

TB(2)1.iii - Response to the Transparency Board's request to the public for additional data sets they would like released

Response to the Transparency Board's request to the public for additional data sets they would like released (July-September 2010)

Summary

- National Parks Members Register of Interests
- Raw Data from Important Experiments
- Civil Service Staff Engagement Survey
- Social Security Recipients at Lower Level Output Area
- National Insurance and PAYE Tables
- Register of Charities
- Postcode Address File
- Boundaries of LNRs, NNRs and SSSIs
- Fares, Timetables and Routing Information
- National Learning Directory Database
- House Prices Paid
- Pre 2009 Boundary Plan
- E-Gov Register
- ACOBA reports in a machine-read format
- All ACOBA data on senior crown servants
- Companies House data
- Inter-Departmental Business Register
- Councillors meetings on planning changes
- Lower layer super output area Isoa boundaries
- Ordnance Survey Mastermap
- List of UK local councils and courts
- Royal Mail Postal Address File
- ONS National Statistics Postcode Directory
- Land Registry records
- Ordnance Survey doorstep grid references
- NHS organisation codes as a web service
- Better data coverage for 19-24 age groups
- Online publication of 'The Diplomatic Service List', 'The Civil Service List', 'The Navy List', 'The Army List' and 'The Air Force List'.
- Environmental data
- Up-to-date local level demographics and census information

TB(2)1.iii - Response to the Transparency Board's request to the public for additional data sets they would like released

- Referrals made by GP surgeries
- Monthly Local Housing Allowance data

National Parks Members Register of Interests

Public Sector Information Holder: **DEFRA**

The problem

National Parks currently have un-elected members on their authority boards. These members often wear a number of other 'hats' that the general public may not be aware of. These other interests could be seen as prejudicial to local decision-making. If these members were locally elected then the general public would KNOW who they were voting for and what other 'hats' these people wear.

My ideal solution

The ideal solution would be to make available to the public, in one place of access, a list of ALL other boards, authorities, councils etc., that National Park Authority Members have an interest in and what remuneration or expenses they draw from each 'seat'.

What I would do

Our group (One Voice) would make a link available from our website. We only inform and empower our members, it would be their decision as to how they may or may not use that information.

Raw Data from Important Experiments

Public Sector Information Holder: **Business Innovation and Skill (& Education?)**

The problem

With the rise of the internet and access to a huge range of often conflicting and unproven physics theories (especially particle physics), it is extremely difficult for individuals to test, prove or disprove these theories as they do not have access to good, raw data, to use to test any hypothesis. Many of these individuals have great ideas (and some mad ones!) but their collective brainpower is not being harnessed as they are outside the normal physics circles. Much of the UK's physics work, especially in particle physics is government funded.

My ideal solution

The provision of sample sets of high quality, well documented experimental data from key experiments. Examples could include: Young's slit experiments with single photons, high energy impacts that lead to the production of subatomic particles, raw atomic spectrum data, experimental quantum entanglement data. Not every experiment, just those seen as key to the corporate body of knowledge, and work that is centrally funded. This would enable those interested in Physics - students and those institutions who do not have access to key facilities - to recreate experiments virtually, and lead to an increased interest in the study of physics. There are a lot of armchair physicists out there who would jump at the chance to test their theories with real data.

TB(2)1.iii - Response to the Transparency Board's request to the public for additional data sets they would like released

<p>What I would do</p> <p>Personally I would focus my own research so that I could use some of the data to help my prove (or disprove) my 'alternative' theories. I would also be happy that physics had been opened up to huge numbers of people in a way that would encourage increased interest in physics and engineering.</p>
<p><u>Civil Service Staff Engagement Survey</u></p> <p>Public Sector Information Holder: Cabinet Office</p> <p>The problem</p> <p>Only partial data has been released from these data - regarding particular aspects of staff engagement. In order to interpret the data more completely the whole anonymised datasets are required please.</p> <p>My ideal solution</p> <p>The whole dataset (anonymised) - past and most recent.</p> <p>What I would do</p> <p>Run analyses and compare against other datasets for staff engagement and career development progress.</p>
<p><u>Social Security Recipients at Lower Level Output Area</u></p> <p>Public Sector Information Holder: DWP</p> <p>The problem</p> <p>This would help with targeting life chances team, home fire safety checks and much more. Merseyside currently has it as a pilot and we'd like it for the whole NW.</p> <p>My ideal solution</p> <p>Anonymous data produced for whole NW region in common GIS formats or even just in Excel similar to that received by Merseyside.</p> <p>What I would do</p> <p>See above.</p>
<p><u>National Insurance and PAYE Tables</u></p> <p>Public Sector Information Holder: HMRC</p> <p>The problem</p> <p>I recently requested the NI and PAYE tables from HMRC in a CSV format (as they only exist in PDF format on their website www.hmrc.gov.uk/payee/forms-publications.htm). I was informed by the SDS Team (SDSTeam@hmrc.gsi.gov.uk) that this information does not exist in CSV form.</p> <p>My ideal solution</p> <p>This tabular information should be available in formats that allow uploading to other systems. I am currently writing a Payroll system and these tables had to be manually created on our systems. When the data changes in the future (new NI/tax rules) then this process will have to be repeated.</p>

TB(2)1.iii - Response to the Transparency Board's request to the public for additional data sets they would like released

<p>What I would do This would vastly simplify the updating of our systems when NI & tax rates change. I am sure other companies would benefit.</p>
<p><u>Register of Charities</u> Public Sector Information Holder: Charities Commission The problem The Register of Charities is not downloadable, which essentially stops mashups/analysis relating to charities, e.g. matching up with spending by central/local govt My ideal solution A download of the register including all core data about the charity (i.e. address, purpose, trustees, relevant dates, annual report location, etc) What I would do Tie to local council spending data</p>
<p>Postcode Address File (PAF) Public Sector Information Holder: Royal Mail The problem The PAF has long been held as secure data which cannot be revealed. However this is data on house and homes owned by the public. How can this kind of information not be freely available to be re-distributed and shared? The information which America has openly shared for many years but it seem the equivalent is under lock and key for us in the UK. My ideal solution Part of the PAF should be shared for address lookup. If Royal mail fear income issues will suffer they should be investing in a range of other services which can support this date, i.e. API's and other services which will show significant benefit of choosing royal mail over other providers. What I would do There are thousands of applications that are possible with the PAF and by limiting data it limits ideas.</p>
<p><u>Boundaries of LNRs, NNRs and SSSIs</u> Public Sector Information Holder: Natural England The problem I have asked for this information through the normal processes and have discovered that despite open OS initiative this data is still unavailable for re-use and is subject to Natural England copyright limitations. Below is a summary of the history... one day the information tax payers pay for can be used by everyone. An organisation that cannot share its data will be seen as adding little value. Dear Ian,</p>

TB(2)1.iii - Response to the Transparency Board's request to the public for additional data sets they would like released

Thank you for your various emails clarifying your position on the question of your using data derived or interpreted from Natural England owned data on Google applications. Having now received feedback from Natural England's Intellectual Property manager and our Ordnance Survey main contact, I would like to clarify where Natural England sits with this.

Obviously, in an ideal world, Natural England owned data would be available to use freely on whatever internet mapping platform the customer requires. This would both promote Natural England's work to a wider audience, and be in keeping with the current trends to put publically funded data into the public domain.

However, due to ongoing issues with Ordnance Survey and Google, which I have outlined in previous emails (the clause in question is in Google's Terms of Service, 11.1, which gives Google more rights than OS are comfortable with) Natural England cannot publish OS derived data on externally available Google platforms, which is not ideal, but is the situation in which we are in.

Your suggestion to redigitise Natural England's data using an OS OpenData source or a Google-compatible source (in this case Infoterra's aerial photographs) are unacceptable to Natural England on the following grounds:

1. This breaches the Ordnance Survey End User Licence, which is supplied with the OS derived data within the Natural England Terms of Use document, specifically:

3.2 (d) The Supplied Data is not copied, adapted, varied or modified except to the extent and only to the extent to which any of those acts are expressly permitted by this Licence.

3.4 The End User shall take all reasonable steps to prevent unauthorised use and exploitation of Supplied Data by any person, including, without limitation, its employees.

2. This also breaches the Natural England Terms of Use supplied with all Natural England's data, specifically:

2 f. Any product derived from the Data must be acknowledged as Natural England copyright using the following statement:

"© Natural England [Year], reproduced with the permission of Natural England, www.naturalengland.org.uk/copyright." If the Data contains derived data additional third party statements may also be required (see below).

3 b. You may not copy or otherwise reproduce the Data without our written consent.

3. In addition, a point you have made yourself, once the data is in the public domain it will be seen by some as the de facto boundaries for the sites you show, and I have reservations about the merit of digitising this data either using a less detailed OS product or using the Infoterra aerial imagery on Google, as, from examination of the sample data you sent and from knowledge of the scale of the products involved, this will not be as accurate as Ordnance Survey MasterMap, nor will it be updated with the same regularity as our data, which could potentially lead to decisions made with erroneous information.

4. Further, using our data on Google constitutes commercial use as there is advertising contained within the Google interface, for which you would require a commercial licence from Natural England.

Our OS contact says:

"Having spoken with the Contracts Manager for Central Government I am advised that publishing redigitised boundaries (via OpenData) onto the free Google API still breeches our terms and conditions, as the original polygons were derived from PGA products."

TB(2)1.iii - Response to the Transparency Board's request to the public for additional data sets they would like released

As I said at the start of this email, in an ideal world Natural England owned data could be made available via publically available interfaces such as Google; however at the current moment implementing what you have proposed contravenes both Ordnance Survey licensing and the Natural England Terms of Use, and provides a dangerous precedent in promulgating inaccurate data. Therefore, and it is with regret, I have to inform you that Natural England cannot agree with your proposed use of NE derived data on Google. If you were to use one of the internet mapping applications which does not contravene OS's terms and conditions then this discussion would potentially have a different outcome.

Regards,

Geographic Services, Evidence Team, Natural England

My ideal solution

Boundary data that can be openly used on Google Earth

What I would do

Use it to help cities understand their sustainability performance by combining this data with social and economic information which is location dependent.

Fares, Timetables and Routing Information

Public Sector Information Holder: **ATOC / National Rail**

The problem

Fares, timetable and routing information can currently only be accessed by National Rail Enquiries and similar web-sites. This is acceptable for individual enquiries, but not for automated enquiries used to assist passengers with more complicated enquiries.

My ideal solution

Free access to the RSP data feeds as available here www.atoc.org/about-atoc/rail-settlement-plan/data-feeds/cost

What I would do

Develop a website to inform passengers about fares, alternative railway routes and allow more flexible journey planning eg:

- advice for obtaining better value train fares such as by using slower, longer routes or trains
- using easier interchanges for those with mobility difficulties
- knowledge of the possibly of getting an earlier train at an interchange

National Learning Directory Database

Public Sector Information Holder: **direct.gov.uk, National Learning Directory**

The problem

The National Learning Directory is "the government's central learning database. It contains approximately one million courses from more than 10,000 providers in England, Scotland, Northern Ireland and Wales"

Whilst organizations such as UCAS have course related data they aren't providing that data to anyone other than Higher Education Institutions to innovated against.

TB(2)1.iii - Response to the Transparency Board's request to the public for additional data sets they would like released

My ideal solution

Ideally I'd love it if this data was published as linked data. I doubt very much if this is available in this format at the moment. However, I have lots of experience, modelling, converting and publishing data as linked data. So if the raw data was made available in csv, xml, or any structured format. I would be able to work with it.

What I would do

This is a huge amount of data about courses being offered in the UK. Unlocking this data would allow people like myself to build new search tools and visualizations and mashups that might better help prospective students find the right course.

House Prices Paid

Public Sector Information Holder: **Land Registry**

The problem

Currently to get this data you have to engage in a business agreement with the HMLR. Since it's possible to search on this data through the HMLR site and through third parties like Rightmove, why not make the raw data available for free too.

My ideal solution

The dataset made available free of charge.

What I would do

Website visualisations of the data.

Pre 2009 Boundary Data

Public Sector Information Holder: **Ordnance Survey or ONS or UKBORDERS**

The problem

We are trying to get hold of pre-2009 electoral ward boundaries as a shape file. Ordnance survey say they can only supply the 2010 BoundaryLine dataset. The data appears to existing under the control of UKBORDERS but it is only available to academics. This is daft.

My ideal solution

Get OS to supply historical datasets as part of Opendata

What I would do

Use it to analyse trends in homeworking and home based business region by region and identify the best locations to provide support services like "workhubs".

E-Gov Register

Public Sector Information Holder: **Socitm/Brent**

The problem

Brent operated the e-Government Register (eGR) for many years and this was openly available to all from their website. The eGR showed which councils purchased which types of systems from which supplier.

TB(2)1.iii - Response to the Transparency Board's request to the public for additional data sets they would like released

The decision was made by Brent and Socitm for this data to be merged with Socitm's own register, and thus would disappear from public view, and from the view of those public organisations, including councils, who do not pay to join Socitm. It has also of course disappeared from view for all those SMEs who could have been interested in finding interesting markets for the many niche markets that exists around and in between the major software houses' applications. More worryingly, the disappearance of this public data a few weeks ago means that the public, or those working on the behalf of the public such as data activists and journalists, can no longer de-reference what payments to suppliers were paying or contributing towards.

There is a fuller history and interesting discussion in the comments on here:

<http://paulgeraghty.posterous.com/closed-data-now-socitm-does-a-times/>

My ideal solution

That the responsibility for maintaining this data be handed back to the DCLG, the collection and management should probably to be incorporated into the esd-toolkit, and powers similar to that regarding LocalGovDirect Redirect service should apply.

That either councils should publish their own open data about which systems they use, or log into the esd-toolkit and update their records just as they have to update their LGSL URLs for the redirect service.

What I would do

We will be able to de-reference payments over £500 by councils, in as much as we will know WHICH councils paid WHICH suppliers HOW MUCH money on WHAT date and if this data was open we could logically and sensibly then be able to see WHAT it paid for.

This will lead to data visualisations of the costs of software and services, leading hopefully to comparisons and more judicious buying decisions.

ACOPA should publish data from their reports in a easy to machine-read format. Currently they publish final reports in PDF only: <http://acoba.independent.gov.uk/publications.aspx> - CSV or Excel would be useful. It would also be useful if they didn't delete the html version of their data (http://acoba.independent.gov.uk/former_ministers_appointments.aspx) when they publish the PDF report.

ACOPA should make available all data on senior crown servants appointment applications, not just the summary tables they provide in their report.

At a minimum we should have by department and year, the number of appointment applications by new employer name. Based on the database schema they revealed in an FOI response last year, the SQL queries to produce this summary are:

```
SELECT DepartmentName, YEAR(DateTakenUp), NewEmployerName, COUNT(*)  
FROM CSAPPLICATION  
  INNER JOIN (CSAPPLICANT  
    INNER JOIN GovernmentDepartment
```


TB(2)1.iii - Response to the Transparency Board's request to the public for additional data sets they would like released

```
ON CSAPPLICANT.DepartmentID=GovernmentDepartment.DepartmentID)
ON CSAPPLICATION.ApplicantID=CSAPPLICANT.ApplicantID
WHERE DateTakenUp IS NOT NULL
GROUP BY NewEmployerName, YEAR(DateTakenUp), DepartmentName
ORDER BY DepartmentName, YEAR(DateTakenUp), NewEmployerName;

SELECT DepartmentName, YEAR(DateTakenUp), NewEmployerName, COUNT(*)
FROM MApplication
INNER JOIN (MAPPLICANT
INNER JOIN GovernmentDepartment
ON MAPPLICANT.DepartmentID=GovernmentDepartment.DepartmentID)
ON MApplication.ApplicantID=CSAPPLICANT.ApplicantID
WHERE DateTakenUp IS NOT NULL
GROUP BY NewEmployerName, YEAR(DateTakenUp), DepartmentName
ORDER BY DepartmentName, YEAR(DateTakenUp), NewEmployerName;
```

Making the ACOBA data freely available is in line with the coalition's "government transparency" agenda. The ACOBA data is critical for the analysis of whether there is a revolving door problem between government and the private sector.

The data held by Companies House is a top priority dataset for me. Access to the Companies House XML gateway (<http://xmlgw.companieshouse.gov.uk/>) and all filings should be free of charge, in order to have data to:

- 1) Lookup director names for transparency reasons. At directors.mapofpower.com a snapshot of directors information has already been available.
- 2) Analyse corporate filings for transparency reasons. The USA made their SEC's EDGAR database of corporate disclosure filings freely available to the public in the 1990s.
- 3) Map from company name to company number, legal status, and SIC code(s). This allows government contract data to be mapped to the legal entity of a supplier in the cases where their registered company name is given.

The Company House registration and filing fees could be increased to cover any revenue the Government loses by making data available at no charge.

TB(2)1.iii - Response to the Transparency Board's request to the public for additional data sets they would like released

	<p>The Companies House dataset is a critical dataset you need to make available if we are expected to make sense of the central and local government spending information that the government is in the process of releasing.</p>
	<p>The Inter-Departmental Business Register is a top priority dataset for me, in order to have data to:</p> <ol style="list-style-type: none"> 1) Map from a VAT number to a company number, company name, SIC code(s), legal status, intrastat marker, and country of ownership. This allows government contract data to be mapped to the legal entity of a supplier in the cases where a VAT number is given. 2) Map from a child company to it's parent company via the register's enterprise groupings of legal entities under common ownership. Note this data is not available from Companies House, as they do not record parent and child company relationships. 3) Compare companies by employee size, country of ownership, SIC codes, legal status, and other fields. <p>Any information considered confidential, such as paye data, tax filings, and turnover figures, could be removed from the register data before public release. In order to release the data, it may be required to change the Code of Practice for Official Statistics, and specific legislation.</p> <p>The Transparency Board has identified the Inter-Departmental Business Register as a frequently requested dataset. However they have incorrectly called it the Integrated Business Register. Can you please use correct name in the future?</p> <p>The Inter-Departmental Business Register is a critical dataset you need to make available if we are expected to make sense of the central and local government spending information that the government is in the process of releasing.</p>
	<p>I think that the councillors meetings which discuss planning changes should be made public - case at present GBC want to change the use of three properties from Agricultural to Residential - simply at first but this could lead to its sale to a member of the public who could then conver to a hotel complex in middle of green belt - the discussion leading to the in house application for change of us are at present hidden from all local objectors as they are 'ancilliry' to the actual application</p>
	<p>lower_layer_super_output_area_lsoa_boundaries</p> <p>I just really wanted to say it seems a shame the LSOA data sets are not available for download. Seems odd having to order a CDROM.</p>
	<p>Ordnance Survey Mastermap(TM)</p>
	<p>I am looking for list of all of the local councils in the UK and also a list of all courts. For the local councils, I'm looking for something like</p>

TB(2)1.iii - Response to the Transparency Board's request to the public for additional data sets they would like released

you find here:
http://www.direct.gov.uk/en/DI1/Directories/Localcouncils/AToZOfLocalCouncils/DG_A-Z_LG
but the version I want isn't hard copy, I want to be able to link my database to the local council/courts database.
Royal Mail Postal Address File
ONS National Statistics Postcode Directory
Land Registry records
Ordnance Survey doorstep grid references
The NHS publishes all its organisation codes as a bunch of excel and access files. It would be great if it could be provided as a proper web service.
In order to identify our baseline for our local needs our community requires joined up anonymised data from all of the provider network across all sectors down to domicile and birth date level for each community. We need to be able to identify travel to learn patterns and challenge the quality of destination information held in databases to include HESA. Those dropping out of Learning or work need to be picked up early to provide the right support mechanisms. We need better data coverage for 19-24 age groups to allow returning NEET's issues to be addressed at each age level within that group. We need to know the cost of producing data and its metadata to ensure that we can select the right source to analyse and determine its value for money.
<p>Until a few years ago, those with an interest could easily access paper copies of annual publications such 'The Diplomatic Service List', 'The Civil Service List', 'The Navy List', 'The Army List' and 'The Air Force List' at public reference libraries. Some of these publications seem to have been discontinued and budgetary constraints on library services mean that those that are still published are often no longer available except in the largest reference libraries.</p> <p>It is believed that information of the scope, quality and quantity as published in the aforementioned publications and similar official publications circa 2000 should be made available online annually as a matter of routine (the possible availability of expensive CD-Roms from TSO not being an acceptable alternative).</p> <p>As an example of the continuing relevance of these publications (or the presentation of similar data in a different format), there is current interest in possible cuts in the senior grades of the Diplomatic Service and senior officers of the Armed Forces. Public debate on this issue should be informed by authoritative and detailed data concerning the baselines against which any cuts would be made.</p>
I am interested in the provision of environmental data and wondered if there are any plans to release a large scale database of environmental data as part of the data transparency steps being taking by the government.
<p>Much of the data on data.gov.uk is great and useful, however the geographic level at which it is available is not detailed enough for local comparison. For example, to attract funding to work with the most deprived neighbourhoods it is often necessary to establish how much more need there is in the neighbourhood, compared to that of the borough it sits within.</p> <p>This is largely not possible with the data provided. While demographics and census information can be put together at a local level</p>

TB(2)1.iii - Response to the Transparency Board's request to the public for additional data sets they would like released

	(now very much out of date), crime levels, educational attainment, benefit claim rates cannot be analysed at the same level. Is this something data.gov.uk could address?
	In response to datasets that should be released - I've been looking for data on the referrals made by GP surgeries, both in terms of prescriptions and onward treatments. This information would make GPs more transparent in their actions.
	Local Housing Allowance data is published for each 'Broad Rental Market Area' on a monthly basis, so we are after some kind of data feed that allows for these monthly updates. A dataset that linked Broad Rental Market Area to postcode might also be useful, as no doubt these areas will be revised at some point (it may be that it is in fact already possible to obtain this data).