

Transmodel and GTFS – Comparison and Convergence

Briefing Paper for the Public Transport Coordination Group (PTIC)

Nick Knowles, Peter Miller and Paul Drummond

Version 4 (Issued), 6 February 2009

Abstract

The UK public transport information community needs to exchange relevant public transport information between its members in order to provide the travelling public with information that would enable them to make intelligent travel choices. Transport operators also need to exchange data between their many different back office functions such as scheduling, journey planning, real-time, operations, fares collection, etc., in order to improve the effectiveness and efficiency of their operations and processes.

TransXChange is currently the standard of choice within the UK for exchanging timetables; it implements a subset of the CEN Transmodel conceptual model and includes detailed information to support both operational and passenger timetable functions. However, two parallel external drivers may have an impact on this position:

- ▶ The Networks and Timetable Exchange (NeTEx) standard is being developed as a European standard also based on the Transmodel conceptual model, with a wider scope than TransXChange;
- ▶ The Google Transit Data Feed Specification (GTFS) is Google's proprietary format for submitting passenger timetable content to Google Transit and is likely to be a significant target format because of Google's international clout. It lacks timing and operational data and some other key features.

This paper compares the GTFS with the Transmodel conceptual model and from this proposes a convergence path so that TransXChange and NeTEx would be fully compatible and interoperable with GTFS. This would open opportunities for the UK transport information community with no major downsides.

The paper also considers some of the real practical issues that arise in managing and cross-checking complex information interchange models based on "flat" CSV files. It advocates the use of XML, which allows the encapsulation of data into versionable, self-describing information "packages" that also facilitate the exchange of incremental ("delta") changes to the underlying information.

Actions Required

- ▶ Note the findings and conclusions of this paper (referencing the full paper as appropriate);
- ▶ Provide feedback, discussion and debate on the proposals;
- ▶ Support the activities needed to converge GTFS and Transmodel;
- ▶ Support the greater adoption of XML-based information exchange standards and processes throughout the UK transport information community and the proposal of Transmodel as a worldwide standard.

Approach Taken

The goal of this paper is to establish a mapping between GTFS and Transmodel terms, to express the contents of the GTFS model in a minimal subset of Transmodel and to suggest an XML schema structure with which GTFS could be encoded – which could be NeTEx. The objectives were to:

- ▶ Establish a mapping between the current GTFS and Transmodel and any identify any gaps in either;
- ▶ Express GTFS as a revised Transmodel-based model;
- ▶ Sketch out a possible schema structure (that could then be supported by NeTEx).

A layered approach was used that recognised that there were differences in the language and semantics used by GTFS and Transmodel. Having established how GTFS could be interpreted using Transmodel, some key process and protocol requirements for a common exchange format were considered. Ultimately, the ability to use a common worldwide model and common multipurpose exchange format would deliver benefits and outcomes that would be valued by all stakeholders.

Language / Semantics

Some of the most familiar and apparently simple concepts (such as a route, a service or a stop) turn out to have a complex set of diverse meanings associated with them:

- ▶ The same term used in GTFS and Transmodel-based standards may refer to different concepts;
- ▶ Just because a term is used with technical precision in one place does not mean it will not be used informally with a different sense elsewhere.

Transmodel assigns more precise technical meanings to its chosen terms, typically reserving the use of a particular word for a single concept. Consequently, Transmodel will be used to provide the baseline definition of terms and concepts, with the GTFS terms translated accordingly.

Protocols

GTFS and Transmodel

Transmodel can be used as a design tool, providing a precise language and conceptual model with which to understand the current capabilities of GTFS. Based on proven representations that correspond to industry practice, limitations of the current model can be identified and future enhancements suggested. The limitations of GTFS include such things as (a comparison of features is provided in Table 1):

- ▶ Rail services with splitting trains.
- ▶ Multi-operator services;
- ▶ Stop areas, connection links and interchanges;
- ▶ Station and transit interchange navigation and accessibility;
- ▶ Connection protection or guaranteed connections;
- ▶ Services across midnight or time-zones;
- ▶ Day types;
- ▶ Stop labelling;
- ▶ Transport modes, vehicle equipment & accessibility.

The ***proposed Transmodel GTFS model is close to that of GTFS***, with a number of refinements:

- ▶ GTFS-*trip* is renamed VEHICLE JOURNEY;
- ▶ GTFS-*service* is renamed to be more specifically a VALIDATION CONDITION;
- ▶ The GTFS-*calendar* (which currently only has DAY OF WEEK) is normalised into a PERIOD and DAYTYPE;
- ▶ GTFS-*zone_id* is translated as a renamed TARIFF ZONE;

- ▶ GTFS-*transfer* is renamed CONNECTION LINK and then, as an enhancement, refined into an additional distinct abstraction: SERVICE JOURNEY INTERCHANGE (this may be GUARANTEED);
- ▶ GTFS-*shape* is considered to be a ROUTE PROJECTION (although strictly speaking it is only the shape of an individual VEHICLE JOURNEY, and not the whole ROUTE);
- ▶ GTFS-*fare_id* is translated as a renamed FARE ELEMENT (which associates a FARE ELEMENT PRICE with a zone-to-zone DISTANCE MATRIX element);
- ▶ A nesting of Stop Place elements is allowed so that large interchanges can be built up from smaller ones;
- ▶ A VEHICLE JOURNEY can reference a JOURNEY PATTERN;
- ▶ GTFS-*agency* is renamed AUTHORITY.

There are some minor points from the translation that should be considered:

- ▶ The GTFS-*route_type* / Transmodel MODES should be refined. Coach is distinct from bus, and Light Rail/Suburban rail is distinct from tram/street car;
- ▶ An AUTHORITY can indicate a default currency type to use on all fares unless overridden;
- ▶ Some additional attributes are suggested for OPERATING DAY (a note so that an explanation can be given to user and a start and end time so that a day boundary other than midnight can be specified);
- ▶ A version frame is suggested for Vehicle Journey (and indeed all other top level elements);
- ▶ Use of the GML ***coordinates*** attribute could be used to encode latitude & longitude;
- ▶ Intervals could be expressed using the XM Duration data type which supports hours, minutes and second formats;
- ▶ It should be clarified whether the Frequency ***headwayInterval*** is the minimum interval or the maximum interval;
- ▶ UTC time zones should be placed on individual times where relevant.

CEN TC278 WG3 SG9 has proposed the creation of a networks and timetable exchange (NeTEx) model based on Transmodel. It will be useful to compare this with both GTFS and TransXChange.

Technologies for Exchanging Data – XML versus CSV

There are very real practical issues in managing and cross-checking complex models serialised into a large number of flat comma-separated variable (CSV) files. Although CSV formats are generally more compact, they put a greater burden on all users to interpret and use the interface correctly and limit the reuse of elements. In effect, systems would be more closely coupled. The continued use of CSV formats is not precluded.

Nevertheless, an alternative, fully compatible, XML representation of GTFS could help with the validation and consistency of data for many users. The use of XML would also make it easier to exchange incremental changes (deltas) to the data rather than the full dataset. XML has some significant technical advantages over CSV for data exchange in offering:

- ▶ A rich object model suitable for representing large and complex models (as found in public transport);

Table 1: Comparison of Features (Transmodel-based and GTFS)

Feature	TransXChange NaPTAN	GTFS	NeTEx IFOPT
Operators	Yes	Yes	Yes
Stops / Stations	Points	Points	Full interchange
Routes	Yes	Yes	Yes
Tracks	Yes	Yes	Yes
Lines	Yes	Yes	Yes
Journey patterns	Yes	No	Yes
Real-Time timing info	Yes	No	Yes
Timetabled journeys	Yes	Yes	Yes
Journey interchange	Yes	No	Yes
Availability conditions	Yes	Limited	Yes
Operational days	Proposed	No	Yes
Full rail support	No	No	Yes
Fares	Stages Only	Basic Zones	Basic Model
Distributed admin	Yes	No	Yes
Peer-to-peer exchange	Yes	No	Yes
Model based	Yes	No	Yes
Reusable elements (CML)	Yes	No	Yes
Loosely coupled	Yes	No	Yes

- ▶ A self-describing, versionable representation suitable for loosely coupled distributed implementations;
- ▶ A degree of encapsulation so that related objects can be packaged together as a coherent whole;
- ▶ The reuse of component models, giving implementation and documentation savings;
- ▶ The delivery over many different data transports (e.g. FTP, email, http, TCP/ IP sockets) both as a bulk file format and as the payload of protocol requests exchanging just groups of elements.

Processes

It is important to pay attention to the robustness and efficiency of the implemented processes:

- ▶ How well the approach works when used to support actual data management processes;
- ▶ Gathering, aggregating and exchanging the diverse data elements inevitably involves the repeated exchange of large datasets over distributed computer systems, using different toolsets;
- ▶ Supporting multiple versions of data exchange formats and toolsets simultaneously.

Outcomes and Benefits

The Transmodel GTFS schema demonstrates full compatibility and interoperability with both Google's GTFS and Transmodel-based schemas such as TransXChange and (in the future) NeTEx. As such, this may be helpful as a contribution towards an ISO standard for public transport models.

The UK transport information community can continue to implement TransXChange to realise the benefits they already expect as well as opening up new opportunities and realising further benefits from the compatibility and interoperability with GTFS. Future Transmodel-based standards, such as NeTEx, will facilitate the exchange of richer datasets without jeopardising the benefits already being realised. GTFS suggests some additional attributes that may be useful in TransXChange and also provides an example of a basic fare model that might be of relevance to both TransXChange and NeTEx.

References and Acknowledgements

This paper is a briefing for PTIC from the more comprehensive paper by Nick Knowles of Kizoom, and Peter Miller of ITO World: "A Transmodel based XML schema for the Google Transit Feed Specification – With a GTFS / Transmodel comparison" (29 December 2008 - <http://www.transmodel.org.uk/schema/doc/GoogleTransit/TransmodelForGoogle-09.pdf>).

The authors of the main paper acknowledge the contributions, guidance and review comments from:

- ▶ Kasia Bourée, the leading French Transmodel expert;
- ▶ Joe Hughes, Google;
- ▶ Roger Slevin, Transport Direct.

PTIC

The Public Transport Information Coordination Group (PTIC) is a community enterprise, created and operated collaboratively by ATCO, RTIG and Traveline, with the active participation of DfT Transport Direct.

Its purpose is to provide coordination on UK activities relating to the compilation and presentation of public transport information, whether to the public or for operational purposes.

Contacts

Paul Goodwin, ATCO
(PGoodwin@BucksCC.gov.uk)

Mark Cartwright, RTIG
(Mark.Cartwright@CentaurConsulting.co.uk)

Peter Stoner, Traveline
(traveline@PJStoner.globalnet.co.uk)

Chris Gibbard, Transport Direct
(Chris.Gibbard@DfT.gsi.gov.uk)