

## **Additional written evidence from the Open Data User Group (ODUG) [OD 29]**

### *Introduction*

The Open Data User Group (ODUG) welcomes this opportunity to provide additional evidence to the inquiry on statistics and open data and to clarify certain points in the light of the oral evidence given to-date, in particular that provided by Ministers from BIS and the Cabinet Office and the National Statistician.

ODUG is impressed by the breadth and depth of the inquiry. Informed questioning from the committee is bringing to light a range of contradictions in government policy which need to be resolved if open data is to contribute fully to the accountability, choice and growth agendas as laid out in the Open Data White Paper<sup>1</sup>.

In this additional evidence we will outline a number of principles that have been raised by and with the committee. We will also provide short case studies that illustrate these principles and other relevant documents as appendices to this evidence.

### *Issues raised as a result of evidence given to the committee*

#### **1. National Information Infrastructure (NII) and Core Reference Data**

A number of witnesses referred to the National Information Infrastructure (NII) as proposed in the Shakespeare Review<sup>2</sup>. ODUG welcomes the strong encouragement for government departments to choose their “most important” datasets and to ensure that those are made available under the Open Government Licence (OGL) as open data.

However, it is our view that the currently selected department data sets will not necessarily provide a coherent “infrastructure”. Also, whilst the number of data sets released by particular departments may be a useful indicator of that department’s commitment to the open data agenda, data sets that have been held back, for reasons other than the protection of personal data, may be a better indicator of a lack of commitment, for whatever reason, to open data.

ODUG also believes that not all data sets are of equal importance. In our view the term “*core reference data*” should be reserved for data sets which are the primary source of an identifier such as a VAT number, a Company Number, a Unique Property Reference Number (UPRN), a Postcode or a code used to identify a statistical or administrative area.

Such core reference data is essential to allow other data sets to be linked and to check the quality of data. This point was well made by Jil Matheson, the National Statistician, in relation to the National Address Register. This is a particularly important core reference data set which we discuss, as a case study, in Appendix 1. Its special importance was also noted in the Open Data White Paper<sup>1</sup> which included a commitment to resolve the issue of national address data.

The Office for National Statistics has a very good record of releasing core reference data relating to statistical or administrative areas as open data. These include the ONS Standard Names And Codes (SNAC) database and the ONS Postcode Directory (see Appendix 3) which has only relatively recently

---

<sup>1</sup> <https://www.gov.uk/government/publications/open-data-white-paper-unleashing-the-potential>

<sup>2</sup> <https://www.gov.uk/government/publications/shakespeare-review-of-public-sector-information>

become available as Open Data because it was previously kept closed under the commercial GridLink agreement between ONS, Royal Mail and Ordnance Survey (OS).

The ONS Postcode directory, while very useful, only provides Postcodes to administrative and statistical area conversions and does not list individual addresses (postcodes on average refer to about 14 addresses).

Where core reference data is not made available this may prevent the release of other data sets as open data in a usable form or the ability to derive the maximum benefit from datasets which are made open. Many of the requests for open data on data.gov.uk are for addressed lists of particular types of facilities, such as schools or libraries, or for data about properties such as Energy Performance Certificate Data. An Open National Address Register would ensure that all such data sets could be referenced using a Unique Property Reference Number from which a correct administrative or postal address could be established. Without access to an Open National Address Register either such data isn't published, or it is published with non-standard addresses making it difficult to match to and combine with other data sets.

ODUG notes that the term National Information Infrastructure might imply that a new "infrastructure" needs to be built. We believe that much of the infrastructure and much of the data that could and should be released as open data already exists. We concur with APPSI (the Advisory Panel on Public Sector Information), who raised many of the issues associated with making the National Information Infrastructure work in their October 2012 Discussion Paper on a National Information Framework<sup>3</sup>.

ODUG believes that the data.gov.uk data request forms submitted by the public provide an equal, or more valid measure of what should constitute either "core data", or "core reference data" in a National Information Infrastructure, rather than the data sets offered by departments. We also question the criteria by which departments have been asked to identify their data inventories for the NII, and suspect that there are many useful datasets which are not, as yet, part of the NII.

However, given that departments are now strongly encouraged to provide an inventory of unpublished data sets we believe that immediate effort is best expended on releasing that data which has been requested, or identified in the NII, with a focus on the data requested as this can be expected to be put to immediate beneficial use, rather than trying to release all data particularly where data has to be documented or refined for release. ODUG can, and does, assist in the assessment and interpretation of open data requests and the prioritisation of further datasets identified in the NII.

## **2. Open policy making and data release obstruction**

Despite a government commitment to open policy making ODUG has noticed a marked disparity between the openness of the decisions made to open datasets with the requirement to make business cases for releasing data and the closed, often unaccountable, process associated with decisions not to release data, or only to make data available under a commercial license.

Several examples characterise the problem. One was the decision to privatise the Royal Mail Postcode Address File (PAF) (see Appendix 2). The case for opening PAF was made publicly, while the advice that led to the decision to sell PAF with the Royal Mail was largely opaque or incomplete. For

---

<sup>3</sup> <http://www.nationalarchives.gov.uk/documents/nif-and-open-data.pdf>

example Ofcom's 2013 consultation on PAF was redacted to the point of being quite useless in assessing whether the price charged for PAF is 'reasonable', as required by the Postal Services Acts (an example of the redaction in the Ofcom consultation is provided in Appendix 6).

Equally, despite substantial lobbying from many sources, over many years, to establish an Open National Address Register the Government's decision, in 2010 to set up GeoPlace LLP a limited liability partnership between Ordnance Survey and the Local Government Association (LGA) was presented as a simple unjustified announcement. GeoPlace's corporate structure is one chosen by many Private Equity companies because it allows the financial dealings of the company to be private. The annual statement of GeoPlace LLP<sup>4</sup> contains no financial information other than the three main sources of revenue and the anticipation that GeoPlace will produce a surplus which will be returned to OS and the LGA.

A final example of closed policy making, which is still under consideration, concerns the obstruction of Her Majesty's Land Registry open data programme. HMLR are obliged, under the European INSPIRE legislation, as transposed into UK law, to make a cadastral (property ownership parcel) dataset publicly available. HMLR are inclined, in line with their exemplary open data policies on house price data, to make the outlines of the land parcels openly available for general re-use, and have released this data as open data. However the open use of the data is being impeded by the Ordnance Survey who claim that the land parcels are OS 'derived data' and any commercial re-use will involve a data royalty paid directly to the Ordnance Survey, despite the fact that OS neither owns, maintains or delivers this data (Appendix 4).

The rationale for the Ordnance Survey's stance and the level of royalty payments has been extraordinarily hard to ascertain and, as a result, the HMLR data is still not available for open re-use. A publicly owned company KentGIS, which should have access to the INSPIRE cadastral data under the Public Sector Mapping Agreement (PSMA), funded by the government, recently created a useful display service for the data which allowed the public to view Land Parcels nationally, for free. However this service has recently been closed down following a protest from Ordnance Survey. So, despite the best efforts of HMLR, the INSPIRE data is not open for either public or private sector use and re-use.

There are many more examples where a government department or agency decides not to make publicly funded data openly available for viewing and re-use. ODUG, on behalf of data consumers, is working to make the case that certain individual datasets of high potential value to the wider economy should be made open. However, despite the presumption that public data should be made open, as the committee has noted, there is no-one in government, outside the Cabinet Office, to ensure that open data policies are enforced and little consistency across government around the decision-making process for opening up data.

### **3. Sustainable funding, cost recovery and the generation of surpluses**

Normally government open data is public sector information which is necessary in order to perform a specific public task. Such data has been fully funded by the budget for that task and can be released as open data at a marginal cost which in an Internet environment is likely to be close to zero. The cost of the release of such data can usually be absorbed into the public task budget that led to the data being created and maintained.

---

<sup>4</sup> <http://www.iahub.net/docs/1377773673189.pdf>

More problematic is data which is not associated with any one specific public task but is used widely across government. Such data includes statistical information and geospatial information. Following the Rayner review which started in 1979 and reported in 1982 an internal market was created for statistical and geospatial information.

Government departments were expected to pay for such information and to act as intelligent customers. In the case of statistical information, the job of re-selling the information, in particular Census information, was given to private, for-profit census agencies, which paid an annual licence fee for the privilege of selling census data. This led to a situation where some government agencies found Census data too expensive to use. Ordnance Survey also moved, more successfully, into cost recovery trading and successfully recovered a larger proportion of its costs than the statistical service.

Around 2000 a review of Crown Copyright data was carried out which found that the statistical service was recovering a very low proportion of its costs, while government departments were being deprived of access to necessary statistical information. Statistical information then came out of the cost recovery / tradable information regime and effectively became open data. Since that time the usage, the variety and the quality of statistical information has increased markedly.

At about the same time the Ordnance Survey was confirmed as a trading fund and was expected to move towards 100% cost recovery and now returns a significant surplus to the Treasury. However much of this cost recovery and surplus is illusory because Ordnance Survey earns a relatively low proportion of its revenue from the private sector. The bulk of its revenue comes from government payments for the Public Sector Mapping Agreement (PSMA) and for Ordnance Survey Open Data.

In the same way that some Government departments could not afford to buy statistical information when the cost recovery regime was in place, prior to the Public Sector Mapping Agreement, departments, including ONS, could not afford access to all the Ordnance Survey Data that they could, or should have used.

A significant amount of Ordnance Survey's supposedly private sector revenue is actually derived from quasi-private bodies such as not-for-profit Housing Associations, the utilities and from the sale of maps for planning applications for which there is little or no competition and which generate disproportionately large revenues for Ordnance Survey and their re-seller partners. All of these supposed trading revenues are high surplus monopoly rents, which are effectively a tax on data use rather than normal competitive market profits. It is important also, to put the surplus OS returns to the treasury into the context of the very substantial market for geospatial information in which it operates which is estimated to be worth £bns a year. Despite their monopoly position OS captures very little value from the overall market. ODUG believes that this is a growth opportunity for the UK and, were OS to open up its data, there would be substantial opportunities for private companies in this sector. We believe that the OS should be funded by government to perform its important public task, as the National Mapping Agency, but that the commercialisation of this data, and the many derivative products and services it can drive is best left to the private sector.

An alternate regime for recovering the costs of currently charged for government information, mainly from the trading funds, could be based on charging those who cause core reference information to change, rather than those who seek to use it. This is the model that allows Land Registry and Companies House to release significant and increasing parts of their data holdings as open data.

#### 4. Societal benefit

Open data is justified in the Open Data White Paper because of the impact it has on Transparency, Efficiency, Accountability, Citizen Choice and Economic Growth. Yet much of the argument and most of the business cases demanded for open data have to be centred on evidenced economic growth. This is difficult to prove as projections are inevitably conjectural and it is very difficult, if not impossible, to identify the enterprises that will grow as a result of an open data release.

However a recent article in the Financial Pages of the New Yorker has been widely acclaimed as explaining why crude attempts to measure the impact of open data on GDP may be incorrect (Appendix 7 – Gross Domestic Freebie).

Witnesses and some members of the committee appeared to believe that an important reason for not releasing open data is that it gives large multinational corporations (such as Google and Microsoft for example) a free ride and that they should not be let off from paying a fair price for the information infrastructure that they use. However this ignores the fact that most of them are choosing not to licence or pay for government information at present. The effect of charging for data doesn't impact on the multinationals at all, as the amounts they are charged have no material impact on their bottom line. However, charging for the data prevents the societal benefits that would accrue to all citizens from easier free access to data as well as the Small and Medium businesses that would generate new products and services on the back of open data, unencumbered by complex and costly licensing regimes.

Charging for data, rather than bringing in addition revenue from large wealthy corporations, limits use of the data to medium sized companies. What is worse is that the revenue isn't being used to maintain the data itself, which is a relatively inexpensive process. It is being used to maintain the bloated sales, marketing, and administrative structures necessary to protect the intellectual property rights of the organisations that are keeping the data closed.

In his evidence to the inquiry the Minister of State for Business and Energy stated that he was seeking ways to differentiate the costs of charged for public data for small businesses. ODUG does not think it is possible to implement a differentiated pricing regime as suggested under the existing Information Fair Trader Scheme<sup>5</sup> and EU State Aid rules<sup>6</sup>. Equally the burden of putting in place such complex licensing regimes has already been noted as an inefficiency in itself and as significant burden on small businesses.

We believe that current levels of government expenditure on government produced public sector information, are already sufficient to release that information as open data. Charging regimes fund an army of data use preventers creating jobs inventing pricing and licensing schemes which are unnecessary. As an example we illustrate this effect for the Royal Mail Postcode Address File (PAF) in Appendix 2, though we are limited in the extent to which we can do so because of the secrecy that has surrounded the real cost of the PAF.

Societal benefit accrues when data is used. The New Yorker article reports massive, though unmeasured, societal benefits accrue to all of those consuming 'Free' services from corporations such as Google, Microsoft, Twitter, Yahoo and Apple. Of course each of these businesses has alternative sources of revenue which fund their 'free' data and 'free' services. However, as this business model now dominates and consumers/citizens are used to consuming and using data at no

---

<sup>5</sup> <http://www.nationalarchives.gov.uk/information-management/ifts.htm>

<sup>6</sup> <https://www.gov.uk/state-aid>

cost, but paying for the services in another way, it is counter-productive for a small number of government, or recently privatised, agencies to swim against this tide.

## **Conclusion**

A number of previous select committees, and government advisers such as Sir Michael Scholar then Chair of the Statistics Authorities (see letter in Appendix 5) who have consistently called for an open data mechanism to maximise the beneficial use of public sector data.

These requests proposals and comments have consistently been ignored in order to chase the illusory revenues, which have never been delivered to their full potential, from closed data regimes.

The UK's leading role in the Open Government Partnership and the rapid move towards open public data in Europe and the US, which in turn is leading to massively greater data re-use, with associated societal benefits and economic growth potential, should lead the committee to recommend a rethink in government over the remaining closed data regimes.

We would also welcome an endorsement by the committee for more open policy making in this area so that, at the very least, closed data decisions are as openly and transparently justified as is the case for open data. We would hope that, in the light of such transparency, open data would prevail more often.

*December 2013*