

Traveline Information Limited

Response to the Government Consultation 'Making Open Data Real'

Traveline Information Limited (TIL) is a not-for-profit company limited by shares and a wholly owned subsidiary of the Confederation for Passenger Transport (CPT). TIL was set up as the legal body for the delivery of the national Traveline service and represents a successful voluntary partnership between transport operators, local government and local authorities that has been established for 12 years.

Traveline provides public transport schedule and route data for use in its own regional and national SMS, call centre, web, and API services as well as to Transport Direct. The Traveline API provides scheduled data for every bus stop in the UK real time data from 34 local real time servers and is offered free to user up to set usage thresholds, which have not yet been exceeded by any user, for 63 developers to date. It receives 750000 hits per month and growth rate is currently at 25% per annum.

Traveline has just let a contract to collate the data from all eight English regions, Scotland, and Wales into a single Traveline National Dataset (TNDS). Its plans are to seek development costs from stakeholders and make the TNDS available free of charge. For further reading, Annex A to this response includes the draft proposal for the TNDS that was recently approved by the TIL Board.

So open data is very high on our agenda and we welcome the debate on opening up data and are grateful for the opportunity to respond to this consultation.

The first observation that we wish to make is that in reading 'Making Open Data Real' in conjunction with the Consultation on Data Policy for a Public Data Corporation the two papers seem quite fundamentally to contradict one another. The former starts with the assumption that data should be free and is very much promoting and encouraging the opening up of public service data and the right to data, whilst the latter assumes that data is charged for and seems to say 'but it's OK for Government to charge for data and here are some of the ways that we can do that'.

TIL would like to see a tighter definition of the term 'Public Service'; this definition helps to define 'Public Data' which underpins the sense of the whole document and is open to interpretation in its current form. For example, where would privately owned bus companies sit in relation to this definition and would it differ between services that are subsidised by local authorities and those that are operated on a wholly commercial basis? In a similar vein clarity over the distinction between data, close to its creation, and information, close to the end-user, would also be welcome in setting the scope for Open Data. This would be fundamental in deeming whether for example the Open Data regime should permit Freemium type arrangements or whether these will be determined by categorisation of base, largely as created Open Data and processed, value added information that may be charged for outwith the Open Data regime.

In relation to Setting Open Data Standards the one to four star ratings for Open Data make sense but what is the business case for 5 star linked data? This content is potentially expensive and time consuming to build and maintain and we would need to be convinced that the additional benefits that this would bring the consumer were equal to or greater than the effort in compiling it and that it would not be more efficient for developers to make and customise these links themselves. The term

'information intermediaries' also needs to be more clearly defined before any accreditation and guidance can be considered for such organisations or services.

It is too early to say whether there would need to be sanctions to enforce a right to data, or obligations on public bodies and public services to publish their data, but where there was the need for them they should sit within the regulatory framework and not as a separate process.

We believe that there should be a sector based data inventory for Public Bodies and about Public Services. Data would be prioritised for inclusion in an inventory by its intrinsic or commercial value and its ease of publication. It is impossible to say what data is published unnecessarily in the absence of an inventory or knowledge of existing datasets. Data should be published without a warranty but with clear metadata about issues such as timeliness, obsolescence and perceived quality and should be fit for its gathered purpose; but whether it is fit for the user's purpose is the responsibility of the user and high quality may not be required.

Prioritisation of datasets for publication should be established by their stated or perceived level of demand and their suitability for/cost of publication. Where publishing a broader set of data is quicker or less costly than a more detailed version of existing data then this should be provided as an interim on the basis that it is better than no data, but again these decisions would be influenced by demand for the datasets.

We do not believe that there is a major ongoing role for government to stimulate innovation but supporting and promoting existing organisations such as the Technology Strategy Board and academia to encourage the use of open data and competition in its use would be welcomed. A further role is in facilitating access to Open Data via data.gov, which could be improved in its presentation, navigation and usability, and also in identifying, highlighting and disseminating best practice. To this end the creation of sector Transparency Boards and the increased connectivity of Open Data to specific sectors, such as transport, is welcomed.

Annex A – TNDS Proposal

A Proposal to the Traveline Advisory Group and National Traveline Board for a Traveline National Dataset

Executive Summary

A Traveline National Dataset (TNDS) can provide a single source of machine readable data in a non-proprietary format at a high level of granularity for all Traveline public transport data for the whole of the UK. It can be distributed via a web-based service with simple and transparent licensing offering free to user basic data with an at cost or low marginal cost for premium access where necessary to cover TNDS delivery costs.

Distribution of a TNDS will be accompanied by an inventory of the data that it includes, clear advice on its format and interpretation, and information about and links where possible to related datasets.

In this way the TNDS can remove existing barriers to access to public transport information, allowing innovation and the development of products and services that will encourage and support the use of public transport services, inform and provide the consumer with choice, and promote growth.

Access to the TNDS will drive data quality improvements through incentivisation of data suppliers and input from TNDS users.

The increased efficiency of distribution of data that the TNDS service can offer will reduce data distribution costs and deliver obligations that Traveline Stakeholders might be required to meet under the government's emerging Open Data policies.

This discussion document allows TAG and NTB to decide how Traveline can deliver the TNDS on a sustainable and affordable basis balanced with minimal restrictions to data access and usage.

TAG and NTB may also wish to consider the risks of not delivering a TNDS.

1. Introduction

The Traveline National Dataset (TNDS) as it exists in its current Beta format is a set of TransXchange 2.1 files for all the public transport services in England, Scotland and Wales that are maintained and supplied by the English Traveline regions, Traveline Cymru and Traveline Scotland. Although the two latter organisations have supplied data to the TNDS, they both reserve the right to collate and distribute their national Traveline datasets independently of the TNDS.

The purpose of this paper is to explore the case for a TNDS in sufficient detail to allow the Traveline Advisory Group (TAG) and National Traveline Board (NTB) to make key decisions about the development of and funding for a TNDS, including the award of a supplier contract for its delivery.

This paper will not address the technical implementation of a TNDS or licensing issues in detail; these will be subject to separate documentation whose creation and content depend to an extent on the decisions made by TAG and NTB in relation to TNDS and in response to this paper.

2. Demand for the Collation and Distribution of a TNDS

There is already significant demand from Traveline Stakeholders and third party organisations for a TNDS. Direct requests to Traveline UK for the dataset have included those from individual developers and open data charities such as mySociety, large transport groups and journey planner suppliers, multi-national organisations such as Google, and from UK government including the Department for Transport (DfT) for use in accessibility and statistical analysis. Many of these organisations are currently testing the Beta TNDS which is described in more detail in section 3 below.

With the exclusion of the Beta TNDS, the only way that third parties can currently get access to a National Traveline dataset is by making agreements with each individual Traveline region, sourcing several datasets from those regions, and combining them to create a national dataset whilst taking into account duplication of files and inconsistencies in data format. The complexity of this task has created a significant barrier to data access; one which after years of trying individual developers through to organisations such as Google have been unable to overcome.

Government consultations on 'Data Policy for a Public Data Corporation' and 'Making Open Data Real', both published in August 2011 and ongoing, look to establishing Public Data Principles which include defining public data and developing access to it using the Open Government License principles of free and freely available for re-supply and re-use including commercial use. For reference, Annex A to this paper includes Annex 2 of Making Open Data Real, which lists Public Data Principles, and Annex B includes the Open Government License.

It is likely that the emerging Open Data agenda, including the outcomes of these consultations and the establishment of a Public Data Corporation later this year, will further drive demand for a TNDS as well as put obligations on Traveline Stakeholders to open access to the data that they gather in order to run the regional and national Traveline services to the public.

On July 7th the Prime Minister announced a series of new data releases for the coming year that included rail data on a weekly basis, roads data on the Highways Agency network, roadworks, cycling and car park data from Transport Direct and more rail performance data from the Office of Rail Regulation.

3. Progress with Developing the Collation and Distribution of a TNDS

Recognising the increasing demands for a national supply of data, the Traveline Review Group has been working with Traveline regions, stakeholders and suppliers to establish:

1. How a TNDS might be achieved technically.
2. What the business case for creating such a dataset is, including how it will be managed and funded.
3. What the terms might be for its use and how it will be distributed.

Progress against item 1) is that a Beta TNDS now exists which is processed each week to include new data from the Traveline regions. Not all regions are able to update their data weekly as required due to financial or strategic barriers; these are included in Annex C. It is the intention of this paper and the resulting discussions at TAG and NTB to remove these barriers, which include the cost to the regions of making weekly updates to TNDS, and the need for them to be convinced that there is a business case.

Several third party organisations are regularly downloading the Beta TNDS for test in their systems including mySociety, ITO World, DfT and the Traveline journey planner suppliers.

Traveline has produced a Code of Practice (CoP) which is now in use with third party organisations that use the NextBuses API. A draft Traveline License Agreement was produced that the CoP could sit within that might be used to more tightly define the terms of use of the TNDS. Both these documents were produced in pursuit of a solution for item 3). The Traveline regions agreed to the use of regional datasets in the TNDS in principle and all have approved third party access to the dataset for test purposes.

Agreement by the Traveline regions for the distribution of the TNDS to third parties in a live environment against an agreed set of license conditions has not been sought formally. The discussions with the regions pre-date the evolution the Open Data agenda, which arguably should have a positive influence on the supply of data to the TNDS and negate the need for complex and varied license agreements.

Having established proof of concept and the CoP, and in order to complete the business case for a TNDS, Traveline went out to tender for a TNDS Collation and Distribution Service that would move the dataset to a robust and scalable service. In July 2011, Basemap was selected by Traveline Information Limited (TIL) as the preferred bidder for a five year TNDS contract and its offer has provided the costs and timescales required for the TNDS business case; these are provided at section 4 below.

4. Costs and Timescales for Delivering the TNDS

The Basemap TNDS core offer includes:

- As frequent as daily updates from each Traveline region or stakeholder
- Daily updates to TNDS
- Data validation and integrity testing
- Import and test reporting to Traveline Stakeholders
- ATCO.cif to TXC 2.1 conversion tool
- GTDF export
- Website for TNDS distribution

The following Table A below shows the costs for years 0-5 for delivery of the above with year 0 costs being development; the contingency costs have been added by TIL for budgeting purposes:

	Y0	Y1	Y2	Y3	Y4	Y5
TNDS Contract	£69,720	£13,401	£13,894	£14,386	£14,986	£15,570
Contingency	£13,944	£2,680	£2,779	£2,877	£2,997	£3,114
TOTAL	£83,664	£16,081	£16,673	£17,263	£17,983	£18,684

The data management role, including liaising with and supporting data suppliers and TNDS customers, will be provided by TIL using existing resource; Peter Stoner is confident that these tasks will fit within the Regional Data Coordinator role both in terms of time and scope.

All Traveline regions currently submit their regional datasets in ATCO.cif to ITO World for data quality testing and use in Traveline approved application such as posters at rail stations. The ATCO.cif to TXC 2.1 conversion tool in the Basemap offer could be used for those regions that have technical or financial barriers to submitting data in TXC 2.1 format on a weekly basis to TNDS. This would effectively remove those barriers as TNDS would simply use the existing regional data export to TIO World, although a TXC 2.1 export direct from the region might be preferred.

The timescales for delivery of the TNDS contract are three months from the start of implementation. It is entirely feasible that the TNDS service could be launched in January 2012.

5. Funding the TNDS

TIL does not have the all funding for the TNDS contract, and the level of funding in its 2011 and 2012 budget will depend on NTB decisions about NextBuses contract funding and other budget items at its meeting on 7th September 2011. The only significant income that TIL has is in the form of revenue from the call plan contract for the 0871 200 22 33.

In order to provide a robust and sustainable service, the TNDS contract should not be dependent on the availability of ongoing funding from existing TIL income, which is diminishing year on year as call volumes decrease.

The proposal is that TIL should seek investment for the Year 0 development costs and the contingency, initially from Traveline stakeholders.

TIL has the funds to cover the annual costs for years 1 – 5 but might wish to consider how it might cover these costs independent of its existing revenue.

Some options for funding TNDS costs might be:

- An annual subscription fee for access to the TNDS. Not all users might be required to pay the fee, for example academic institutions and those wishing to test the data in their applications before formalising access may be exempt. As an example a subscription fee of £300 for 45 users, which is the current number of licensed NextBuses API users, would generate £13,500 of annual income.
- Free access to the basic TNDS service which might be the complete TNDS on a weekly basis, but monthly support charges for premium services such as tailor made subsets of the TNDS or a daily update. Note that the TNDS will be updated each day as Traveline regions update their regional datasets on different days of the week, and some update more than once a week.
- A combination of the above

The formation of a PDC and responses to the government's Making Open Data Real consultation could mean that TIL may adjust any charges made upwards or downwards. But an interim decision

on funding, pending the outcomes of those consultations, should be made by TAG and NTB to prevent a delay in the implementation of a TNDS.

6. Licensing of TNDS

The proposal is that the TNDS is made available using the principles of the OGL but with at cost or marginal cost charges for premium access in order to cover the costs to TIL of the TNDS collation and distribution where TAG and NTB take the decision to proceed in that direction

7. Recommendations

The recommendations are that TAG/NTB:

- Agree to award the TNDS contract to Basemap subject to contract
- Aims to launch the TNDS service in January 2012
- Seeks investment, initially from Traveline Stakeholders, for development costs
- Agrees a position on licensing and funding annual TNDS costs

The Traveline Review Group will continue to manage the TNDS implementation and liaise with Traveline regions and stakeholders as appropriate throughout.