

## Open Data Provision: Position Paper

### 1 Introduction

- 1.1 The past decade or so has seen some dramatic developments not only in technology, but how it is used. The creation and use of data-based services is much more pervasive, since the rise of social networking and greatly simplified opportunities (commercial and non-commercial) for “app” developers. The opportunities have not been lost on Government and there has been a significant shift in policy towards “making public data public”.
- 1.2 Within public transport, much information is captured by the public sector for operational purposes. In the current policy environment, the question arises as to how, what, where and why this should be released. However there are some significant questions on how, what, where and why data should be released.
- 1.3 This paper has been developed as a result of work carried out by the RTiG Open Data Working Group, with involvement from central Government, local authorities, operators and the software industry. It outlines the issues involved in making data openly available, summarises RTiG’s current position and suggests ways that members can make sure that their data release strategies are both robust and effective.

### 2 Current situation

#### *Policy context*

- 2.1 Government policy has been guided by the influential report “The Power of Information Review”, co-authored by Tom Steinberg and Ed Mayo in June 2007 under the auspices of the Prime Minister’s Strategy Unit. The vision created in this has been sustained following the change of Government in 2010 and the Cabinet Office has now set a policy and mechanisms to achieve it:

Freeing up public data and putting them in people’s hands can help them have more of a say in the reform of public services. Making more public data available for re-use will also realise social and economic benefits by enabling the growth of new, innovative information-based businesses. This will help make the UK an attractive location for the global service and software business sectors.<sup>1</sup>

Making public data public takes forward the Power of Information Taskforce’s report recommendations. As part of this initiative Cabinet Office is committed to making available data that is in a form that is easily found; openly accessible; freely available for reuse by all for any purpose, and easy to licence.<sup>2</sup>

---

<sup>1</sup> Taken from <http://www.cabinetoffice.gov.uk/content/transparency-overview>.

<sup>2</sup> Taken from <http://www.cabinetoffice.gov.uk/content/making-public-data-public>.

### ***Data release activity***

- 2.2 One of the first major moves into publication of public sector data was made late in 2009 with a commitment by the then-Prime Minister to publish (a significant amount of) Ordnance Survey data. This was followed through over the next few months and now a large amount of OS data is publicly downloadable.
- 2.3 However the drive quickly became much wider, and the data.gov.uk initiative now collates the publication of many data sources. In the public transport sector, this includes the full NaPTAN database as well as, more recently, the National Public Transport Data Repository (NPTDR) – a week’s “snapshot” of full route and timetable data.
- 2.4 Separate publication initiatives have been undertaken at local level, notably within TfL, which has published some of its public transport data through the London Datastore (data.london.gov.uk). As well as stop location, route and timetable information, this includes some real-time data – namely live rail departure boards.
- 2.5 “Open data” is a changing and dynamic landscape, and is likely to be so for some time to come as new data are published (and potentially other data withdrawn). While the overall policy looks set to be sustained, in many contexts the practicalities of achieving open publication are taking a considerable amount of thought, both technically and commercially.

### ***Response of the transport sector***

- 2.6 While transport is just one of many sectors affected by this policy, it is more complex than many in that a significant fraction of transport is delivered in the UK through the activities of commercial, rather than public sector, organisations. Achieving this smoothly and coherently in a substantially de-regulated environment has, over the past three decades, required some careful relationship building. Because of this, the open data agenda has led to considerable discomfort in the transport community, particularly among operators but also among a number of local transport authorities and system/service providers.
- 2.7 Data like OS mapping is clearly collected and owned by the relevant public sector body, but the source of much public transport data is in commercial operations. While clearly operators will wish to make some information public, they feel that data being released without their explicit approval represents a loss of control and therefore, potentially, may be damaging to shareholder value. Some datasets (eg those linked to pricing and revenue) are likely to be more commercial sensitive than others (eg basic timetables). There is variation between operators; for instance, some are relatively relaxed about publishing real-time information, while others regard this too as highly sensitive.
- 2.8 There is a very different perspective in some parts of the software community. Charitable organisations like mySociety, as well as commercial organisations, are vocal about the “need” to release data held by the public sector, to encourage/facilitate the development of innovative services. For these people, it is irrelevant whether the data originated in (and remain owned by) private organisations. This gives commercial operators major concerns.

### 3 RTIG's position

#### *Principles of releasing data*

- 3.1 The open data agenda opens new challenges and poses real conundrums for stakeholders in both public and private sectors. Entrenched interests will inevitably be apparent for some time to come.
- 3.2 Fortunately, most local authority officers work quietly and patiently with local commercial operators to maximise mutual benefit. For at least five years in many areas (and a lot longer in some), this has included delicate discussions about detailed operator data being made available to local authorities, on the understanding that it was used with appropriate care and respect. Typically, this is codified and governed by a formal data sharing agreement; RTIG endorses this approach, since – although it can be challenging to agree – it provides a clear agreed basis for operations (and indeed a basis for negotiated change).
- 3.3 There is a risk that, if overplayed, the high-profile policy direction might lead to a stand-off (particularly in the current economic environment): LAs insisting on making data free to public access, and operators retaliating by stopping the flow of data to LAs. However at local level there is a high degree of pragmatism, and we do not see this as likely.
- 3.4 “Making public data public” covers a wide range of very different possibilities, with very different cost and risk profiles. It is unrealistic to expect public authorities simply to provide an open-access port to their corporate data networks. In practice, therefore, the provision of public data feeds will be a step at a time.
- 3.5 RTIG strongly supports this pragmatic approach. Rather than a blanket “presumption of release”, or indeed a blanket “presumption of withholding”, it is essential for public and private sectors to work locally and nationally for mutually acceptable solutions. That means both sides’ views of the likely cost and risk can be legitimately taken into account in setting strategy.
- 3.6 Within this context:
- LAs will want to share experiences among themselves, and operators will be watching what has been found to work. Cross-sector groups like RTIG, and single-sector bodies like ATCO and CPT, will be important to lubricate this process as it unfolds.
  - LAs have much clearer powers over tendered services. (This includes all TfL-controlled services.) If an LA determines that it wishes to make tendered service data public, it can do so unilaterally. However, this needs to be done on an equitable basis: it should be a clear part of the tender contract from the outset.
  - LAs should have an open door to the potential users of public transport data. This will give them a clearer understanding of how much potential use a data supply would get (and avoid releasing data that no one wants). It also provides a channel through which the LA can explain why particular data is *not* being made available.
  - Operators also have an opportunity to consider what of their private data they could beneficially release, either directly or using LAs as portals/aggregators. This is likely to vary both between operators and over time.

### ***Scope and "rules"***

- 3.7 Prior to releasing data, a stakeholder will need to identify clearly what is being considered:
- What information/data is covered.
  - Whether it is to be provided as base data, as a query service, as an API, or through some other mechanism.
  - Whether any restrictions will be applied on who can access the information – for instance, whether users are required to register before access is granted, or whether there are limits on the amount of data that can be accessed.
  - Whether any restrictions will be applied on what users can do with the data they receive, including restrictions on who the recipient may pass the data (and any processed data derived from them) on to; if so, how these restrictions can be made “fair and equitable” and how they will be enforced.
  - Whether any restrictions will be applied on the use of data over time – for instance, ensuring time-dependent data are not used after they have become outdated (eg year-old timetables or hour-old real-time information).
  - Whether any charge will be levied for use of the data; if so, how the charge will be determined and levied.
- 3.8 There are a number of licensing models that can help this discussion. Principles of data exchange were identified in the transport context as far back as 2000 for projects such as the Highways Agency’s Traffic Control Centre and, a little later, Transport Direct. Among current models the following are the most relevant:
- The Cabinet Office cites the “Open Government Licence for Public Sector Information” for “most” data released under on [www.data.gov.uk](http://www.data.gov.uk).<sup>3</sup>
  - The London Datastore also has various licensing models, but the one applicable to transport is the TfL model.<sup>4</sup>
  - There is a Code of Practice developed by the Public Transport Information Stakeholder Group, an *ad hoc* collaboration between Traveline, Transport Direct and others.<sup>5</sup>
  - The model Data Sharing Agreement provided as an annex to the Guidance for Bus Punctuality Partnerships, developed collaboratively by DfT, CPT, ATCO, LGA, CSS (now ADEPT), Passenger Focus, VOSA, and the Senior Traffic Commissioner.<sup>6</sup>

---

<sup>3</sup> Available at <http://www.nationalarchives.gov.uk/doc/open-government-licence>. This is what Cabinet Office asks public authorities to use for public data.

<sup>4</sup> Available at <http://www.tfl.gov.uk/termsandconditions/11402.aspx>.

<sup>5</sup> Available at [www.pti.org.uk/docs/tnds/Code\\_of\\_Practice.doc](http://www.pti.org.uk/docs/tnds/Code_of_Practice.doc).

<sup>6</sup> Available at <http://www.dft.gov.uk/pgr/regional/buses/buspunctuality>.

- The terms and conditions developed and published by ATOC for national rail data.<sup>7</sup>

3.9 RTIG recommends that any stakeholder – public or private – which is publishing data openly should also publish associated terms and conditions of use, based on one or more of these core models<sup>8</sup>.

### ***Delivery mechanics***

3.10 Once the decision has been taken to publish data, there are a number of practical issues that need to be considered:

- There needs to be a clear understanding of the data volumetrics, timeliness, complexity etc, and of the requirements this will impose on both LA systems and those of upstream data providers.
- Third party users that make poor use of information (for instance, providing time-relevant services based on out-of-date data) are a potential risk, to both public and private sector data owners. The publishing organisation needs to determine whether to control this kind of problem or ignore it. This is particularly important if a third-party service is advertised as “based on” the data (or similar).
- The publishing organisation needs to establish and sustain the technical and organisational resources to support the data feed service, in a way that data consumers understand. This includes mechanisms to review the quality of service and any problems that arise.
- The publishing organisation needs to be clear what liability it is taking on for any use made of these data. In particular this includes whether the user (where providing a third-party service) needs to acknowledge the data source, or even submit the service to scrutiny/approval by the publishing organisation.
- The publishing organisation needs to consider whether it wishes to publicise any or all third party services which are based on its data. Publicising may yield benefit, but may also become a liability by implicitly underwriting the services. Conversely the “approval” aspect provides an opportunity for the publishing organisation to encourage high quality services at minimal cost.
- The costs of the service need to be established, and ideally budget support gained for it to be sustained over the long term. Although “value for money” may not be essential if this becomes a political imperative, the publishing organisation would normally need to be able to measure the benefits to justify its expenditure.

---

<sup>7</sup> Available at [http://www.atoc.org/clientfiles/File/NRE/NRE\\_COP.pdf](http://www.atoc.org/clientfiles/File/NRE/NRE_COP.pdf) for NRE data, and at <http://www.atoc.org/about-atoc/rail-settlement-plan/data-feeds/conditions> for RSP data.

<sup>8</sup> These models vary in scope, depth, specificity and rigour, and are unlikely all to be relevant or acceptable in any given situation. Data publishers need to consult their own data partners and potential consumers before adopting a specific model.

### ***Private data owners***

- 3.11 Private data owners (operators or information service providers) are not covered by Government policy directly. However they may become affected if they:
- Use, in their business, public data which carry obligations (for instance, NaPTAN data usage which obliges users to notify and correct errors);
  - Are operating under contract to a public body with “flowed down” conditions (for instance, tendered services which oblige the provision of constraint-free operational data);
  - Receive fiscal support associated with specific publication obligations, for instance through the Bus Service Operators Grant “uplift” for AVL/smartcard-equipped vehicles.
- 3.12 The licence conditions may vary among the various obligatory publications. It is very likely – and the organisation should seek – that the variant conditions should be based, as far as possible, on one of the generally accepted frameworks identified in paragraph 3.8, which have benefited from extensive and wide-ranging scrutiny (though see caveat at footnote 8).
- 3.13 Beyond the minimum to which these aspects oblige them, data publication is of course a private decision. However, it may be convenient to use licence conditions which are similar to those used for the obligatory releases.