the license is not required for elements of the material in the public domain."[25] In the case of databases, fair use may permit copying of a copyright-protected database for the purpose of extracting uncopyrightable material in many circumstances.[26]

d. To further emphasize this point, the Creative Commons Wiki page on the subject of "NonCommercial interpretation" further emphasizes that the "NonCommercial term does not limit uses otherwise allowed by limitations and exceptions to copyright . . . . In such cases, the CC license never comes into play and the NC restriction (and other limitations or conditions contained in the license) may be disregarded."[27]

e. When the Creative Commons license does in fact apply, the "non-commercial" term could be of concern for any for-profit company that is using the licensed data in a way that ultimately may lead to a commercial product. The Creative Commons "share-alike" sharing requirement, which requires derivative works of the licensed work (data or databases, as applicable) could be read to apply only to a new version of the licensed database to which new data has been added, but also could be read more broadly to apply to an AI model that was trained using the data.[28]

f. Accordingly, for databases licensed under a Creative Commons license, the data user needs to determine: (1) whether the data itself is protected by copyright; (2) whether the database is protected by copyright; (3) whether, if the database or data is protected by copyright, the doctrine of fair use would permit the data user's contemplated uses; (4) if copyright or database rights apply and the doctrine of fair use does not apply, whether the contemplated use is non-commercial or commercial; and (5) how to comply with the attribution and share-alike requirements. The risk of making a wrong decision (a potential copyright infringement claim) can be a barrier to use of the data for smart mobility products and services. This isn't an issue for the very permissive Creative Commons 0 1.0 Universal (CC0) license, but can be for other varieties of Creative Commons licenses.

g. The fact that Creative Commons licenses are so frequently used for transportation data that appears to be purely factual, and for databases of transportation data that do not appear to reflect a unique selection, coordination, or arrangement, can give data users pause when considering whether to use that data, particularly when the Creative Commons license governing the data limits the use to "non-commercial" or includes a "share-alike" requirement. While the potential data user may take comfort that the restrictions may not in fact apply because of the lack of copyright protection in the underlying material or the applicability of the fair use doctrine, the publisher of the data, by virtue of selecting this license, appears to have intended that the restrictions apply and data users may be reluctant to use the data for smart mobility products and services in that circumstance.

h. Based on this analysis, Creative Commons licenses (other than CC0, which is very permissive) are not the optimal choices for data providers to use for databases because of the ambiguity they create for data that is not subject to, or may not be subject to, copyright or sui generis database rights protections.

3. Open Data Commons Open Database License (**"ODbL"**)[29] and Database Contents License (**"DbCL"**).[30] The ODbL license is broad and allows the data user to share,

create, and adapt the database through transformation, modification, or utilization of the database or the production of works from the database, and expressly allows commercial use. There is a sharing requirement that applies only to derived databases, but does not apply to works resulting from use of the database contents. This means the sharing requirement would not apply to an AI model trained on the data, or a product that incorporates that AI model. Similarly, the attribution requirement applies only to the database and derivative databases.

a. The ODbL licenses the copyright rights in the database but not the data itself, so if the data itself is subject to copyright protection, this license alone would not be adequate for a user of the data (unless fair use applies). The DbCL is a companion to the ODbL and grants a license to the database contents, and also expressly permits commercial use. The DbCL also grants a license under copyright, and allows commercial use. The sharing requirement does not appear to apply to database contents.

b. The ODbL and DbCL are both likely to be acceptable to a range of users of open data.

IV. OTHER STANDARD LICENSES

There are standard terms besides the PDDL, ODbL, and DbCL discussed above that have been prepared specifically for data and therefore avoid the potential ambiguities that arise from using terms or licenses that are not specifically drafted for data use. The ones listed here vary in complexity.

C. Open Use of Data Agreement (O-UDA)

The O-UDA[31] permits unrestricted use of the data and databases to which it has been applied, and does not impose any restrictions or limitations on distribution of the results of use of the licensed data, including AI models trained on the data, so long as the results contain no more than a de minimis portion of the data. There are sharing requirements such as obligations to apply attribution and disclaimers to the licensed data itself on redistribution; this requirement is likely acceptable to a wide range of data users.

D. Computational Use of Data Agreement ("C-UDA")

The C-UDA[32] allows for full computational use of the licensed data. Redistribution of the data, unmodified or modified, must be on the same data terms and with attribution to the original, but the C-UDA imposes no requirements on use or distribution of the

results, including AI models trained on the licensed data. This agreement is useful for data providers who want to grant rights specifically for computational use.

E. Open Data Commons Attribution License v1.0 ("ODC-By")

The ODC-By[33] license only covers copyright and database rights in the database itself, but grants no rights to the contents of the database; the user would have to separately ensure that it can use the data itself. This license grants broad rights to the database, and requires that the licensed database and any derivative databases or collective works that include the database be distributed on the same data terms as the ODC-By license. Works produced using the database do not need to be distributed under the data terms of the ODC-By license, but the user must include an attribution stating that the produced work was based on the licensed database with a link to the database, and a link to the ODC-By license. Without an express right to use the data itself, this would not be an optimal choice if the goal of the data provider is to facilitate wide use of the data for smart mobility products and services, but may be a useful choice for a data provider who sees value in the compilation of the data though it does not have the rights it would need to convey a license to the data itself.

F. Community Data License Agreement - Permissive, Version 2.0 ("CDLA-Permissive")

The CDLA-Permissive[34] license allows use of the data made available under the license, and requires only that the same license be applied to redistribution of the data, with or without modifications. This requirement expressly excludes any results of use of the data, including any outcome by computational analysis of the data such as machine learning models, and therefore gives clarity to some of the concerns raised by other licenses.

G. Data Use Agreement for Open AI Model Development ("DUA-OAI")

The DUA-OAI[35] is designed to be a signed agreement entered into between a data provider and a data user. Use of the data is limited to development of AI models, and the trained AI models must be made publicly available under an open source software license. This agreement requires confidential treatment of the data, and prohibits attempts to re-identify any individual from anonymized or de-identified personal data. This template is useful for data providers who want to control who their users are and to ensure that privacy-protecting terms are in place.

V. CONSIDERATIONS IN SELECTING LICENSE TERMS FOR TRANSPORTATION DATA

Government entities seeking to make transportation data available for use by businesses and other users need to consider the following in relation to their objectives. For example, if their objectives are to make transportation data available for mere academic research, that would dictate one set of choices. If their objectives include making transportation data available to businesses for use in developing smart mobility products and services that could ultimately benefit their citizens, that would dictate another set of choices. Regardless of the objective, having clear lanes and guardrails will facilitate the use of transportation data by the data provider's users.

A. Use Rights

Rights to use the transportation data and database need to be clearly spelled out. In particular, if use is limited to non-commercial use, the line between commercial and non-commercial use needs to be understandable. For example, if transportation data is used by a commercial entity to train an AI model that ultimately may be used in a commercial product or service, but the transportation data itself would not be included in that commercial product or service (other than perhaps a de minimis amount), the data provider needs to specify whether that is commercial or non-commercial use of the data.

B. Use Restrictions

Apart from the commercial/non-commercial distinction, the data provider needs to consider whether to impose any restrictions on use of the transportation data. An obvious example is a prohibition on use for any illegal purpose, but there could be other sensitive uses that a data provider wishes to prohibit.

C. Privacy Protections

If the transportation data made available by a data provider has been derived from personal information, the data provider may want to include a prohibition on attempts to re-identify an individual or other privacy-protecting terms.

D. Sharing Requirements

If a data provider wishes to have users of its transportation data also make publicly available the results of their use of the transportation data, it is critical to specify the reach of this requirement, e.g., whether it applies to (i) an expanded database that includes the licensed data, (ii) an AI algorithm or model developed using the licensed data, (iii) the weights of such a model, or (iv) the source code that embodies an algorithm or AI model

developed using the licensed data. In each case, the trigger for any sharing requirement also needs to be specified, such as commercial sale, release for testing purposes, or the like. The term "derivative" or "derivative work" should be avoided unless its meaning is clearly defined in the context of the particular transportation data and data use at issue, since "derivative work" is a copyright term and thus its application to data and databases can be ambiguous when the data and/or database may not in fact be subject to copyright protection.

E. Intellectual Property Matters

Data providers should consider specifying which IP rights are licensed. For example, if the database is licensed but the data it contains is copyright-protected material such as images, then it would be important to specify whether the images are licensed, or whether the data user has to determine for itself whether fair use applies or whether it needs to seek a license from the copyright owner. As another example, if a data user might choose to seek patent protection for an AI model, and if the license includes a sharing requirement for that AI model, it would be important to specify whether that patent is licensed as part of that sharing requirement.[36]

In addition, because the doctrine of fair use can be complicated to apply in practice, transportation data providers may want to consider specifying that they are not seeking to limit the applicability of this doctrine or other exceptions to the applicable copyright to database law.

F. Attribution

If the data provider wishes to have the source of the transportation data identified by data users, it is important to specify when this requirement applies. For example, a data user needs to understand whether the attribution requirement applies only upon distribution of an enhanced database that includes the licensed transportation data and additional data, or whether it also applies to commercial distribution of a smart mobility platform that incorporates an AI model that was developed in part using the licensed transportation data.

G. Disclaimers; Limitations of Liability

A data provider generally wants to include disclaimers regarding the transportation data and to limit its liability for use of the transportation data; these tend to be non-controversial.

H. Governing Law; Dispute Resolution

Decisions as to whether to include a governing law/dispute resolution provision, and what

to include in such a provision, are specific to the nature of the data provider. For example, government entities may be required by law to use their states' laws as the governing law and their states' courts as the forum for dispute resolution.

VI. CONCLUSION

Careful consideration of the points outlined in Part V and clearly drafted data terms will facilitate the use of transportation data, regardless of the data provider's objectives in publishing the transportation data. If the data provider's objective is to foster the development of smart mobility products and services, then a number of the standard licenses covered in this paper could be good choices such as the Open Use of Data Agreement (**"O-UDA"**) and the Computational Use of Data Agreement (**"C-UDA"**). It also is, of course, possible to craft a custom license that is comprehensive and clear enough such that data users can understand and rely on the data terms in using the data.

Of the terms reviewed in this paper, Creative Commons licenses are generally not the best choices for transportation data because Creative Commons licenses were not designed to apply to data that is not protected by copyright or database rights. It can be difficult to determine when copyright protection does and doesn't apply to databases and data; much of transportation data is likely factual and therefore not protected under U.S. copyright law, and the published databases containing transportation data are unlikely to show sufficient originality in the selection, coordination, or arrangement. Similarly, it can be difficult to determine whether the sui generis EU database protection rights apply. Further, when Creative Commons licenses do apply to data and databases, it can be difficult to interpret share-alike and non-commercial-use requirements in the context of certain likely uses of transportation data such as machine learning and AI model development.

Mere disclaimers and limitations of liability, without an express grant or other statement about the right to use the data, are also not optimal, as they may not provide specific enough guidance for a data user to be confident that their contemplated uses of the published transportation data are authorized.

It is also important to have the relevant data terms easily accessible to a potential data user, such that when they identify transportation data that they believe may be useful to them, they can also easily see, or easily see how to navigate to, the applicable data terms.

APPENDIX

A. Michigan

Several of the 50 most populated Michigan cities make transportation data publicly available on their websites, including Grand Rapids, Ann Arbor, and the Michigan Department of Transportation.

1. Grand Rapids

The material provisions of the Grand Rapids data terms[37] for bus stop data are disclaimers and a dispute resolution provision. The disclaimers disclaim any liability for usage of the published data and any association of the city with the data or products produced from the data post-download, and state that the data may not be accurate and that it is not representative of any views of the city. The data terms also state that the city provides data as a "complementary service to its residents," without further explanation as to whether the intent is to limit use of the data to residents.

2. Ann Arbor

The Ann Arbor data terms[38] for transportation data provided by the city[39] state that the transportation data is provided for "informational purposes only." There is a requirement to notify the City if the user modifies, uses, or presents data supplied by the City "in a manner other than originally presented," and to include a disclaimer with the data indicating that the data has been modified from its original source. The data terms further include (i) a statement that the content of the Ann Arbor-provided databases is in the public domain unless it has a copyright notice, and a commitment on the part of the city to make reasonable efforts to ensure that any third party copyrighted information—such as imagery—is labeled clearly; (ii) disclaimers, including as to accuracy; and (iii) limitations of liability, including a release.

3. Michigan Department of Transportation

The Michigan Department of Transportation provides a number of transportation data sets. The Terms of Use for the GIS Open Data portal state that the data sets are a public record and "there are no restrictions on the use, reproduction, or distribution."[40] There is a statement on the portal, however, that maps and other material are protected by copyright and that the data terms do not apply to other materials or content, including maps or logos, that may be located on the site or portal containing the data sets and may be protected by intellectual property rights such copyright, trademark, or patent. Accordingly, while the data sets are made available without restriction, maps and other copyright-

protected materials are not licensed.

B. California

Of the top 50 most populated cities in California, several cities made transportation data available pursuant to data terms.

1. San Diego, San Francisco and Santa Ana; Napa County

San Diego,[41] San Francisco,[42] and Santa Ana[43] use the standard Open Data Commons Public Domain Dedication and License v1.0 (**"PDDL"**) license,[44] as does Napa County.[45]

2. Los Angeles and Oakland

Los Angeles[46] uses the standard Creative Commons 0 1.0 Universal (**"CC0"**) license.[47] Oakland[48] has an extensive data portal but individual data sets on that portal may have different data terms; one data set uses the CC0 license.[49]

3. Salinas

Salinas[50] uses the Open Data Commons Open Database License (**"ODbL"**).[51]

4. Riverside County, San Jose, Long Beach, Anaheim, Chula Vista, Fremont, San Bernardino, Hayward, Visalia, and Victorville

The County of Riverside,[52] San Jose,[53] Long Beach,[54] Anaheim,[55] Chula Vista,[56] Fremont,[57] San Bernardino,[58] Hayward,[59] Visalia,[60] and Victorville[61] use custom terms for their publicly available transportation data. The Long Beach data terms are permissive, explicitly allowing commercial and/or personal use; the Long Beach data terms also include disclaimers and limitations of liability stating that usage of the data does not indicate endorsement by the city, that there are no warranties, and that the city has no liability. Anaheim has brief data terms that expressly allow commercial use and state that users are free to copy, redistribute, and adapt the data. Other data terms for the cities listed above are permissive but indicate that there may be other data terms that apply, so further investigation would be required for individual data sets to know the full scope of applicable provisions; Riverside and Victorville are examples of this approach. Victorville's data terms further state that if the city claims or seeks to protect any patent,[62] copyright, or other intellectual property rights in the data, including the derivative work, the city's website will indicate on the webpage on or from which the data is accessed, and that the data terms do not grant to the user any right, title, or interest in or to any patent, copyright, or intellectual property right that the city and/or any third party may have in the data, including the derivative work. Visalia's data terms include a disclaimer and an attribution requirement, and a clause that states that there may be fees charged for some of the costs associated with the processing, handling, and distribution of the data. A few

cities and counties listed in this section use data terms that consist merely of disclaimers and limitations of liability. The San Jose data terms simply state that the data is provided "as is" with a disclaimer that the data user takes full risk and responsibility over data usage. Similarly, the Chula Vista data terms merely set forth an "as is" warranty and a disclaimer for liability relating to errors in the data. San Bernardino's data terms have an "as is" warranty and a statement that the county has no liability for errors or any use by the user. Other listed cities' data terms are even briefer. Fremont's data terms are one line and simply state that the data is "open data;" Hayward's are also one line and merely say "for public use."

C. Pennsylvania

Six Pennsylvania municipalities were identified that made data publicly available pursuant to data terms. No two cities had the same data terms.

1. Pittsburgh

Pittsburgh makes transportation data available through the Western Pennsylvania Regional Data Center (**"WPRDC"**), which allows each data provider to assign a license to each database that the provider chooses to share.[63] Pittsburgh has released its transportation data primarily under the CC0[64] license described above. Pittsburgh has also released some transportation data under the Creative Commons Attribution 4.0 International (**"CC-BY"**)[65] license. The CC-BY license, like all Creative Commons licenses as noted above, applies to copyrightable material and material protected by database rights. The material covered by this license may be copied, redistributed, transformed, or built upon, including for commercial purposes. This license includes an attribution requirement. To satisfy the attribution requirement, any creator identification (unless the creator requests that it be removed), copyright, warranty information, reference to the license, and a link to the license must be retained. All uses must also indicate whether the material was modified and note that it is licensed under the CC-BY. The WPRDC's own license[66] (which must be agreed to before accessing any data set) includes the usual disclaimers and limitations, but also includes a prohibition on including "Non-Public Information" (defined below) in any data set that is published, prohibits users from using any Non-Public Information that has accidentally been released, and requires the user to notify the WPRDC of any Non-Public Information that the user discovers.[67]
"Non-Public Information" means information that may not be disclosed to the public for the following reasons: "The information is exempt from disclosure or the information is prohibited from being disclosed under State and Federal Laws and regulations including the Pennsylvania Right to Know Act, 65 P.S § 67.101 et seq., the Criminal History Record Information Act, 18 Pa.C.S. §§ 9101-9183, the Health Insurance Portability and

Accountability Act (HIPAA), and other applicable privacy laws; The information is covered by a contractual non-disclosure obligation; The information is covered by confidentiality and fiduciary obligations; or The information is private, proprietary or privileged."

2. Reading

Reading has an Open Data Platform[68] that includes some transportation data. One sample data set included on that platform did not specify any data terms with respect to the use of that data.[69]

3. Harrisburg, Philadelphia, Centre County, and York County

Harrisburg,[70] Philadelphia,[71] Centre County,[72] and York County[73] all make some transportation data available as part of their open data programs. Harrisburg's data terms consist merely of an "as is" disclaimer and Centre County's data terms are the same as Harrisburg. York County's data terms consist of disclaimers. Philadelphia's data terms include an obligation for the data user to hold the City harmless from claims that arise from use of the data.

D. Arizona

In Arizona, of all cities with a population over 10,000, eleven cities have released data under data terms. Of these eleven, two—Tempe[74] and Phoenix[75]—use the CC-BY license. Seven use custom terms and the other two use two different standard terms.

1. Scottsdale

Scottsdale[76] licenses its data under the ODbL, discussed above.

2. Maricopa Pinal

Maricopa Pinal[77] uses the Creative Commons Attribution-ShareAlike 2.0 Generic[78] ("CC BY-SA") license, discussed above.

3. Tucson, Mesa, Glendale, Goodyear, Maricopa, Pima County, Cochise County

Tucson,[79] Mesa,[80] Glendale,[81] Goodyear,[82] Maricopa[83], Pima County,[84] and Cochise County[85] all make data available using custom terms. In Tucson, the data terms are merely a disclaimer and limitation of liability. Glendale, Goodyear, Maricopa, Pima County, and Cochise County all use data terms similar to Tucson, with some variations between cities and some variation among data sets. Notably, a Glendale data set[87] requests that all non-proprietary data created, collected, or modified be provided back to the city for public use. The Mesa data terms permit broad use of the data for commercial and non-commercial purposes. They include an attribution requirement for "derivative works," defined as "any work that is based in any manner or to any extent upon the [d]ata, including, without limitation, any work that uses the [d]ata in a modified form."

About Mcity

Mcity at the University of Michigan is leading the mobility transformation. Home to world-renowned researchers, a one-of-a-kind test facility, and on-road deployments, Mcity brings together industry, government, and academia from across disciplines to advance transportation safety, sustainability, equity, and accessibility for the benefit of society.

REFERENCES

1. Traffic Accidents by date, Los Angeles City, https://data.lacity.org/Public-Safety/Traffic-Accidents-by-date/2mzm-av8t (last accessed Jan. 17, 2022).
2. 2019 Traffic Volumes, Michigan Dep't of Transp., https://gis-mdot.opendata.arcgis.com/datasets/2019-traffic-volumes/explore (last accessed Jan. 17, 2022).
3. Bike Volume - Manual Counts, Data San Francisco, https://data.sfgov.org/Transportation/Bike-Volume-Manual-Counts/v4v2-5x7s (last accessed Jan. 17, 2022).
4. Arts District Pedestrian and Bike Counts - LA CoMotion, Los Angeles City, https://data.lacity.org/Transportation/Arts-District-Pedestrian-and-Bike-Counts-LA-CoMoti/mbz9-j2zk (last accessed Jan. 17, 2022).
5. Smart Bus Stops, City of Detroit, https://data.detroitmi.gov/datasets/6ec0b22bc67e4068af4c2f09cb7f31b4_0/explore?location=42.466644%2C-83.141838%2C10.00 (last visited Jan. 17, 2022).
6. Slow Streets Map, Data San Francisco, https://data.sfgov.org/Transportation/Slow-Streets-Map/8rsm-shen (last accessed Jan. 17, 2022).
7. Traffic Signals, Data San Francisco, https://data.sfgov.org/Transportation/Traffic-Signals/ybh5-27n2 (last accessed Jan. 17, 2022).
8. Miles of Streets, Data San Francisco, https://data.sfgov.org/City-Infrastructure/Miles-Of-Streets/5s76-j52p (last accessed Jan. 17, 2022).
9. Los Angeles International Airport (LAX) - Closures, Los Angeles City, https://data.lacity.org/Transportation/Los-Angeles-International-Airport-LAX-Closures/63kx-necw (last accessed Jan. 17, 2022).
10. Though "data" is a plural term, this paper uses the more colloquial and easier-to-read singular formulation.
11. Teri Figueroa, Mayor orders San Diego's Smart Streetlights turned off until surveillance

ordinance in place, The San Diego Union-Tribune (Sep. 9, 2020), https://www.sandiegouniontribune.com/news/public-safety/story/2020-09-09/mayor-orders-san-diegos-smart-streetlights-turned-off-until-surveillance-ordinance-in-place.

12. The descriptions and summaries of licenses and terms in this paper are not complete and are not substitutes for reviewing the licenses or terms themselves. This paper does not constitute legal advice as to any such licenses or terms, or any other matter. Refer to the full licenses and terms at the links in the footnotes and consult legal counsel before choosing to use any data or database made available by a third party, or before choosing which license or terms to apply to any data or database that you choose to make available to third parties.
13. Privacy concerns regarding publicly available data will be addressed by my colleagues in a separate paper.
14. The enforceability of data terms is beyond the scope of this paper; for purposes of this paper, we have assumed they are enforceable, but legal advice should be sought before publishing or using open data.
15. 17 U.S.C. § 102 ("In no case does copyright protection for an original work of authorship extend to any idea, procedure, process, system, method of operation, concept, principle, or discovery, regardless of the form in which it is described, explained, illustrated, or embodied in such work.").
16. Id. A compilation is defined as a "collection and assembling of preexisting materials or of data that are selected in such a way that the resulting work as a whole constitutes an original work of authorship." 17. U.S.C. § 101.
17. Id.
18. 17 U.S. Code § 107 sets out four factors to be considered to determine whether use of a copyrighted work is protected as fair use:

    (1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;

    (2) the nature of the copyrighted work;

    (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and

    (4) the effect of the use upon the potential market for or value of the copyrighted work.
19. When municipalities make transportation data available publicly online, it can be surprisingly difficult to find and identify whether in fact there are data terms that apply to that data set.
20. Open Data Commons Public Domain Dedication and License (PDDL) v1.0, Open Data Commons, https://opendatacommons.org/licenses/pddl/1-0/ (last accessed Feb. 27, 2022).
21. CC0 1.0 Universal, Creative Commons, https://creativecommons.org/publicdomain/zero/1.0/legalcode (last accessed Feb. 27, 2022).
22. Data, Creative Commons Wiki (Oct. 23, 2019), https://wiki.creativecommons.org/wiki/data ("The license terms and conditions apply to the database structure (its selection and

arrangement, to the extent copyrightable), its contents (if copyrightable), and in those instances where the database maker has sui generis database rights, to the rights that are granted those makers.").

23. The European Union database directive is outside the scope of this paper.
24. 17 U.S. Code § 106.
25. Frequently Asked Questions, Creative Commons, https://creativecommons.org/faq/ (last accessed Feb. 27, 2022).
26. See Authors Guild v. Google, Inc., 804 F.3d 202, 214 (2d Cir. 2015).
27. NonCommercial interpretation, Creative Commons Wiki, (Oct. 15, 2017), https://wiki.creativecommons.org/wiki/NonCommercial_interpretation.
28. See generally Attribution-ShareAlike 2.0, Creative Commons, https://creativecommons.org/licenses/by-sa/2.0/legalcode (last accessed April 1, 2022).
29. Open Data Commons Open Database License (ODbL) v1.0, Open Data Commons, https://opendatacommons.org/licenses/odbl/1-0/ (last accessed Mar. 19, 2022).
30. Open Data Commons Database Contents License (DbCL) v1.0, Open Data Commons, https://opendatacommons.org/licenses/dbcl/1-0/ (last accessed Feb. 27, 2022).
31. Open Use of Data Agreement v1.0, Community Data License Agreement, https://cdla.dev/open-use-of-data-agreement-v1-0/ (last accessed Mar. 19, 2022).
32. Computational Use of Data Agreement v1.0, Computational Use of Data Agreement v1.0 - CDLA (last accessed Mar. 19, 2022).
33. Open Data Commons Attribution License (ODC-By) v1.0, Open Data Commons, https://opendatacommons.org/licenses/by/1-0/ (last accessed Mar. 19, 2021).
34. Community Data License Agreement--Permissive, Version 2.0, Community Data License Agreement, https://cdla.dev/permissive-2-0/ (last accessed Mar. 19, 2021).
35. Data Use Agreement for Open AI Model Development, Microsoft, https://news.microsoft.com/wp-content/uploads/prod/sites/560/2019/12/DUA-OAI-1.0.pdf (last accessed Mar. 19, 2022).
36. Patent matters are beyond the scope of this paper.
37. Grand Rapids Bus Stops, City of Grand Rapids, https://grdata-grandrapids.opendata.arcgis.com/datasets/93aaf591cac94ccb9f7cc1b856f0027a_0/explore?location=42.941755%2C-85.704232%2C11.56 (last accessed Feb. 27, 2022).
38. Policies and Notices, City of Ann Arbor, Michigan, https://www.a2gov.org/Pages/Policies-and-Notices.aspx (last accessed Feb. 27, 2022).
39. Ann Arbor's data catalog can be found here: https://www.a2gov.org/services/data/Pages/default.aspx (last accessed Feb. 27, 2022); it includes transportation data such as streetlights and bus stops.
40. 2017 Traffic Volumes, Michigan Department of Transportation, https://gis-mdot.opendata.arcgis.com/ (last accessed Feb. 27, 2022).
41. Traffic Volumes, Data San Diego, https://data.sandiego.gov/datasets/traffic-volumes/ (last

accessed Feb. 27, 2022).

42. Traffic Signals, Data San Francisco, https://data.sfgov.org/Transportation/Traffic-Signals/ybh5-27n2 (last accessed Feb. 27, 2022).
43. Traffic Volumes, Santa Ana GIS Open Data, https://gis-santa-ana.opendata.arcgis.com/datasets/927f0b948fcf400eaf1c13d35239a7b1_0/explore?location=33.742165%2C-117.816097%2C12.30 (last accessed Feb. 27, 2022).
44. Open Data Commons Public Domain Dedication and License (PDDL) v1.0, Open Data Commons, https://opendatacommons.org/licenses/pddl/1-0/ (last accessed Feb. 27, 2022).
45. addresses_all, Napa County, http://gis.napa.ca.gov/giscatalog/viewXML.asp?Name=maingis.GIS.Addresses_All&meta_style=fgdc#Identification_Information (last accessed Feb. 27, 2022).
46. LADOT Traffic Counts Summary, LADOT, https://data.lacity.org/Transportation/LADOT-Traffic-Counts-Summary/94wu-3ps3 (last accessed Feb. 27, 2022).
47. CC0 1.0 Universal, Creative Commons, https://creativecommons.org/publicdomain/zero/1.0/legalcode (last accessed Feb. 27, 2022).
48. Abandoned Autos, City of Oakland, https://data.oaklandca.gov/Infrastructure/Abandoned-Autos-CCD7-Starting-7-1-17/kuxb-xxnt (last accessed Nov. 14, 2021); Oakland Street Trees, City of Oakland, https://data.oaklandca.gov/Environmental/Oakland-Street-Trees/4jcx-enxf (last accessed Feb. 27, 2022).
49. Datasets, City of Oakland, https://data.oaklandca.gov/browse?limitTo=datasets (last accessed Feb. 27, 2022).
50. Traffic Counts at Signalized Intersections, City of Salinas, https://cityofsalinas.opendatasoft.com/explore/dataset/trafficcountsatsignalizedintersections/custom/ (last accessed Feb. 27, 2022).
51. Open Data Commons Open Database License (ODbL) v1.0, Open Data Commons, https://opendatacommons.org/licenses/odbl/1-0/ (last accessed Mar. 19, 2022).
52. County of Riverside, California, https://data.countyofriverside.us/Administrative-and-Fiscal-Services/Open-Data-Terms-of-Use-Final/ymmr-7dcv (last accessed Feb. 27, 2022).
53. Transportation Analysis Zones, City of San Jose, https://gisdata-csj.opendata.arcgis.com/datasets/624255ba86e64eb4b0a25124d61f1f2b_532/explore (last accessed Feb. 27, 2022).
54. Centerlines, Data Long Beach, https://datalb.longbeach.gov/datasets/05a58d8596d24613b1f77e8758d98c19_0/explore (last accessed Feb. 27, 2022).
55. Bus Stops, City of Anaheim, https://data-anaheim.opendata.arcgis.com/datasets/54058d7103044cb69b269d046c16646c_37/explore?location=33.830196%2C-117.881650%2C12.71 (last accessed Feb. 27, 2022).
56. Traffic Volumes, City of Chula Vista, https://chulavista-cvgis.opendata.arcgis.com/datasets/85e926165c3d4c81800ad398959a4863_5/explore?location=32.628598%2C-117.018000%2C12.89 (last accessed Feb. 27, 2022).
57. Intersections, City of Fremont, https://fremont-ca-open-data-cofgis.hub.arcgis.com/

datasets/ceded4848c094e2f98a555556c252f6a_0/explore?location=37.523858%2C-121.997129%2C12.08 (last accessed Feb. 27, 2022).

58. Street Network Current, County of San Bernardino, https://open.sbcounty.gov/datasets/9766aa6294ee40bf9ec5eadcf05105cc_0/explore (last accessed Feb. 27, 2022).
59. Public Works, City of Hayward, https://opendata.hayward-ca.gov/datasets/e44cd379bb76479daa8a5541f5f48303_0/explore?location=37.644750%2C-122.076900%2C12.98 (last accessed Feb. 27, 2022).
60. GIS Digital Data Release Agreement/Disclaimer, City of Visalia, https://geodata.visalia.city/ (last accessed Feb. 27, 2022).
61. Open Data Terms of Use, City of Victorville, https://opengis-victorville.opendata.arcgis.com/pages/bd0fc0ed2cf34374ac82c91fdba04954 (last accessed Feb. 27, 2022).
62. Though patent rights are outside of the scope of this paper, this is one of the few licenses that addresses patents specifically.
63. Data Licenses, Western Pennsylvania Regional Data Center, http://www.wprdc.org/data-licenses/ (last accessed Feb. 27, 2022)
64. Right-of-Way Permits and Traffic-Obstruction Permits, Western Pennsylvania, Regional Data Center, https://data.wprdc.org/dataset/right-of-way-permits (last accessed February 27, 2022).
65. 311 Data, Western Pennsylvania Regional Data Center, https://data.wprdc.org/dataset/311-data (last accessed Feb. 27, 2022).
66. Data Licenses, Western Pennsylvania Regional Data Center, http://www.wprdc.org/data-licenses/ (last accessed Feb. 27, 2022)
67. This is one of the few licenses reviewed that addresses privacy.
68. Reading Open Data Platform, City of Reading, https://data.readingpa.gov/ (last accessed Feb. 27, 2022).
69. Public Bike Parking Facilities, Open Data Platform City of Reading, https://data.readingpa.gov/Quality-of-Life/Public-Bike-Parking-Facilities/phn5-jf (last accessed Feb. 27, 2022).
70. Disclaimer of Liability, City of Harrisburg, https://harrisburg-open-data-cohbg.opendata.arcgis.com/apps/16d0e3c2233f4b98b30d49eb240bb32a/explore (last accessed Feb. 27, 2022).
71. City of Philadelphia Terms of Use, OpenDataPhilly, https://www.opendataphilly.org/organization/about/city-of-philadelphia (last accessed Feb. 27, 2022).
72. Edge of Pavement, Centre County Government, https://data-yorkcountypa.opendata.arcgis.com/datasets/5771e1fc17884d83a10fe115749c084a_0/explore?location=39.968387%2C-76.689750%2C10.00 (last accessed Feb. 27, 2022).
73. Roads, York County, Pennsylvania, https://data-yorkcountypa.opendata.arcgis.com/datasets/5771e1fc17884d83a10fe115749c084a_0/explore?location=39.970150%2C-76.689750%2C10.29 (last accessed Feb. 27, 2022).
74. 1.08 Crash Data Report, City of Tempe, https://data.tempe.gov/datasets/0c333bd164d64d62aa0ee6f99b1ccf82_0/explore?location=33.389171%2C-

111.927944%2C12.31 (last accessed Nov. 14, 2021).

75. Street Maintenance Resurfacing Projects, City of Phoenix Open Data, https://www.phoenixopendata.com/dataset/street-maintenance-resurfacing-projects (last accessed Nov. 14, 2021).
76. Open Database License Agreement, Scottsdale, Arizona, https://www.scottsdaleaz.gov/AssetFactory.aspx?did=69351 (last accessed Nov. 14, 2021).
77. Freeways in Maricopa and Pinal Counties, Arizona, Maricopa Association of Governments, https://geodata-azmag.opendata.arcgis.com/datasets/068d325725e842db9aed0be43a4892f6_0/explore?location=33.542426%2C-113.602439%2C17.30 (last accessed Nov. 14, 2021).
78. Attribution-ShareAlike 2.0, Creative Commons, https://creativecommons.org/licenses/by-sa/2.0/legalcode (last accessed Nov. 14, 2021).
79. Traffic Signals - Open Data, City of Tucson, https://gisdata.tucsonaz.gov/datasets/3046f4c7e4e547358f07c6aeb86a3350_27/explore?location=32.199750%2C-110.907750%2C11.65 (last accessed Mar. 19, 2022).
80. Open Data Terms of Use, Mesaaz, Smart City, https://data.mesaaz.gov/stories/s/Terms-of-Use-page/2dcd-j2nx/ (last accessed Mar 19, 2022).
81. Glendale Post Offices, City of Glendale, https://glendaleaz-cog-gis.hub.arcgis.com/datasets/0014db9ccd954084bf0001363fc7e91d_3/explore (last accessed Mar 19, 2022).
82. Street Light Poles, City of Goodyear, https://city-of-goodyear-gis-hub-goodyearaz.hub.arcgis.com/datasets/e0567bb62a8a4047a1392f82cde27012_0/explore?location=33.283683%2C-112.338519%2C11.28 (last accessed Mar 19, 2022).
83. MapIDs Updated 05/11/2021, Maricopa County, https://data-maricopa.opendata.arcgis.com/content/8d7eca5236824709895d8d0ae97dceec/about (last accessed Mar 19, 2022).
84. Dissolved Street Network, Pima County, https://gisopendata.pima.gov/datasets/f0e1a65ecf2f4a99a3fce4f278bc4ba7_3/explore?location=31.978411%2C-111.875000%2C8.82 (last accessed Mar 19, 2022).
85. Trn CountyMaintainedRoad, Cochise County, https://gis-cochise.opendata.arcgis.com/datasets/a5063cfbaee44d96b017c08a2687d5d0_0/explore?location=0.000000%2C0.000000%2C0.00 (last accessed Mar 19, 2022).
86. Glendale Parks (Poly), City of Glendale, https://glendaleaz-cog-gis.hub.arcgis.com/datasets/glendale-parks-poly/explore?location=33.615378%2C-112.121472%2C9.24 (last accessed Mar 19, 2022).