TOWARDS AN INTERNATIONAL STANDARD FOR TRANSPORTATION PLANNING DATA

Megan Katsumi, University of Toronto[1]
Mark Fox, University of Toronto

Introduction

Transportation planning activities require the collection and integration of data from a variety of sources. As a result, a considerable amount of work involved in transportation planning activities must be expended on data cleaning, semantic reconciliation, and processing. Despite the increasing availability of data, its sources are often siloed; this makes the task of integration particularly challenging. It is difficult to determine how two datasets are related beyond a comparison of unique identifiers, timestamps, or locations, and even these alignments are subject to ambiguity. Often, data is published with little to no metadata and where metadata is provided, textual descriptions still leave many opportunities for misinterpretation.

These challenges for data use and integration may be alleviated with the adoption of a standard data model. Indeed, such standards have been pursued in some limited areas of transportation (particularly at the device-level, for example). The domain of transportation planning has – until recently – not been addressed. This paper presents an overview of the approach that has been adopted for the development of a standard for transportation planning data, and the current status of the specification. The purpose of this paper is to bring attention to this effort in order to maximize the engagement and involvement of stakeholders.

Ontology-Based Standards

An agreed-upon syntax(es) and format for data sources is not sufficient to guarantee that they will be correctly used or combined; it does not ensure that the meaning of the content itself is well-understood. Beyond syntactic interoperability, data silos introduce challenges due to a lack of *semantic* interoperability. That is, in order to support the correct use and integration of data from various sources, it must be possible to combine and exchange data between sources in a meaningful, unambiguous way. The provision of a standard alone does not guarantee semantic interoperability. Any standard may be ineffective if it is not uniformly understood and adopted. Data standards typically rely on detailed documentation or representation tools such as UML in order to convey their intended use. Even in the most detailed efforts, these traditional approaches are subject to ambiguity and thus may be unintentionally misinterpreted and misused.

For example, consider the General Transit Feed Specification (GTFS) (MobilityData, 2020). The GTFS has been widely adopted across various public transit providers and is supported by a variety of third-party applications. It is well-maintained and well-documented – an exemplary data specification. Even still, some inconsistencies arise in its use owing to different interpretations of the field definitions. The definition of loop routes is an interesting example: there are varying interpretations as to whether the first/last stop in a route should be represented once or twice, in the case of a looping route. There has recently been an attempt to address this and other potential issues with the prescription of GTFS Data Best Practices. The Best Practice recommendation is to include the first/last stop twice, however this cannot be determined from the specification itself. It does not assert either that the first stop in a route should be the same as the last, or that the first stop should be captured distinctly from the last stop. As a result, some GTFS data sets may include the first/last stop twice whereas others may not. A minor point to be sure, however it could introduce issues when combining route data from multiple sources. To be clear, this is not a criticism of the GTFS; it is an inherent characteristic of natural language descriptions to be subject to some degree of ambiguity. Best

[1] Presented at the 55th Annual Meetings of the *Canadian Transportation Research Forum*, May 24-27, 2020 at Montreal, QC

practices are one way to address specific issues, nevertheless it would be impractical to attempt to address all possible misinterpretations in this way. Further, such natural language specifications do not facilitate automated compliance checks based on the specification. With ontologies, it is possible to express the specification in such a way as to avoid ambiguity and to enable this sort of automated verification.

Ontologies provide a means of addressing these challenges by making the semantics of a standard computationally explicit, precise and unambiguous. The most widely used definition of an ontology was presented by (Gruber, 1993) and states simply: "An ontology is an explicit specification of a conceptualization". In the literature, there exist a range of artefacts that are identified as ontologies, from basic glossaries to formal theories specified in highly expressive logics. The ontology-based standard described here relies on the use of formal logic with explicitly defined, computer-interpretable semantics. As a result, ontologies are able to support the various knowledge management and reasoning services, including: semantic integration, data validation, and inference. In a given application, an ontology may be employed for one or many of these services.

The Standard

The iCity Transportation Planning Suite of Ontologies (TPSO) described in (Katsumi & Fox, 2019), defines an ontology-based standard for transportation planning. Both the ontology-based formalism and the scope distinguish the TPSO from existing standards efforts. The standard is broad in scope as it has to cover the transportation system, its users, and how they use it. It includes city infrastructure such as road and transit data, as well as demographic information such as descriptions of persons and households.

The standard is proposed in three parts, according to the level of generality of scope:

1. Foundational Level: The first level covers the foundational data: concepts such as time and space that could be expected to appear in datasets of nearly any domain.
2. City Level: The second level covers data specific to the city domain, but generic in the sense that such data could be expected to be both generated and consumed by many city services.
3. Service Level: The third level will be comprised of multiple standards, including – the first such standard – a standard for transportation planning. Each standard at this level will cover data particular to a city service domain. The data modelled at this level may be consumed by a variety of city services, but would only be expected to be produced by one such service and thus will be defined in the standard for the service that produces it.

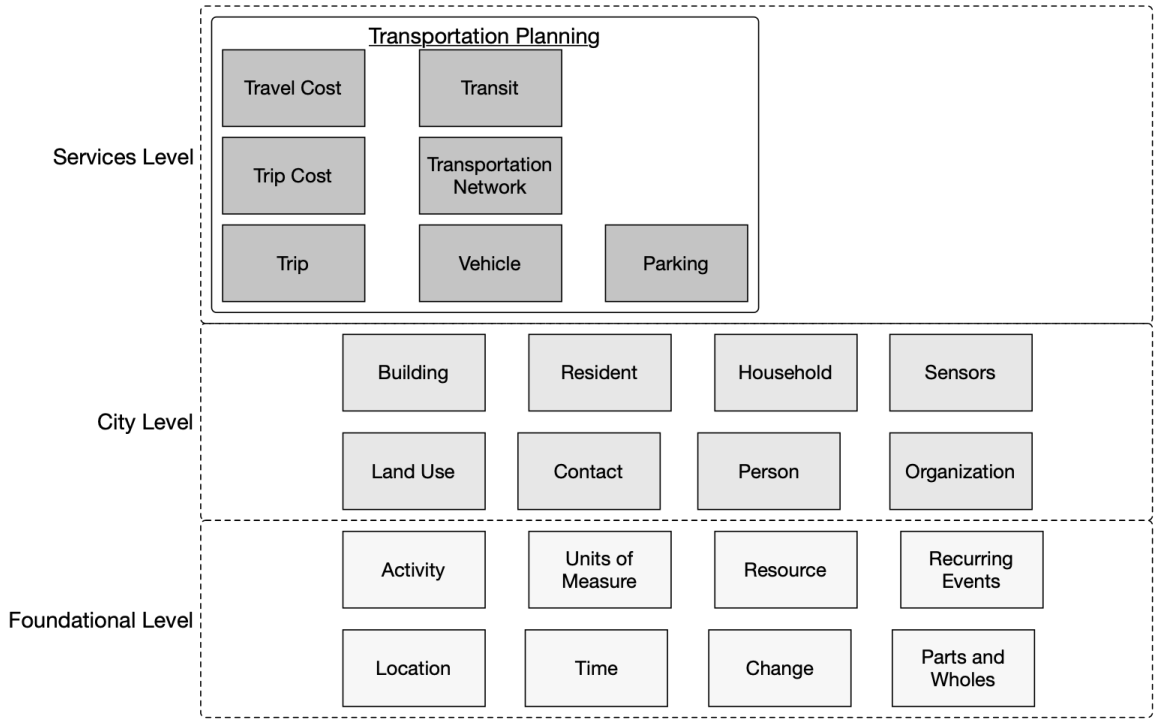

Figure 1: Levels and initial modules in the proposed transportation planning standard.

As illustrated in Figure 1, the transportation planning data model is being developed as the first in a series of city service-level data model standards.

A Collaborative Approach

The iCity TPSO is only a first step toward a standard for transportation planning data, because an intensive effort on the collection of feedback and input from stakeholders remains to be completed. Rather than take a prescriptive approach to the standard's specification, our aim is to develop this standard in an open way, with input from international standards organizations and other stakeholders. To this end, the content of the iCity TPSO will provide the basis for the development of a transportation planning standard – including foundational and city level concepts – that will be developed through collaboration and input open to all interested stakeholders and standards groups.

We have created a web-based platform that will facilitate this collaboration. The platform supports an ongoing conversation as to what concepts and properties should be included until convergence is achieved. The platform, available at citydata.utoronto.ca, allows registered users to propose new concepts and properties. It also allows users to comment on proposed concepts and properties. Any proposed content must have an associated use case to substantiate its purpose in the standard.

A key aspect of the process will involve the specification of definitions for the concepts and properties that are proposed for inclusion in the ontology upon which the standard will be based. These definitions will ensure that all involved contributors have a shared understanding of the terms that are being considered. Further, the definitions will evolve through the collaborative process, resulting in artefacts that will then form the basis of the standard.

The platform encourages standards development via an open, global conversation, where the global community of stakeholders can converge on a set of concepts and properties to be included in the standard. A key aspect of this approach is that it is not expected that the community will converge on a single

definition for the terms in the standard. Instead, the system will support the creation and inclusion of multiple different definitions for a term. The standard will then identify the *maximal common definition* for each such term. A maximal common definition identifies the semantics for a term that are shared between all of the proposed definitions. The purpose is to clearly identify the shared meaning that all stakeholders have in common regarding a particular term. Distinct interpretations can then be clearly identified and related to one another. A key requirement in these definitions is the provision of a use case. This provides clear motivation for new terms and definitions, and will serve as useful documentation for the resulting standard.

Looking Forward

As noted, the transportation planning standard will be one part of a larger city data model. It will be a first step toward a solution for the challenges of resulting from data silos in cities. The city data model standard that results from this effort will serve as the foundation for other standards for city data services. The structure and functionality adopted by the website will serve as a template to facilitate communication for the development of these standards as well.

At the time of writing, the official standardization of the transportation planning standard is in the early stages. Currently, a main objective is to promote awareness of and participation in this initiative.

Acknowledgements

This project was supported by the Ontario Ministry of Research and Innovation through the ORF-RE program.

References

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition, 5*(2), 199-220.

Katsumi, M., & Fox, M. (2019). An Ontology-Based Standard for Transportation Planning. To appear in: Joint Ontology Workshops (JOWO), Graz, Austria.

MobilityData. (2020). *GTFS: Making Public Transit Data Universally Accessible*.