

This is a pre-copy-editing, author-produced PDF of an article accepted for publication in International Journal of Public Law and Policy, 2, (3) 299-333. See further: <http://www.inderscience.com/jhome.php?jcode=ijplap>

## **Public sector information and open data: which way forward for the UK?**

Stephen Saxby  
The Law School  
Faculty of Business and Law  
Southampton University  
SO17 1BJ, UK  
Email: s.j.saxby@soton.ac.uk

**Abstract:** Since 2009, the move towards Open Data policies in the UK, also currently under review in proposals to replace PSI Directive 2003/98/EC on access and re-use of data, is having a profound impact on UK policy towards public sector information (PSI) that, as a resource, goes to the core of its function and purpose. Driven by principles of openness and transparency, the process now supports the systematic release into the public domain of PSI in the form of datasets. The Government believes that collective scrutiny of such data, while contributing to transparency, may also offer new insights into policy. Expectations have grown and new partnerships are emerging that blur traditional distinctions as to what is ‘public’ and ‘private’ in this regard. Government has also been listening to the ideas of the founder of the World Wide Web that much more can be secured from today’s Web via adoption of new techniques for linking data. The important contribution that location data, “information rich” in content, can make towards policy development has been recognised and acted upon. The EU has also observed the need for process in sharing spatial data in Europe.

© 2012 Steve Saxby. All rights reserved.

**Keywords:** public sector information; Open Data; Linked Data; INSPIRE Directive; PSI Directive; UK Location; Digital National Framework; Public Data Corporation; Ordnance Survey; making open data real; place matters; data.gov.uk

**Reference** to this paper should be made as follows: Saxby, S. (xxxx) ‘sector information and open data: which way forward for the UK?’ *Int J. Public Law*, Vol. X., No. Y, pp. 000-000

**Biographical notes:** Stephen Saxby is Professor of IT Law and Public Policy and REF Champion at The Law School, Faculty of Business and Law, Southampton University. He is Director of the Institute for Law and the Web (ILAWS) at Southampton and Conference Chair of [www.lspi.net](http://www.lspi.net). He is Editor in Chief of *The Computer Law and*

*Security Review – The International Journal of Technology Law and Practice* (Elsevier) and the *Encyclopedia of Information Technology Law* (Sweet and Maxwell).

## 1. Introduction

If there was an equivalent to Moore's Law<sup>1</sup> capable of measuring the complexity of PSI policy in 2012, then it would probably relate well to the rule implicit in Gordon Moore's prediction concerning chip design that it would double in transistor capacity every 12 months. That prediction is now heading for the 'silicon wall' and new techniques must be developed to enable it to retain its accuracy. But in terms of policy towards PSI, the present pace of change and the challenges this imposes is outpacing anything that has been seen before.

Up to now policy towards what can be produced and distributed has had to overcome cumbersome regulatory controls, technical shortcomings, administrative obstacles and proprietary restrictions. The result has been that the raw material of PSI – the data itself – has gradually broken loose from its moorings as a passive commodity under government control, into a freely available and potentially data rich resource capable of linking to other data. Further development of the policy is needed to enable that potential to be realised.

This paper explores these changes and examines the present state of play, the current policy initiatives and data standards requirements. It will assess what is happening, what more needs to be done and the challenges for PSI that exist in responding to the pace of change. Particular attention will focus on the emergence of location data and its growing role as a policy tool, made subject to specific regulatory objectives via the INSPIRE Directive that establishes an infrastructure for spatial information in Europe.

## 2. The genesis of data policy

Until the Millennium, PSI policy development sat very comfortably within a fairly straightforward debate about the terms upon which government would open up access to its vast store of public data. Access to and exploitation of PSI was under government control via Crown copyright, with licensing arrangements in place that would permit limited scope for the private sector in adding value to the data. Broadly speaking this involved its accretion with other data so as to enrich the data resource, combined with improvements in its presentation and accessibility. Government would obtain revenue from the use of the original dataset(s) and preside over the integrity of the process.

Over time, the mood for change has intensified, characterised by the *The Guardian newspaper's* 'free our data' campaign which argued that public data, already paid for by the tax payer, should be accessible to the latter without the public having to "pay for it a second time". With increasing regularity, however, the data did become available under more relaxed licensing

---

<sup>1</sup> This has become known as 'Moore's Law' following an observation made by the founder of Intel Corp. Gordon Moore in 1965, that the density of transistors on an integrated circuit chip would double every 12 months. That has largely proved correct although the time span has slipped over the years to something nearer 18 months.

arrangements. The exception lay with core datasets e.g. Ordnance Survey's *MasterMap* series, that required on-going investment and for which access would need to be paid for to recoup the development and maintenance costs involved. The process was administered, first by *Her Majesty's Stationery Office* and subsequently the *Office of Public Sector Information* (OPSI), until the entire task was subsumed within *The National Archive*. By this time government had begun to better understand the links between policy towards PSI and what could be achieved via modernisation and reform of government services. This was particularly so in relation to information use and re-use and led to a number of policy reviews and subsequent reaction to reflect the value and evolving nature of information as a resource.<sup>2</sup>

The private sector, however, wanted more uninhibited access to PSI. This would enhance business activity and profits, either directly through the creation of information products from the data, or simply as a result of the private sector better informing itself about a range of business advice available from PSI sources. The initial story then was one of a gradual process whereby government began to act upon these principles. For the reason just given, HM Treasury, however, needed to be convinced as to the cost benefits of such policy changes applicable to *Trading Funds*, such as Ordnance Survey,<sup>3</sup> and the Meteorological Office, that funded much of their work from the licensing of products and services that were within their remit to produce and maintain.

By this time, however, in addition to pressures on PSI policy associated with efficiency and cost, was growing enthusiasm for more openness and transparency. Limited steps to open up government and government decisions to scrutiny had taken place via legislation such as the *Human Rights Act 1998* (c.42) and *Freedom of Information Act 2000* (c.36).<sup>4</sup> To this was added *Directive 2003/98/EC on re-use of public sector information* (PSI Directive)<sup>5</sup> which led to the *2005 Regulations on PSI*.<sup>6</sup> From now on a wide range of public sector bodies within the UK would be required to publish clear statements as to the arrangements in place for release of data, the materials available and the licence terms and conditions applicable including, where relevant, clarity as to any access charges to be made for re-use. A fundamental objective of the Directive was to "unlock the economic potential of government owned data"<sup>7</sup> but it also represented a further step towards transparency and accountability.

---

<sup>2</sup> See for example: *Crown Copyright in the Information Age*, Cm. 3819, (1998); *The Future Management of Crown copyright*, Cm 4300, (1999); *Cost Cutting Review of the Knowledge Economy*, HM Government (2000); *Selling Government Services into Wider Markets*, HM Treasury (2003); *Electronic Government Services for the 21<sup>st</sup> Century*, Cabinet Office (2000); *Transformational Government: Enabled by Technology*, HM Government (2005); *Service Transformation (Varney Review)*, Cm 6683 D. Varney (2006); *Commercial Use of Public Information Office of Fair Trading* (2006).

<sup>3</sup> OS operates as a Trading Fund under the Government Trading Funds Act 1973 and the *Ordnance Survey Trading Fund Order* 1999 SI 1999 No. 965.

<sup>4</sup> There have been 27 statutory instruments under the Act since 2000 (see <http://www.legislation.gov.uk/>) including specific rules on access to Environmental Regulations at SI 2004 No. 3391.

<sup>5</sup> The PSI Directive was adopted on the basis of Art. 114 *Treaty on the Functioning of the EU* (TFEU); 95 *Treaty Establishing the European Union* (TEC) "as its subject matter covers the free circulation of services and the proper functioning of the internal market". Op. cit. note 9 post.

<sup>6</sup> SI 2005 No. 1515 on the *Re-use of Public Sector Information Regulations*. See further: *The Re-use of Public Sector Information: A Guide to the Regulations and Best Practice*, OPSI, June 2005.

<sup>7</sup> Brussels, 12.12.2011 SEC(2011) 1551 final, Commission Staff Working Paper – *Executive Summary of the Impact Assessment accompanying the document Proposal for a Directive amending Directive 2003/98/EC on the re-use of public sector information*, COM(2011) 877 final, para 1.2.

Following the Millennium, the debate strengthened with research on the issues leading to more precise data. Such input came from studies such as the *Power of Information Task Force Report*<sup>8</sup> of 2007. This challenged government assumptions about the value and role of information in the online environment of the Internet. For the first time evidence began to be brought forward suggesting that the lifting of restrictions on data, more systematic and transparent information about what data existed and charging structures based on marginal cost, was a better way forward in the long term for government and the nation. The 2009 *Digital Britain Report*<sup>9</sup> continued this line of argument, noting that data was “an innovation currency” and the “lifeblood of the knowledge economy”. It recognised more clearly than had been the case before, that data was a vital raw material for new information products and services, particularly those capable of visualising and analysing data from different sources.

Within the past five years, the basic model governing the economics of access and exploitation of PSI has largely been settled, and that is increasingly the position too with EU policy.<sup>10</sup> However, a European Commission review of PSI rules in 2009 reported that, while Member States had enacted measures to generate increases in re-use of PSI, particularly within the geographical and meteorological sectors, more action was needed. In September 2010, during a public consultation on the EU Directive, the Commission reflected on the fact that PSI now covered:

“All sorts of data generated by public sector bodies – e.g. maps, meteorological, legal, traffic, financial and economic information – that can be re-used by anyone else in innovative products such as car navigation systems, weather forecasts and travel information (‘apps’) that can be downloaded on smart phones..... [but] to realise the full potential of PSI for the EU economy, EU Member States must remove remaining barriers to re-use. These include discrimination between potential users, excessive charges for public sector information re-use and complex licensing policies..... lack of awareness of what public sector is available and failure to realise the economic potential of their data”.<sup>11</sup>

At the European level, it is now clear that further reform of the PSI Directive is now a “key action” within the *Digital Agenda for Europe* programme.<sup>12</sup> Achievement of this is seen as a source of “potential growth of innovative on-line services” and, since December 2011, central to its new *Open Data Strategy* for Europe. The Commission hopes this will generate up to €40 billion annually for the EU economy.<sup>13</sup> The strongest evidence of a fundamental shift away from past policies can be found in a remark by Commission Vice President Neelie Kroes: “We are sending a strong signal to administrations today. Your data is worth more if you give it away”.

---

<sup>8</sup> See: *The Power of Information: An independent review by Ed Mayo and Tom Steinberg*, June 2007; *The Government’s Response Cm 7157*, Cabinet Office, June 2007; and *The Power of Information Task Force Report* (February 2009).

<sup>9</sup> *Digital Britain: The Final Report* of the Department for Business, Innovation and Skills and Department for Culture, Media and Sport, 16 June 2009.

<sup>10</sup> According to a survey on existing findings on the economic impact of PSI by the European Commission in 2011 (*Review of recent studies on PSI re-use and related market developments*, G. Vickery August 2011) the “overall direct and indirect economic gains are estimated at €140 billion throughout the EU. Increase in the re-use of PSI generates new businesses and jobs and provides consumers with more choice and more value for money”.

<sup>11</sup> *Digital Agenda: Commission consults on re-use of public sector data*. IP/10/1103, Brussels, 9 September 2010.

<sup>12</sup> See: IP/10/581, MEMO/10/199 and MEMO/10/200.

<sup>13</sup> IP/11/1524 *Digital Agenda: Turning government data into gold*.

Under the programme, a new data portal will be established to host the release of Commission material and €100 million will be allocated in 2011-13 to fund research into improved data-handling technologies. Alongside will come a new PSI Directive as a “building block of *Digital Agenda for Europe*<sup>14</sup> and the *Europe 2020* strategy<sup>15</sup> for smart, sustainable and inclusive growth”.<sup>16</sup> Responses to the Consultation show that clarification and guidance is needed in respect of the charging and licensing principles and on data formats. As a result, the European Commission has indicated that the new Directive will:

- Make it a general rule that all documents made accessible by public sector bodies can be re-used for any purpose, commercial or non-commercial, unless protected by third party copyright;
- Establish the principle that public bodies should not be allowed to charge more than costs triggered by the individual request for data (marginal costs);<sup>17</sup>
- Make it compulsory to provide data in commonly-used, machine readable formats, to ensure data can be effectively re-used;
- Introduce regulatory oversight to enforce these principles; and
- Massively expand the reach of the Directive to include libraries, museums and archives for the first time. Existing 2003 rules will be applied to data from such institutions.<sup>18</sup>

### 3. UK readiness for PSI reform

If the European Commission is right in its assertion that the importance of Open Data, in particular government data, “is now more widely recognised,”<sup>19</sup> as the basis for new information services and products, how prepared then is the UK to respond to this challenge? It is important to recognise, when attempting to answer this question that the issue extends beyond economic analysis of UK performance within the digital economy. PSI is also “an innovation currency” for government policy development. The extent to which UK public sector has understood the full implications of this in relation to PSI is an issue worth exploring.

In the early days of digitisation, as government began to convert from the offline to the online world, the focus of ‘*eGovernment*’ was upon the efficiency savings to be gained from the introduction of information and communications technology (ICT) into public sector administration. Less thinking took place, however, as to how to utilise digitisation to improve the data flows required for policymaking, beyond continued deployment of traditional channels of consultation well known to Parliament and the Civil Service during pre-Internet days. Consultation remained an invitational process, based on prescribed terms as to what government wanted to consult about. Government had traditionally conducted policy development on those terms and saw no particular need for change.

---

<sup>14</sup> See: Digital Agenda (1 c).

<sup>15</sup> Launched on 3 March 2010 with the aim of turning Europe “into a smart, sustainable and inclusive economy delivering high levels of employment, productivity and social cohesion”.

<sup>16</sup> Op. cit. note 7, ante.

<sup>17</sup> In practice the European Commission believes that this means “most data will be offered for free or virtually for free, unless duly justified”.

<sup>18</sup> Op. cit. note 7 ante.

<sup>19</sup> Op. cit. note 7 ante, p. 3 - rationale for EU action, EU added value and subsidiarity.

Several factors have come together, however, to change the situation. Policy on re-use of PSI has matured at both EU and domestic level. The assumption that PSI can be more productive when widely distributed is now accepted and the duty to do so is likely to strengthen with the proposed reforms to the PSI Directive just described. Developments in ICT and potential advances in how the Internet delivers content are gathering in strength, and attracting government attention on political and economic grounds as to the implications for policy.

This paper will now consider these matters in relation to the introduction of Open Data portals; the extension of transparency and wider participation in policy development; improved public access to data; the impact of advances in PSI policy in specific sectors, such as spatial data; the erosion of barriers to cross-border use of PSI; the potential of 'Linked Data'; and conclusions as to the challenges that lie ahead.

### 3.1. Open data portals

The concept of Open Data Portals has “become mainstream” over the past three years, adopted by governments in pursuit of largely political objectives, mainly centred around openness and transparency in what they do.<sup>20</sup> ‘Data.gov’ was the first such portal, launched by President Obama in January 2009. In doing so he called for “more transparent participatory and collaborative government”. Executive departments and agencies would now “harness new technologies to put information out about their operations and decisions online and [make these] readily available to the public”.<sup>21</sup>

The UK equivalent, ‘data.gov.uk’ went online in September 2009, prior to an official launch in January 2010. This was in response to the outgoing Labour Government’s commitment announced in May 2009 to create a data service through OPSI. This would “expose government’s data feeds in a well ordered and useful way [and] provide a focal point for development using government information”.<sup>22</sup> A subsidiary objective lay in “holding public service providers to account for their activities”.<sup>23</sup> Leading the project, to provide the initial impetus, was Sir Tim Berners Lee, founder of the World Wide Web and Professor Nigel Shadbolt, both of Southampton University.

In the first instance, the primary focus was to “provide the data as quickly as possible”, with less emphasis on the “cleanliness of the data”. In cases where datasets had been specifically intended for a single original purpose, it was recognised that there might be difficulties in attempts to re-use such data in another context. To enhance understanding of these and other issues about the portal, data.gov.uk was supported by a wiki, blogs and a forum.

In April 2011 the portal reported that it had 6,200 datasets in machine-readable formats (CVS, XML, etc) licensed under the *Open Government Licence*<sup>24</sup> permitting commercial and non-

---

<sup>20</sup> European Commission, Information Society and Media Directorate General, *Pricing of Public Sector Information Study – Open Data Portals (E) Final Report*.

<sup>21</sup> President Obama’s *Open Government memo*, January 21, 2009.

<sup>22</sup> *Digital engagement: Update on power of information*, May 2009, Recommendation 14.

<sup>23</sup> Op. cit. note 20, ante, p. 39.

<sup>24</sup> See further: <http://www.nationalarchives.gov.uk>. The OGC has been developed by HMSO “as a tool to enable Information Providers in the public sector to licence the use and re-use of their information under a common open licence. The Controller invites public sector bodies owning their own copyright and database rights to permit the use of their information under this licence. ... These terms have been aligned

commercial use of data. These include COINS - HM Treasury's database of public spending and other datasets emanating from the Department of Health; Communities and Local Government; UK Statistics Authority; and Environment, Food and Rural Affairs.

Key to the impact of the dataset is Sir Tim Berners Lee's support for *Linked Data* in which "real world items and activities are given Uniform Resource Identifier (URI) addresses on the Web, and data about them is published in machine readable formats at those locations". The objective here is to enable individuals to "find out more about a particular item without information being copied into the original dataset".<sup>25</sup>

This transparency agenda *Making Public Data Public*, of outgoing Prime Minister Gordon Brown was continued after the election by David Cameron and the new Coalition government. In a letter to departments in May 2010 the Prime Minister announced further plans to open up government data including, inter alia, the publication of new central government ICT contracts, the salaries of Civil Servants above £150,000, central government tender documents over £10,000, new items of local government spending over £500, all new government contracts and crime data down to street level. These have since been added to the site.

In June 2010, responsibility for the project was handed to a new Transparency Board,<sup>26</sup> chaired by Francis Maude, the Minister for the Cabinet Office. Its remit includes responsibility for the implementation of the UK Government's *Transparency Agenda* articulated in set of public data principles that it published:<sup>27</sup>

- public data policy and practice will be clearly driven by the public and businesses who want and use the data, including what data is released when and in what form;
- public data will be published in reusable, machine-readable form;
- public data will be released under the same open licence that enables free re-use, including commercial re-use;
- public data will be available and easy to find through a single easy to use online access point (data.gov.uk);
- public data will be published using open standards, and following relevant recommendations of the World Wide Web Consortium (w3.org);
- public data underlying the government's own websites will be published in re-usable form for others to use;
- public data will be timely and fine grained;
- public data will be released quickly and then re-published in Linked Data form;
- public data will be freely available to use in any lawful way
- public bodies should actively encourage the re-use of their public data and

---

to be interoperable with any Creative Commons Attribution Licence, which covers database rights and applicable copyrights".

<sup>25</sup> Op. cit. note 20, ante, p. 41. See also note 123 et al, post. See: <http://linkeddata.org/> .

<sup>26</sup> Members include Sir Tim Berners Lee, inventor of the world wide web; Professor Nigel Shadbolt of Southampton University, an expert in open data; Tom Steinberg, founder of mySociety (mysociety.org); Andrew Stott, former Director of Digital Engagement at the Cabinet Office; and Dr Rufus Pollock, of Cambridge University, an economist who was one of the founders of the Open Knowledge foundation. As at October 2011, the data.gov.uk website was still operating as a 'beta version'.

<sup>27</sup> *The United Kingdom Report on the Re-use of Public Sector Information 2010 – unlocking PSI potential*, The National Archives, January 2011, para 2.7. The current text of the principles can now be observed at: <http://data.gov.uk/opendataconsultation/annex-2>.

- public bodies should maintain and publish inventories of their data holdings.

During this early stage of the process, perhaps the most significant point here is the political commitment to openness and transparency, as opposed to the combined impact of the datasets thus far released. Releasing datasets without context is in any case a limited exercise in transparency. As a ‘beta’ exercise, however, to test public enthusiasm to do something with the data, the results have been interesting. ‘Apps’ have been created that can locate post boxes, care homes and determine air quality, map road accidents, crime, landfill sites and pharmacies. An ‘*Unlocking Service*’ has since been introduced to “gather and assess requests to re-use information”, and this has stimulated some useable products,<sup>28</sup> applicable particularly to the mobile Internet environment.

A recent European Commission POPSIS study concluded that:

“While Open Data Portals appear to offer an important step in pushing forward the Open Data agenda and delivering its policy impact, their impact on opening up high added-value datasets is modest and their direct short term economic effects have been so far limited. Their largest impact to-date is indirect: the portals stimulate creativity and innovation and pave the way to unanticipated value creation. In this context the ‘start small’ approach appears to be the most effective. Open Data are a way to kick-start a process of cultural change that ultimately leads to the application of these high-level policy goals”.<sup>29</sup>

### 3.2. Tensions in data proliferation

A point that merits comment at this stage is whether it remains true that, despite the political commitment towards greater transparency, unnecessary controls still operate in respect of the release of PSI in the UK. From a legal perspective traditional copyright, contract, confidentiality and more recently, data protection regulation must always be considered. Taking account of such legal duties and responsibilities can create their own problems for open access. Work continues on a number of fronts, both nationally and internationally, to consider the suitability of these regimes in the digital environment for rights holders and consumers. In addition, where data is to be relied upon as authoritative of its contents, there need to be systems in place to demonstrate that provenance.

A consequence of open access is the proliferation of information. Some of this may find its way into poor quality or misleading online data resources or advertorial content, so it is important to preserve the integrity of trusted data where the importance of trust exists. To minimise the risks here, verifiable mechanisms must be established to protect the integrity of core data, by imposition of appropriate and measurable levels of provenance according to the profile and standards for that data. This raises further questions about access and archiving.<sup>30</sup>

---

<sup>28</sup> *MetroParis* and *London Tube* applications have generated jointly €400,000 in revenues. More revenue opportunities are likely to arise from apps that integrate different data sources alongside more value-added datasets and those that provide real-time data. The market for mobile apps is estimated to reach \$US 35 billion by 2015.

<sup>29</sup> European Commission, Information Society and Media Directorate General, *Pricing of Public Sector Information Study, Summary report*, October 2011, p. 8.

<sup>30</sup> See: Saxby, S, *National archives and records – the legal and policy considerations for the UK* Int. J. Private Law, Vol 3, Nos. 1/2, 2010.

How to move forward in a digital environment, where provenance may be a pre-condition of value of a dataset, is a complex question as there are many types of threat. The World Wide Web is a medium both for locating and accessing information, but it is also becoming increasingly a tool for data matching and integration, where the user is free to draw meaning. There will be many instances where, despite the proliferation of data sources, the lack of proven systematic integrity in content causes no particular harm. An individual seeking financial or medical advice for example would be advised to use either professional or government sources of information. Alternatively they can revert to the private sector where positive reputations exist. The issue is one of education – equipping the data user with the online judgement and skills to act intelligently when searching online for such content.

Innovation designed to enhance online searching, while intrinsically desirable as a means of sharpening the quality and accuracy of information flows, become still more powerful when *Open Data* concepts are applied that may add further relational links. What this shows is the importance of policy makers and relevant stakeholders beginning to understand these dynamics, so as to identify where action is needed. It is perhaps an issue that has received little analysis up to this point, at least from a policy perspective, but one that has been exposed by the move to more Open Data.

At the heart of this question for PSI is how to ensure the accuracy, integrity and appropriateness of data that is released. One can readily see examples where categories of data are subject to controls. For example, oversight of the use of personal data is subject to data protection legislation that requires adherence to a set of sound data principles. A regulatory framework has been created within which the data subject enjoys certain rights and choices and in which data users and controllers must comply with certain obligations. The desire for data accuracy and integrity in this case is just one of the factors motivating regulation. The broader picture is found in the desire to protect fundamental rights to privacy and respect of personal data, such as those articulated in the *European Convention on Human Rights* and other such statements of principle.

In this context it is the interests of the data subject and data controller, rather than due to the intrinsic importance of the data itself, that is the motivating factor for regulation. Whilst the prevalence of inaccurate personal data may be harmful to the data subject, the damage is likely to be confined. Different consequences ensue where data integrity converges upon the wider context. For example, in November 2005 the Chancellor of the Exchequer announced the Government's intention to legislate for "independence" in statistics.<sup>31</sup> The Consultation that followed explained the rationale:

“Statistics make a crucial contribution to good government in a modern democracy: assisting in the formulation and evaluation of policies; in the management of the services for which the Government is responsible; encouraging and informing debate; and allowing people to judge whether the Government is delivering on its promises. High quality statistics are also a key resource for business, academia and the wider community. With increasing emphasis on evidence-based policymaking and effective performance management, statistics have a greater importance than ever before, and an ever increasing scrutiny is placed upon them”.<sup>32</sup>

---

<sup>31</sup> HM Treasury (March 2006) *Independence for statistics: A consultation document*.

<sup>32</sup> *Ibid.* para 1.3.

A previous statement dating back to 1998 described the need for quality and integrity in such data in these terms:

“Quality needs to be assured. Official statistics must be sufficiently accurate and reliable for the purposes for which they are required ... the production and presentation of official statistics needs to be free from political interference and to be seen as such, so that the objectivity and impartiality of statistics is assured”.<sup>33</sup>

The net result of this exercise has been the *Statistics and Registration Services Act 2007*,<sup>34</sup> which replaces the previous non-statutory framework<sup>35</sup> and Office of National Statistics (ONS)<sup>36</sup> with a new body – the *Statistics Board*. This has statutory responsibility to “promote and safeguard the production and publication of official statistics that serve the public good”.<sup>37</sup> The latter has become the legal successor to ONS, a non-ministerial department with powers to “produce statistics, provide statistical services and promote statistical research”.<sup>38</sup>

The Act has left unchanged the continuation of the *Government Statistical Service*, a professional grouping of up to 7000 civil servants, working within the Board and with government departments and agencies in the collection, analysis and dissemination of statistics. However, a *National Statistics Code of Practice* has since been passed, governing implementation of the Act within government administration. One of 12 supporting protocols on Data Management, Documentation and Preservation, applies to all national statistics and requires managers to: “guard the integrity and security of the data in accordance with the organisation’s overall policies....; ensure that statistical resources are documented in a standard manner to increase usability and understanding of the data; and archive their resources in line with the organisation’s overall policy on data retention, preservation and destruction”.<sup>39</sup>

One can understand the importance of data provenance in this context where national policy, in many different areas of responsibility is contingent upon the accuracy of the data source in terms of both content and compilation methodology. Where a national dataset within the regulatory ambit of the legislation has been released as Open Data, it will have gone through these benchmarking requirements as to provenance, prior to distribution. Once classified as Open Data, however, these checks will no longer exist in respect of derivative data that represent the product of spontaneous individual research and creativity. The rationale of Open Data is deliberately set in this way. The objective is to expose the data to precisely this form of scrutiny, so as to capture any analysis or conclusions that government itself has failed to see. By such methods Open Data philosophy devolves the task to a relatively indiscriminate but potentially rewarding process of participation and recommendation.

---

<sup>33</sup> Office for National Statistics (1998a), *Statistics – A Matter of Trust: A Consultation Document*.

<sup>34</sup> *Statistics and Registration Services Act 2007* (c. 18) received Royal Assent on 26 July 2007.

<sup>35</sup> Office for National Statistics (2000) *Framework for National Statistics*.

<sup>36</sup> The Office of National Statistics (ONS) had been an Executive Agency accountable to the Chancellor of the Exchequer. It was headed by the National Statistician who was concurrently Registrar General for England and Wales within the General Register Office (part of ONS) which administered the system for the registration of births, deaths and marriages and civil partnerships in England and Wales.

<sup>37</sup> *Statistics and Registration Service Act 2007 Explanatory Notes* para. 6.

<sup>38</sup> *Ibid.* para 7.

<sup>39</sup> *National Statistics Code of Practice – Protocol on Data Management, Documentation and Preservation*.

So the question here is what to do about data inputs that draw upon original or derived datasets where provenance cannot be guaranteed? As a straightforward contribution to policy debate, one might imagine that this is not going to be a problem. Government can exercise its own judgement as to the value of such contributions. That experience will have built up over time, although under more benign conditions than those which encourage the torrent of data being circulated today. Prior to this, submissions were channelled and directed through public consultation. Views and opinions were invited from interested parties in direct response to such requests. Today, Open Data concepts induce a more indiscriminate submission of ideas and analysis, on a wider range of topics and from a broader constituency with varying degrees of value.

Perhaps the real issue here is about ensuring, in the most general sense that the right data gets into the right hands at the right time. That is not a statement about censorship, but one about ensuring that transparency, as a policy objective of government, works both ways. It is no longer simply an issue of accountability for actions taken: it is one that requires scrutiny of the data upon which policy is made. That data can then be evaluated and perhaps provide a more information rich resource for policymakers. For those that have professional, academic or other specialist contributions to make, the challenge for government must be to ensure such groups have access to appropriate data. Leaving politics aside, implicit in this is the assumption that if they do then policy, or the data upon which it relies, will acquire added value. If that premise is accepted then the task for government must be to define its data holdings, develop protocols to ensure ease of transfer and use, establish channels for communication, data assimilation and archiving, and train its public personnel to understand and apply these processes.

Illustrations of the problems that can occur in this context are reported by the inquiry into the integrity of scientific research concerning raw temperature data. This research was allegedly misrepresented by the *Intergovernmental Panel on Climate Change* (IPCC) in its advice to government. The university-based research unit that conducted the analysis for IPCC may not have used the most appropriate statistical techniques in its work, but an enquiry by leading scientists, nominated by the Royal Society, “criticised government for ‘impeding the flow of processed and raw data to and between researchers’ by adopting a policy of charging for access to environmental data collected by publicly funded researchers”.<sup>40</sup>

Another trend can be perceived. Thinking is now beginning to focus upon the data access needs of specific groups of users. The findings of such research may help in the development of distribution channels for data, including format and content, that may add value to the Open Data initiatives reported above. In 2012, the European Commission has indicated that it intends to “adopt a communication and recommendation on access to and preservation of scientific information in the digital age ...and set out the actions that the Commission intends to take on open access to publications and data in the context of research projects funded by the European Union budget”.<sup>41</sup> The need for this is indicated in the report of a recent online survey on scientific information, conducted as part of the European Commission’s 7<sup>th</sup> *Framework Programme on the European Research Area*.<sup>42</sup> The central question concerned the areas in

---

<sup>40</sup> *The Times*™ Online (14 April 2010) Climate scientists at East Anglia University cleared by inquiry.

<sup>41</sup> European Commission, DG for Research and Innovation (January 2012), *Online Survey on scientific information in the digital age*, p.8.

<sup>42</sup> *Ibid.* Responses were secured from a variety of stakeholders including: national, regional and local governments; research funding organisations; university/research institutes; libraries; publishers; international organisations; individual researchers; citizens; and others that included NGO’s; industries; charities; learned societies and scientific and professional bodies.

which the EU could “best contribute to improving the circulation of knowledge and, specifically, access to and preservation of scientific information”.

Of the 1140 responses received from 42 countries 84% of respondents believed there was an “access problem” to scientific publications in Europe, while 90% agreed or agreed strongly that “publications resulting from publicly funded research should, as a matter of principle, be in open access mode. Broadly similar results were reported in respect of access to research data where “lack of funding to develop and maintain the necessary infrastructures (80%); insufficient credit given to researchers for making research data available (80%); and insufficient national/regional strategies/policies (79%)” were cited. Strong support (90% of responses) favoured the public availability of research data and “results from public funding to be, as a matter of principle, available for re-use and free of charge on the Internet”.<sup>43</sup> Among barriers not addressed in the questionnaire, respondents flagged “a lack of skills and capacity for data management, a lack of standards for making research data exchangeable and re-usable and issues related to digital preservation and the long-term accessibility of digital content”.<sup>44</sup>

The broad merits of Open Data were summed up by one responder who said:

“whether you are a patient seeking health information, an educator wishing to enliven a lesson plan or a researcher looking to formulate a hypothesis, making papers freely available online provides you with the most current peer-reviewed scientific information and discoveries”.<sup>45</sup>

It is important to recognise that the ownership and management of data and research maintained by commercial publishers raise fundamentally different issues from those that arise with the release by government of PSI and public datasets. The most obvious distinction of course lies in the fact that the primary focus of journal publishers is in the publication of research and not in the release of primary data. The contribution to public policy of journal papers lies in their impact in persuading policy makers to access and use the analysis in the course of policy development. These are commercial organisations that seek a return on their investment in the information resource to which they are right holders viz. peer reviewed authored research.

Publisher investment is also now extending to the development of better tools for its discovery, exploitation and use. Elsevier B. V. for example, which supports a global community of 7,000 journal editors,<sup>46</sup> maintains ScienceDirect – a full text database of journal articles and book chapters from some 2,500 journals and 11,000 books – currently 9.5 million sources growing at 0.5m per annum. Elsevier states that this platform offers:

“sophisticated search and retrieval functionality that enables the user to maximise the effectiveness of their knowledge discovery process. New tools facilitate research work flow aids, such as access to content, at an early publication stage and efficient multiple document downloading of content that can be stored, printed and passed to colleagues. The web environment offers new ways to present information as well as enhancing it with other content sources based on semantic technologies, e.g. NextBio. In addition,

---

<sup>43</sup> Ibid. Executive summary, p. 5.

<sup>44</sup> Ibid. p. 30.

<sup>45</sup> Ibid.

<sup>46</sup> Elsevier reports that the community also includes 70,000 editorial board members, 300,000 reviewers and 600,000 authors.

since 2003, many authors have been submitting extra value-added content associated with the research, such as audio and video files, datasets and other supplementary content, effectively accelerating research beyond the print format”<sup>47</sup>.

Such investment needs to be paid for, but that does not mean that access options cannot be agreed by commercial providers to ensure that open access principles are respected, particularly to the output of publicly funded research. In many instances that research will be freely available on peer-to-peer sites, including those maintained by universities themselves. But tensions can arise between the research community and publishers in respect of these arrangements as arose early in 2012 with Elsevier.<sup>48</sup>

These are not questions likely to engage government or regional institutions, such as the EU, unless at the macro level of competition policy and reform of intellectual property law, specifically in the field of copyright. In that regard there is on-going analysis, within the framework of international treaties and domestic legislation, to look at specific issues such as the copyright treatment of ‘orphan works’,<sup>49</sup> so as to free up data for archiving in datasets and databases. Consideration of the terms of engagement between open access policies and Linked Data principles<sup>50</sup> will also concern both public and private information providers as these initiatives accelerate in popularity. Beyond that, the questions surrounding development of research analysis and its impact on public policy remain centred upon the level of investment in primary research, the quality of data accessible to those researchers and the political context in which such research is received and applied by government. Otherwise, the work of publishers in this field can be readily separated from the objectives of PSI policy already articulated.

### 3.3. Current proposals on Open Data

The UK Government now seems set upon consideration of the merits of *Open Data* having launched a consultation on ‘*Making Open Data Real*’ in August 2011.<sup>51</sup> This is intended to inform the Government’s overall approach towards transparency across government and public

---

<sup>47</sup> See: <http://www.info.sciverse.com/sciencedirect/about>

<sup>48</sup> Elsevier states that the essence of its work “is to create and sustain journals that make it possible for researchers to have their work efficiently reviewed, enhanced, validated, recognised, discovered and made highly accessible, in perpetuity, to readers in virtually every country in the world.

<sup>49</sup> The Government announced on 14 December 2011 that a Consultation following up the recommendations of ‘*Hargreaves Review*’ *Digital Opportunity – A Review of Intellectual Property and Growth* (May 2011) was to run until 21 March 2012 that would consider establishing “licensing and clearance procedures for ‘orphan works’ (material with unknown copyright owners). This would open up a range of works that are currently locked away in libraries and museums and unavailable for consumer or research purposes. Source: Intellectual Property Office (14 December 2011) Government plans to improve UK’s copyright laws”. See also *Proposal for a Directive on certain permitted uses of orphan works* COM(2011) 289 final, Brussels 24.5.2011 which proposes a legal framework for cross-border digitisation and dissemination of orphan works in the single market.

<sup>50</sup> Op. cit. note 127, post.

<sup>51</sup> Cabinet Office, (4 August 2011) *Making Open Data Real*. The consultation closed on 27 October 2011. In this context Open Data means: “Data which can be freely used, re-used and distributed by anyone (<http://www.opendefinition.org/government>). In relation to public services Open Data means data available under the terms of Open Government Licence. The presumption is that data about public services will be Open Data. It may be that some data held in relation to public services is made ‘available’, but is charged for. Source Op. cit. note 58 post Section 7. Annex A.

services.<sup>52</sup> The stated intention is to “to embed transparency and Open Data as core operating principles of the public services”. The focus is upon the conditions that need to be in place to ensure access to such data is accomplished, so as to enhance “public service outcomes and productivity, social and economic growth”.<sup>53</sup>

Initial results of the consultation<sup>54</sup> suggest widespread support for the principles of Open Data and a belief in its ability to enhance performance, deliver public services and foster economic growth. Data should be available free of charge with the cost implications met by government. This is based on the rationale that the focus is upon “the value added by individuals or organisations using data” rather than “recouping the costs of making data available”. There was almost unanimity of view that government should play a lead role in establishing common Open Data standard(s) in collaboration with the wider Open Data community. The principle should be that the standard(s) is “accessible across a variety of organisations and systems and that user needs are accounted for”. In addition, in order to remedy the present “piecemeal” arrangements for tracking data, an “effective data inventory” would need to be created, although views differ as to whether this should take the form of a “centrally held catalogue” or “a series of inventories that reflect the diverse nature of the sectors organisations, potentially subject to Open Data requirements, operate in”.<sup>55</sup>

The Government has since announced that, from March 2012, it is to co-chair a new partnership, the *Open Government Partnership*, to develop discussion on transparency and Open Data at international level.

A parallel and quite significant component of the UK Government’s move toward Open Data has been the proposal to establish a *Public Data Corporation* (PDC).<sup>56</sup> If established, this will bring together, under single departmental sponsorship, those organisations “whose primary purpose is collecting, managing and disseminating data and providing value-added services based on that data”.<sup>57</sup> What one perceives here, perhaps overtly for the first time, is the perception that some data are ‘public data’ in a deeper sense than its PSI origins imply. This occurs when they comprise “essential ‘core reference data’ (such as addresses, company numbers etc.) that have become “accepted identifiers for physical and non-physical entities in the economy and wider society”. Whereas successive Governments “sought to improve delivery of public services through openly publishing public data, sharing these datasets has had as powerful an effect on those producing it, as it has on those receiving it”. The fact that such data are now more accessible “has stimulated the development of an information market”.<sup>58</sup>

Ordnance Survey, the Meteorological Office and HM Land Registry<sup>59</sup> are singled out as fulfilling these criteria, in contrast with those departments and agencies for whom the collection

---

<sup>52</sup> The policy is intended to be a “lever for a wide range of positive outcomes: increasing accountability, building public confidence in government bodies, stimulating efficiency gains within the public sector, promoting greater citizen engagement and stimulating economic growth”. Op. cit note 54 post.

<sup>53</sup> <http://data.gov.uk/blog>, 20 October 2011.

<sup>54</sup> Cabinet Office, (February 2012) *Making Open Data Real – A Government Summary of Responses*.

<sup>55</sup> Ibid. paras 1.54-1.57.

<sup>56</sup> The Government announced plans to establish a Public Data Corporation in January 2011.

<sup>57</sup> Statement by the Cabinet Office announcing the closure on the PDC Consultation (October 2012).

<sup>58</sup> HM Government (August 2011) *A Consultation on Data Policy for a Public Data Corporation*, para 1.11.

<sup>59</sup> The PDC will comprise public sector organisations “which are subject to the requirements set out in HM Treasury’s publication *Managing Public Money*.”

of public data is simply a “by-product” of public sector service delivery e.g. health, education, transport and criminal justice. The Government believes that single departmental sponsorship of these organisations, within the PDC, will provide “a more consistent approach towards access to and accessibility of public sector information; ... create a centre of excellence driving further efficiencies in the public sector; and create a vehicle that can attract private investment”.<sup>60</sup>

In August 2011, the Government launched a Consultation<sup>61</sup> on the data policy framework under which the PDC will operate. This concerns the charging, licensing and regulation of PSI that the PDC will produce for re-use. For critical observers of PSI policy over two decades or more, it is interesting to see how thinking has matured from the days when the principle focus, under *Trading Funds* regulation,<sup>62</sup> was the licensing of PSI. This ensured that the income from such data products and services be applied as a contribution towards operational costs.

Since then, economic and policy analysis of the issue has pointed more strongly to the benefits to the wider economy from collaboration and sharing of data. This, in turn, has led to “the vast majority” of Crown copyright material now being made available for re-use under *Open Government Licence*. Under this proposal the emphasis will be upon “how best to balance affordability considerations and the implications for attracting external capital into a PDC with the ambition to release more data for free”. The Government now sees potential for private investment in this task. Barriers to the use and re-use of PSI are also under scrutiny, such as “licensing and issues around access and release”.<sup>63</sup>

Up to this point one observes the effort by government, EU and the wider community to fashion new ways to access and utilise PSI resources particularly online. The instigator of this has been the rapidity of technological advance in ICT which, in turn, has generated new information products and services. This has been stimulated by innovative data tools built around the Internet and World Wide Web. Such change has occurred alongside pressure to open up access to data and information resources, previously subject to varying forms of control that have since, to some extent, evaporated. In addition, within all sectors, the search for funding has continued in order to cover the costs of investment in data creation, identification and recovery, as well as tools for its exploitation, access and use.

Public sector motivation for this has primarily resided in the desire for greater efficiency in the delivery and consumption of public services and more recently as a direct contributor to policy. Private sector engagement has been profit driven in the collective desire to support, via publication of peer-reviewed content, the stimulation of economic activity. Led by a variety of

---

<sup>60</sup> Op. cit. note 54, ante.

<sup>61</sup> Op. cit. note 58, ante.

<sup>62</sup> Fundamentally the Trading Fund model is one of “breaking even one year with another after allowing for operating costs, investment needs, loan repayments and agreed levels of dividend.” Revenue shortfalls could in theory be recovered “by an appropriate combination of increased productivity, efficiency savings, reduced costs, lower dividends and curtailing loss making on core activities”. See: *Spatial Data Infrastructures in the United Kingdom: State of Play 2010*, Spatial Applications Division, K.U. Leuven Research and Development, September 2010, para 2.3.7.

<sup>63</sup> These entities will be subject to the existing policy and regulatory framework for re-using information held by public sector bodies (Crown copyright and Crown database rights; Re-use of PSI Regulations 2005 and The Information Fair Trader Scheme (IFTS)) but “the extent to which these regulations and policies apply individually, and in combination, differ for different public sector bodies”. The Consultation indicates that it is likely within PDC that “all will apply to some degree, depending on the ultimate structure and classification of PDC or its constituent parts”.

motivations, individuals and groups have further formed themselves into movements to encourage relaxations of restrictions on access and re-use of data held by both the public and private sectors. Others have sought to improve World Wide Web capability to link data to other data, and to improve the value of the latter as a tool for enhancing knowledge and understanding. The voluntary and routine release of datasets for general public scrutiny is likely to stimulate this process further. Over time the application of the collective mind to a wide range of issues may in turn help inform and shape the actions that governments take. In the broadest sense one observes here a desire to use the tools of ICT, constantly advancing through innovation, to produce a society that is better informed and more engaged and a government that is more accountable and better advised.

#### 4. Shared standards for datasets

In addition to the analysis of proposed EU and domestic reforms of PSI policy already discussed, it is important to consider the issue of shared standards. These reside among different categories of data, and may have a role to play in the integration of data and data analysis leading to shared insights into policy.

Some datasets, as a category of data, are clearly more powerful and influential than others and this is particularly true in respect of data linked to location. Since 2005, when the Government established a *Geographic Information Panel (GIP)* to advise Ministers on “geographic information issues of national importance”, the link between such data and policy formation and management has become ever more clear. At that time, however, “no overarching strategy” existed for managing geographic information across the UK.

Consequently, in April 2006, as part of its *Transformational Government Implementation Plan*,<sup>64</sup> Ministers asked the GIP to create a strategy. The outcome was the launch in 2008 of *UK Location Strategy (UKLS)*.<sup>65</sup> Shortly after it was also tasked with UK implementation of an EU Directive to create an *infrastructure for spatial information*<sup>66</sup> across Europe (INSPIRE).<sup>67</sup>

At the heart of location strategy is the premise that “everything happens somewhere”. As a result:

“When different types of information about a particular place are compared or related to each other, this can increase considerably the understanding, and hence the power to make decisions about a particular place .... for central and devolved government geographic information supports effective policy formation and evaluation .... key areas where the Location Strategy will be of benefit are in policy and operational areas of the public and private sector where shared and integrated place-based information is valuable for decision-making. These include planning for communities, environment, health,

---

<sup>64</sup> *Transformational Government Implementation Plan*, Cabinet Office, April 2006.

<sup>65</sup> *Place matters: the location strategy for the United Kingdom*, Communities and Local Government and Geographic Information Panel, November 2008. Following discussion with senior officials in government it was decided to rename ‘geographic information strategy’ with its new name - ‘location strategy’ – at this time.

<sup>66</sup> Article 3.2 of Directive 2007/2/EC defines spatial data as data with a direct (e.g. grid coordinates) or indirect reference (e.g. place name, postcode) to a specific location or geographic area. This is considered as wider in scope than previous definitions.

<sup>67</sup> *Directive 2007/2/EC establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)*.

education, security, construction, transport, crime prevention, insurance, retail, energy, climate change, agriculture, heritage, sport, employment and statistics”.<sup>68</sup>

From the outset one of the key tasks that UKLS set itself, was to tackle the problem of information from a range of sources that were neither joined-up nor integrated. Investigation found that there were too few government-owned datasets incorporating location data that “could be easily assembled and analysed with reliability”. In addition, there was “too much duplication, too little re-use and too few linkages across datasets needed to support policy development and implementation.”<sup>69</sup> A “consistent reference framework” was needed, which in turn would lead to “more effective cross-organisational processes, far greater sharing and re-use across the public sector and beyond”.<sup>70</sup>

#### 4.1. INSPIRE implementation

Since 2008, the UK response to these challenges can be found within the broader framework of INSPIRE, following the entry into force of the Directive in May 2007. UK implementing Regulations SI 3157 came into effect on 31 December 2009, commencing a process of transposition and implementation through the *Location Programme* that will continue until 2019. While the Directive does not require the collection of new spatial data, it does set a number of common *Implementing Rules* (IR) in the specific areas of metadata, data specifications, network services, data and service sharing, and monitoring and reporting.<sup>71</sup> Between 2008 and 2012 such IRs have been adopted in Commission Decisions and Regulations and are binding in their entirety.

INSPIRE regulation applies to the spatial datasets and services held by or on behalf of public services and to their use by public authorities in the course of their public task.<sup>72</sup> It is said that, when fully implemented, the Directive will “theoretically, enable data from one Member State to be seamlessly combined cross-border with data from all other States”. From an EU perspective, it is deemed particularly important in tackling issues related to the environment such as “planning, pollution control, environmental protection” as well as climate change concerns e.g. “flood control, water management, extreme weather event contingency and many others”.<sup>73</sup>

Within the UK there are a number of influences now coming together to aid implementation of UKLS and INSPIRE. The public face of the programme is located on the website of the

---

<sup>68</sup> Op cit, note 65 ante pp. 8 &10.

<sup>69</sup> *Getting Started – Initial Guidance to Data Providers and Publishers – Guide 1: UK Location*, Edition 2.0, (uk location) July 2011

<sup>70</sup> Op. cit. note 65 ante, p. 9.

<sup>71</sup> INSPIRE *Metadata Regulation* No. 1205/2008, 3.12.2008 and *Corrigendum*, 15.12.2009; *Commission Decision* 2009/442/EC regarding INSPIRE monitoring and reporting, 5.06.2009; *Commission Regulation (EC) No 976/2009 as regards Network Services*, 19.10.2009 and *amendment* No 1088/2010, 23.11.2010; *Commission Regulation No 268/2010 on INSPIRE Data and Service Sharing*, 29.03.2010; *Commission Regulation 1089/2010 as regards interoperability of spatial datasets and services*, 10.12.2010 and *amendment* No 102/2011, 04.02.2011.

<sup>72</sup> For UK guidance on ‘public task’ see: *Public task principles – Principles defining public task for public sector information under PSI Regulations*, The National Archives, 6 February 2012) and *Guide to drawing up a statement of public task – Information for public sector bodies on producing a statement of public task under the Re-use of Public Sector Information Regulations 2005*, The National Archives, Version 1.0, August 2011). For definitions of public authority see circa note 100, post.

<sup>73</sup> *The INSPIRE Directive – A Brief Overview*, Association of Geographic Information, 2007.

Department of Environment, Food and Rural Affairs (DEFRA). Programme co-ordination lies with the *UK Location Council* (UKLC) - a cross-government group of local, devolved and central government organisations,<sup>74</sup> with the remit to deliver UKLS, implementation of INSPIRE and associated initiatives. This includes “monitoring technical advances, interoperability and information exchange within which geographic information (GI) is collected and managed; promoting best practice and supporting innovation in the collection and use of GI; facilitating a coordinated position on potential legislation; ... tackling data quality and integrity issues [including] investment requirements; articulating the economic, national productivity and competitiveness benefits of an effective location-information infrastructure for the UK; and identifying other medium and long-term location information issues”.<sup>75</sup>

Beyond UKLC, participants in policy development include the *Cabinet Office*, in respect of its responsibilities for e-government and, through OPSI, policy towards PSI; the *Intra-Governmental Group on GI* (IGGI), representing central government departments; *The Improvement and Development Agency for Local Government* (IDEA); and the *Association of Geographic Information* (AGI), representing more than 1000 data producers and users from public and private sectors.<sup>76</sup> Input via the Scottish Executive and Welsh National Assembly will also be secured.

In terms of engagement with the policy framework surrounding location data strategy, the UK was an early starter. In the late 1990's initial impetus came from the *Digital National Framework* (DNF), in effect an industry body established to “promote the integration and sharing of location-based information from multiple sources”.<sup>77</sup> The work programmes of DNF, in contributing to theoretical underpinnings, technical guidance and best practice in the development of location-based data, have proved invaluable.<sup>78</sup> Since then, convergence has centred upon *UK Location Information Infrastructure* (UKLII) as the framework to deliver implementation of INSPIRE and the need to “share and access location information from across the UK, using core data provided by trusted sources”.<sup>79</sup> A number of UK initiatives and services, operating within the sphere, are now working towards the “coordinated publishing, discovery, sharing and re-use of location information”.<sup>80</sup> These will contribute to the development of UKLII

---

<sup>74</sup> The most active organisations include: Registers of Scotland; Local Government Association; DEFRA; Ministry of Defence; Welsh Assembly Government; Communities and Local Government; Statistics Board; Environment Agency; Ordnance Survey; Land and Property Services (Northern Ireland); Office of Public Sector Information; Land Registry; British Geological Survey; and Department of Transport.

<sup>75</sup> UK Location Council, *Revised Governance Arrangements*, Version 1.0, 14 September 2009, Appendix 1: Terms of Reference.

<sup>76</sup> Op. cit. note 62, ante.

<sup>77</sup> *Digital National Framework (DNF) – Overview* Version 3.0, 28.10.2010.

<sup>78</sup> *Implications of the INSPIRE Directive – A DNF White Paper* Version 1.0c (Digital National Framework, April 2008) p. 7 - For example: providing a framework for the better integration of data; using location as a common denominator with objects referenced through unique identifiers; linking and exchange of information; consistent forms of geo-referencing to provide information integrity; associating data from different sources; supporting information transfer and sharing or using data in conjunction with others in cross organizational applications.

<sup>79</sup> *UK Location Programme – A guide to the UK Location Information Infrastructure* (UKLII), uk location December 2009, p. 3.

<sup>80</sup> Ibid. para 70. These are broad ranging in terms of purpose and scope. Examples of SDI-related initiatives include: MAGIC – A DEFRA led project providing a single base for rural and countryside information; the UK Environmental Observation Framework (UK-EOF); the Marine Environmental Data and Information Network (MEDIN); the National Underground Asset Group (NUAG); the Marine Data and Information Partnership (MDIP); Vertical Offshore Reference Framework (VORF); Integrated

along with significant service providers such as *Ordnance Survey* and *Land & Property Services* in Northern Ireland.

At the heart of UKLII, and the adoption of the *Implementing Rules* within INSPIRE, are the data sharing<sup>81</sup> arrangements that centre upon metadata<sup>82</sup> and the services available to enable users to search and display their contents. It is implicit to the success of INSPIRE that mechanisms be put in place to “allow ‘GI search engines’ to search the catalogues of producers and custodians” of spatial datasets. The approach towards metadata is set out in Article 5 of INSPIRE together with a timetable for implementation in Article 6.<sup>83</sup> This makes it clear that public authorities will be responsible for the “establishment, management, maintenance and distribution of spatial datasets and services”, while Member States must ensure that metadata are created for the latter, as defined in Annexes I, II and III of the Directive. Some 34 data layers are specified and the Annex in which they feature is determined by which set of slightly varying implementation conditions apply:

### **Figure 1 INSPIRE Spatial Data Scope**

#### *Annex I*

1. Co-ordinate reference systems
2. Geographical grid systems
3. Geographical names
4. Administrative units
5. Addresses
6. Cadastral parcels
7. Transport networks
8. Hydrography
9. Protected sites

#### *Annex II*

1. Elevation
2. Land cover
3. Identifiers of properties
4. Ortho-imagery (Aerial Photography)
5. Geology

#### *Annex III*

1. Statistical units
2. Buildings

---

Coastal Hydrography project – a metadata discovery portal to identify hydrographic surveys and SeaZone – providing access to hydrographic and other marine and coastal data.

<sup>81</sup> See: *UK Location Data Sharing Operational Guidance, Part I – Policy Context*, Edition 1.0, March 2011.

<sup>82</sup> *Metadata and GIS – An ESRI White Paper* (October 2002) p. 2: “Metadata is a summary document providing content, quality, type, creation, and spatial information about a dataset. It can be stored in any format such as a text file, Extensible Markup Language (XML), or database record”.

<sup>83</sup> Article 3.6 of INSPIRE defines ‘metadata’ as “information describing spatial datasets and spatial data services and making it possible to discover, inventory and use them”.

3. Soil
4. Land use
5. Human health and safety
6. Utility and governmental services
7. Environmental monitoring facilities
8. Production and industrial facilities
9. Agricultural and aquaculture facilities
10. Population distribution – demography
11. Area management/restriction/regulation zones & reporting units
12. Natural risk zones
13. Atmospheric conditions
14. Meteorological geographical features
15. Oceanographic geographical features
16. Sea regions
17. Bio-geographical regions
18. Habitats and biotopes
19. Species distribution
20. Energy resources
21. Mineral resources

The scheme established for the UK sets out that data providers will publish their data and online services into UK Location “by creating and publishing discovery metadata”<sup>84</sup> through its central metadata catalogue. The *UK Discovery Metadata Service (DMS)* is one of a number of business services supporting UKLII that underpins “the coordinated and regulated publishing of public sector location information to INSPIRE and UK Location specified standards”. It will provide the “discovery component for a set of online services that will allow data users to evaluate and use public sector location information”: this includes to “view, download and invoke as part of an end-to-end business application”.<sup>85</sup>

Whereas in the past the task of creating metadata:

“has often been part of a later documentation activity, for example as part of the production of a catalogue, or directory of information resources ..... created by someone removed from the creation of the data and thus lacking knowledge about the data.... UK Location encourages metadata to be produced as part of the data production process itself, as part of the same tools, and stored in parallel with the data, ideally in the same data storage. It seeks to retain a clear distinction between the master record held by the Data Publisher and the copy held by the UK Location central metadata catalogue service”.<sup>86</sup>

To facilitate this service the public data publishing platform ‘data.gov.uk’ will extend its functionality to encompass DMS within its portal.<sup>87</sup> The assessment is that:

---

<sup>84</sup> Discovery metadata is information about a data or service resource, used to discover and access its suitability for sharing or re-use.

<sup>85</sup> *Operational Guide – UK Location Discovery Metadata Service*, Edition 2.0, July 2011, uk location.

<sup>86</sup> *Getting Started – Initial Guidance to Data Providers and Publishers Guide 4: Publishing Discovery and View Services*, Edition 2.0 (uk location, July 2011).

<sup>87</sup> To publish the data and services metadata into UK Location, providers must first register as a Data Publisher using the central registration services of data.gov.uk.

“In the spirit of openness and transparency, data.gov.uk has been set up to provide a one-stop shop for public data. INSPIRE and UK Location Programme (UKLP) are both initiatives aimed at providing easy access to environmental public data. Rather than create a separate portal for accessing geospatial data, this principle acknowledges that in many senses there is nothing special about such data, and that as a result it should be accessed through a common public data portal. This ensures that such data is exposed to a wider audience, and is treated consistently with other public data”.<sup>88</sup>

UK Location discovery metadata will use an application profile of the UK industry metadata standard UK GEMINI2, version 2.1 [7].<sup>89</sup> The next stage will be to create and publish a View Service for published datasets, both under INSPIRE obligations and voluntary distribution. INSPIRE Regulation for a View Service defines quality of service criteria based on the OGC standard for Web Mapping Services (WMS) in which “data users can access and view location information published by a wide range of data providers”.<sup>90</sup> The Government has asked Ordnance Survey to take on the technical delivery role of the services that are required to meet Britain’s obligations under INSPIRE.

## 5. Analysis

Observing PSI policy in 2012, one perceives a much more energetic policy than was ever the case during in the early years of the Internet and of ‘transformational government’ approaches designed to respond to it. First of all, there is a firmer commitment to greater transparency, articulated in the Coalition Government’s *Transparency Principles*. Then there are the proposals for a new PSI directive in the wake of a review that concluded that the progress and implementation of Directive 2003/98/EC was “uneven” and that a number of “remaining barriers” existed to be tackled. But perhaps most important of all is the explosion of ideas about data and how best to exploit it that one can see illustrated in ‘*Making Open Data Real*’.<sup>91</sup>

Policy began with the modest and unambitious belief that PSI really only had value in terms of the revenues it could generate from distribution under licence. This led in turn to more detailed scrutiny of the wide range of data available and the possible uses to which it could be applied. The initial focus upon the economics and consequent reduction of barriers to exploitation, gradually gave way to an acceptance that re-use of PSI, while clearly beneficial economically, had more to offer in terms of democratic engagement.

---

<sup>88</sup> UK Location Programme – Location Information Interoperability Board, Design Co-ordination Group - Design Principles for UKLII (uk location, March 2011) para 2.1.

<sup>89</sup> UK Location Programme – Location Information Interoperability Board, Design Co-ordination Group - Technical Architecture Overview (uk location, October 2010) para 3.1.2 - While UKLP is providing a supported tool for creating UK GEMINI2 metadata, use of this tool is not mandated. Any other tool that can create metadata can be used. However, the tool will need to support the publication of data as UK GEMINI2-compliant eXtensible Markup Language (XML) either through an Open Geospatial Consortium (OGC) Catalogue Service for the Web (CSW) interface or via a Web Accessible Folder (WAF) i.e. an HTTP-accessible directory of files, in which all files and their time-stamps are visible to a web browser or client. All the files in this folder must be UK GEMINI2 XML files.

<sup>90</sup> Op. cit. note 83 ante, p. 11. This can be accomplished through a single map viewer application or one which a computer application can access and use within an end user business application. The OGC WMS standard has been adopted as an ISO standard, ISO 19128.

<sup>91</sup> Op.cit, note 51, ante.

The delivery mechanism for innovation in PSI is ICT, an obvious pre-condition for the development of new applications. Beyond PSI, the results of such connection can be observed in abundance in social networking, cloud computing, mobile Internet and many other applications. This has led to a desire to look more deeply into data as an entity and into its holistic relationships, in the belief that there is much to be gained if these layers can be exposed and understood.

The collection, integration and re-use of location data, in particular, as a tool for beginning to understand and reach decisions about what is happening in the world around us, is one of the main reasons why arrangements for the collection and distribution of spatial data has been among the first to be examined. What one observes within the EU is a highly structured regime, with a series of stages set out in implementing rules and a schedule for their introduction. Core geospatial data is defined by theme and specification, supported by a mandatory framework of discovery metadata. The stated aim is to create a “pan-European Spatial Data Infrastructure” and to “improve the interoperability of spatial information across the European Union at a local, regional, national and international level”.<sup>92</sup> The objective is to facilitate improvements in sharing spatial information between public authorities and in so doing improve public access to it.

An obvious question that needs to be considered here is whether such highly structured models are necessary to achieve these stated objectives and whether they represent the way forward for handling different categories and types of data. At the heart of INSPIRE is the desire to offer access to datasets that can be defined, categorised and validated. The disciplines involved in this require considerable strategic and operational planning. UK data providers must first establish their own compliance policy, identify relevant datasets and their strategy for each, culminating in decisions as to what data will be published into UK Location and how INSPIRE policy and standards will be integrated into the organisation.<sup>93</sup> Once this is decided stage two requires decisions as to “how to publish” i.e. the operational arrangements involved, including whether to do so directly or through a third party.<sup>94</sup> Even if the latter route is chosen the considerable task of supplying the raw data to that party remains.

While the initial focus will be on publishing the data as it is now, this must lead, in due course, to making such data “INSPIRE compliant” i.e. legally compliant as bound by INSPIRE Regulation. Conformance with the technical requirements involved will require investment in skill, training and resources.<sup>95</sup> Where third party intellectual property rights (copyright and database right) in

---

<sup>92</sup> James Reid, *The EU INSPIRE Directive: An Infrastructure for Spatial Information in the European Community* Version 2.1, EDINA National Data Centre, June 2011.

<sup>93</sup> *Getting Started – Initial Guidance to Data Providers and Publishers Guide 3: What Needs to Happen and When*, Edition 2.0 (uk location, July 2011). Data referred to in INSPIRE Regulation must be compliant, but data providers may also wish to publish additional spatial objects not yet defined in the Regulations.

<sup>94</sup> Op. cit. note 86, ante.

<sup>95</sup> Data providers are responsible for the costs of ensuring their data complies with INSPIRE metadata and data interoperability standards and making their data available via network services. However, the right to impose conditions for access and use of data is retained. Charges must be kept to the minimum, “to ensure necessary quality and supply of datasets and services”, which means under normal circumstances that public authorities will seek to recover only marginal costs.

spatial data reside, these must be accommodated.<sup>96</sup> PSI Regulations<sup>97</sup> are complementary to INSPIRE Regulations: any licensing of the re-use of public sector spatial data will be subject to the former.<sup>98</sup> Obligations under *Freedom of Information* legislation and any other applicable statutory duties remain unaffected.

As the model unfolds, uncertainties have arisen that will need to be resolved. One relates to what happens when, under Article 4.2 Directive 2007/2/EC, “multiple identical copies of the same spatial data are held by or on behalf of various public authorities”. The Directive applies “only to the reference version from which the various copies are derived”. Compliance with INSPIRE would only be expected of the reference version, which would normally be that which is held by the organisation responsible for the data. However, no structure is presently in place to determine which is that copy. This means “multiple versions of the same (or similar) datasets can/may (and likely will) exist”.<sup>99</sup>

Uncertainties also exist as to the significance, under INSPIRE, of products derived from these data.

Another issue that has attracted debate is whether universities can be data providers as well as users within the terms of INSPIRE. Article 3.9 Directive 2007/2/EC states that:

“‘public authority’ means:

- (a) any government or other public administration, including public advisory bodies, at national, regional or local level;
- (b) any natural or legal person performing public administrative functions under national law, including specific duties, activities or services in relation to the environment; and
- (c) any natural or legal person having public responsibilities or functions, or providing public services relating to the environment under the control of a body or person falling within (a) or (b).”

Contrary to UK PSI Regulations,<sup>100</sup> this definition is in line with Freedom of Information regulation,<sup>101</sup> and is likely to include universities and research councils. The *Guide to INSPIRE*

---

<sup>96</sup> Article 4.5 of Directive 2007/2/EC states that in the case of spatial datasets to which the Directive applies: “but in respect of which a third party holds intellectual property rights, the public authority may take action under this Directive only with the consent of that third party”.

<sup>97</sup> SI 2005 No.1515, Op. cit. note 6 ante. However, the latter imposes no general obligation on public sector bodies to make their data available for re-use. If re-use is permitted then the regulations must be complied with. OPSI, part of The National Archives, “regulates the information trading activities of government Trading Funds and other PSI holders using two key tools: the Information Fair Trader Scheme (IFTS) and the disputes resolution service.” In addition to the full IFTS process, ‘IFTS online’ exists “to help public sector organisations in meeting their responsibilities as holders of re-usable information”. (op. cit. note 27, ante, Part 3: Governance and supervision regulation).

<sup>98</sup> Op. cit. note 81 ante, p.10.

<sup>99</sup> Op. cit. note 92, ante.

<sup>100</sup> Regulation 5(3) SI 2005 No. 1515, op. cit. note 7 ante, “excludes education and research establishments such as schools, universities, archives, libraries and research facilities (such as research councils)”. See: *The Re-use of Public Sector Information: A Guide to the Regulations and Best Practice*, Office of Public Sector Information, June 2005, para 3.6. As such they are not required to draw up statements of public task.

<sup>101</sup> *The Freedom of Information Act 2000* (c.36), Schedule 1 part 4, paras 53(1) (a) to (e) and 55(1)(a) and (b).

*Regulations*, which deals with its application to public authorities, interprets Regulation 3 as *including*, for example, “government departments, local authorities, the health service, police or other public bodies; those carrying out public administration functions; or those with responsibilities, function or the role of delivering services relating to the environment”.<sup>102</sup>

If defined as *public bodies*, universities will be required to comply with the Directive in respect of spatial datasets produced, received, managed or updated by them within the scope of their public tasks.<sup>103</sup> The original proposal for the Directive, however, refers only to “the private sector, universities, researchers and the media” as likely *beneficiaries* of INSPIRE. Moreover, universities were not directly included in the work plan for implementation and *EDINA* reports that, aside from itself, “there are no special Spatial Data Interest Communities related to education and very few people involved from the educational sector in the technical development of the Implementation Rules”.<sup>104</sup>

However, the European Commission has stated:

“Whether or not a dataset falls under the INSPIRE obligations does not depend on the scale, the specificity of the datasets, or the level of government involved in their management. When the datasets, at any level of government, are relevant for developing, implementing or monitoring laws or regulations which may have an impact on the environment, INSPIRE obligations should apply. Such conditions could equally apply to datasets collected by a *research project activity* as the INSPIRE Directive makes no distinction between ‘operational’ and ‘research’ datasets. INSPIRE could be considered a positive incentive to safeguard valuable research datasets after the ending of a project.”<sup>105</sup>

*EDINA* believes that, whatever the position, in terms of their public task,<sup>106</sup> it is unlikely that university institutions will, “in the first instance”, hold much geospatial data within the terms of INSPIRE, at least under Annex I and II Themes.

---

<sup>102</sup> *A Guide to the INSPIRE Regulations*, SI 2009 No. 3157, DEFRA, Land & Property Services, Welsh Assembly Government, December 2009. The definition of ‘public authorities’ is that used in the *Environmental Information Regulations* SI 2004 No 3391 (EIR). The Guidance states “‘as a rule of thumb’, if your organization responds to requests for information under EIR, you may assume it is also a ‘public authority’ for the purposes of these Regulations”.

<sup>103</sup> Article 4.1. Directive 2007/2/EC states: “This Directive shall cover spatial datasets which fulfill the following conditions: (a) they relate to an area where a Member State has and /or exercises jurisdictional rights; (b) they are in electronic format; (c) they are held by or behalf of any of the following: (i) a public authority, having been produced or received by a public authority, or being managed or updated by that authority and falling within the scope of its public tasks; (ii) a third party to whom the network has been made available in accordance with Article 12; (d) they relate to one or more of the themes listed in Annex I, II or III”.

<sup>104</sup> *Op. cit.* note 92, ante. *EDINA* is a UK national academic data centre, designated by Joint Information Systems Committee (JISC) on behalf of UK funding bodies “to support the activity of universities, colleges and research institutes in the UK, by delivering access to a range of online services through a UK academic infrastructure, as well as supporting knowledge exchange and ICT capacity building, nationally and internationally”. See: <http://edina.ac.uk>.

<sup>105</sup> *Ibid.* *Report of the Workshop on the Legislative Transposition of the INSPIRE Directive 2007/2/EC*, 17 April 2008.

<sup>106</sup> The Scottish Information Commissioner has noted: I would question, however, whether it is possible to say that a university will never have public tasks for the purposes of ... [INSPIRE] ... it is not unknown for EU law to deal with universities on the basis that they do discharge public functions”. Source: *Op. cit.* note 92, ante.

However:

“As the focus shifts to the third annex, it is *possible* that data held within universities might come within scope e.g. species distribution, habitats and atmospheric conditions”. Moreover, “studies of environmental change require an understanding of how phenomena change over time. This requires access to historic data and earlier editions of data which may be held by universities (or rather researchers and research teams within universities). In both cases, universities would be required to make these data available.”<sup>107</sup>

While the position of universities within the INSPIRE regime, in terms of the definitions of ‘public authority’ and ‘public task’, may ultimately be a matter for the courts, EDINA believes the main benefit will be that a large number of geospatial datasets will become available for research and use in higher education. However, given dispensation on charges, “researchers could find themselves in the strange position of getting data for free from one country but paying for a similar type of data in another”.<sup>108</sup>

There is an additional concern about INSPIRE that relates to ‘function creep’. INSPIRE defines 34 SDI data layers set out in the three annexes to the Directive. These are subject to *Implementing Rules* and a timetable for inclusion. The objective is for core geospatial data described there to conform to a strict product specification. But this does not mean that it must necessarily become an EU wide specification, but merely a recorded specification of an available dataset within the relevant Annex category held by a public authority within a Member State. At the heart of INSPIRE is the desire to ensure that Member States have accessible datasets that are known and can be validated against published criteria. That is not the same as the creation of an ‘EU wide standard’ in which, for example, deviation in excess of such compliance criteria is deemed to be in breach.

*Land Cover Map 2007*,<sup>109</sup> for example, is a UK product developed by the *Centre for Ecology & Hydrology* (CEH)<sup>110</sup> of *Natural Environment Research Council* (NERC).<sup>111</sup> It has a specification and may be interpreted as a national dataset within the scope of Annex I. INSPIRE would have it meet specification requirements and to document these in a particular way via UML. That is one level of standardisation, but beyond that, any attempt to impose an EU standard, e.g. as to particular map characteristics, beyond sensible voluntary agreement, could affect the value of the data to UK policy – in this example in relation to land/biodiversity development.

---

<sup>107</sup> The EU INSPIRE Directive – And what it might mean for UK academia, EDINA. EDINA also points out what the European Commission has said: that it is “a fundamental right of third parties to enrich the European Spatial Data Infrastructure with datasets that are currently hidden or difficult to find” and that this is also the philosophy underpinning UKLP for spatial data infrastructure.

<sup>108</sup> Op. cit. note 92, ante.

<sup>109</sup> CEH has produced over the past 22 years three digital land cover maps: LCMGB 1990; LCM 2000 and LCM 2007. Each of these has been produced as a number of different products with varying data formats and spatial resolutions. LCM2007 has improved thematic and spatial accuracy over its predecessors, providing continuous coverage of habitat distributions across the UK. These products are available under licence for academic, non-commercial and commercial use.

<sup>110</sup> CEH “hosts the Environmental Information Data Centre for terrestrial and freshwater sciences which brings together wide-ranging, nationally-important datasets and expertise in managing diverse types of environmental data. See further: [www.ceh.ac.uk](http://www.ceh.ac.uk).

<sup>111</sup> NERC is the UK’s main agency for funding and managing research, training and knowledge exchange in the environmental sciences. It is a non-departmental public body and receives funding from the Department for Business, Innovation and Skills.

There were attempts back in 1985 to standardise an EU digital land cover database (CORINE “Co-ordination of Information on the Environment”) but problems arose with this in relation to variable data access policy, lack of consistency with other data, irregular updating, a lack of long term perspective, quality/reliability and synchronization with other MS Data. INSPIRE is a *Framework Directive*: detailed technical provisions can be found in the *Implementing Rules*. But this does not and should not be interpreted as a ‘one-size-fits-all’ direction: the merits of retaining principles of subsidiarity and the flexibility this offers as technology advances should be among the lessons learnt from the CORINE experience.

## 6. Removing barriers to re-use

As this paper has shown “the world of information does not stand still”.<sup>112</sup> In the case of PSI its currency has grown in stature as digitisation has matured, so that it is now perceived as “primary material for digital content products and services with a large hitherto unexploited potential”.<sup>113</sup> Removing barriers to re-use has become the dominant theme so as to release PSI to fulfil a wider ambition. Access restrictions and licensing policies have been relaxed in parallel with cost reductions. Commitments in the form of *public data principles* now express the Coalition’s approach to bring PSI into use. This is supported by the Government’s transparency agenda ‘*Making Open Data Real*’.<sup>114</sup>

The EU is also now moving on this with the launch, in December 2011, of an ‘*Open Data Strategy for Europe*’, which the European Commission believes will “deliver a €40 billion boost to the EU’s economy each year”.<sup>115</sup> It talks about Europe’s public administrations “sitting on a goldmine of unrealised economic potential”.<sup>116</sup> Reform of the PSI Directive 2003/98/EC is intended to carry this through.

But this approach towards unlocking access to PSI can achieve little unless systematic effort is applied to the identification, cataloguing, formatting and distribution of the data itself. Engaging ministers as to the merits of developing a unified and integrated approach to the management of PSI assets has not always been easy. Since 2006, that task as well as responsibility for the management of PSI has been in the hands of OPSI, backed by the policy on transformational government<sup>117</sup> and an *Advisory Panel on Public Sector information (APPSI)*.<sup>118</sup> This comprises experts drawn from commercial, government and academic sectors. Evidence of the challenges for PSI policy can be found in APPSI’s Annual Report for 2006:

---

<sup>112</sup> Op. cit. note 27, ante p. 3.

<sup>113</sup> Op. cit. notes 7 & 16, ante, p. 4.

<sup>114</sup> Op. cit. note 51, ante. For the principles see note 27, ante.

<sup>115</sup> Op. cit. note 13, ante.

<sup>116</sup> Ibid.

<sup>117</sup> *Transformational Government – Enabled by Technology* (HM Government 2005). The policy, announced in November of that year, was about the “design of IT services more around the citizen, and the move to a shared services culture” – i.e. simpler, faster, more cost effective services for citizens and businesses. Op. cit. note 8 ante, para 39.

<sup>118</sup> APPSI’s role is “to advise Ministers on how to encourage and create opportunities in the information industry for greater re-use of public sector information; to advise the Director of OPSI and Controller of Her Majesty’s Stationery Office about changes and opportunities in the information industry, so that the licensing of Crown copyright and PSI is aligned with current and emerging developments; and to review and consider complaints under Re-use of PSI Regulations 2005 and advise on the impact of the complaints procedures under those Regulations.

“Most APPSI members have been disappointed in the past year with our inability to stimulate and secure ministerial interest in PSI at the Cabinet Office. It will be recalled that many of our recommendations in last year’s report required ministerial engagement. Perhaps because APPSI did not make its case forcefully enough or perhaps because Cabinet Office Ministers had other more pressing and mainstream demands on their time, the reality is that APPSI has not met with any Ministers over the past 18 months, despite attempts to set up meetings. Still less have Ministers pursued PSI initiatives”.<sup>119</sup>

So the question arises what has been the catalyst for change to bring PSI to the fore? Founder of the Web and *World Wide Web Consortium* (W3C), Sir Berners Lee, believes that it began:

“with lunch at Chequers [in 2009] when the Prime Minister [Gordon Brown] asked me what I felt the UK should do to make the best use of the Internet, and I said, you should put all your government data on the web. And he said, ‘okay then, let’s do it’. So when one has spent a lot of one’s life persuading people to put things onto the web, and persuading people to be open, it’s almost disarming to have somebody say that straight away. The result of that was a team in the Cabinet Office under Andrew Stott [Director of Digital Engagement at the Cabinet Office at that time]. Various people in the UK government had experience of this already so it was a question of how to accelerate this as much as possible”.<sup>120</sup>

The Coalition Government continued the strategy so that the Chancellor of the Exchequer, Rt Hon George Osborne could say in May 2011 that:

“Over the next 12 months we’re going to unlock some of the most valuable datasets still locked away in government servers. That is the raw data that will enable you, for the first time, to analyse the performance of public services, and of competing providers within those public services”.<sup>121</sup>

But behind this rhetoric are some fundamental questions that need to be considered. PSI can exist in a variety of forms and for a variety of purposes. Matching form with purpose in an appropriate manner is as important as the decision to release the data in the first place. From the outset it is important to consider the context in which such data is to be created, held, used and distributed. In the case of spatial information the objective has been to establish an infrastructure for spatial information in Europe that will support the purposes of EU environmental policy. The data disciplines that represent the core of what INSPIRE stands for are considered necessary to ensure that the dissemination of environmental information held by public bodies across Europe can be located, matched and used authoritatively and effectively by policymakers and researchers who can rely on its format and provenance.

---

<sup>119</sup> (APPSI) 2006, *Realising the Value of Public Sector Information- Annual Report 2006*. For a full discussion see: Saxby S, *Public sector information and re-use – where is the UK now?* Int. J. Private Law, Vol 1, Nos 3/4, 2008.

<sup>120</sup> Comment made during an Interview on 5 January 2010 with *Prospect Magazine* editor, Tom Chatfield, just before the official public launch of data.gov.uk on 21 January 2010. W3C aims to bring “diverse stakeholders together, under a clear and effective consensus-based process to develop high-quality standards based on contributions from the W3C Members, staff, and the community at large.

<sup>121</sup> Speech by Rt. Hon. George Osborne at *Google Zeitgeist 2011*.

The 34 themes, core reference geographies and the *Implementing Rules* on metadata, interoperability of spatial datasets, network services, data and service sharing, monitoring, reporting and co-ordination, are in place because the objectives of Community policy and use of the data are thought to require it. Far from being an isolated case, the European Commission believes that this infrastructure has the potential to extend to other policy areas in the future, such as agriculture and transport.

## 7. Linked Data

The structured nature of INSPIRE is in sharp contrast to the unstructured release of more than 7000 datasets for public scrutiny and analysis under the Government's Open Data initiative, '*data.gov.uk*'. The rationale here is clear. It is built around transparency, openness and a desire to encourage input and exchange of ideas about policy. While the distribution of datasets needs a context to fulfil its purpose, this does not necessarily require the kind of structures for so doing implicit in INSPIRE.

Chancellor of the Exchequer, Rt Hon. George Osborne has observed:

“If the first impact of the internet age on government has been to change accountability, the second has been to change the nature of policymaking itself. Just as the old asymmetries of information have been eroded, so too have the perceived asymmetries of wisdom. I genuinely believe that in almost all areas of government, we do a better job when we open up ourselves to the ideas of the crowd .... To those that say people are disengaged from the work of government, and want their representatives to take care of everything, this is a powerful riposte”.<sup>122</sup>

The aim of *data.gov.uk* and its Open Data philosophy then is not simply about passive delivery of transparency, but more active engagement by government with the community it serves. The latter is invited to apply its collective mind to published data, extract new information from it, draw connections not previously identified, spot errors of fact and potential efficiencies to be gained, as well as defects in drafting and implementation of policy. Identification of defects and shortcomings in the data itself is also a benefit to be gained.

Encouraging Government to open its doors to data scrutiny, however, is only part of the agenda of Sir Tim Berners Lee, supported by the organisation he founded, *World Wide Web Consortium* (W3C). His broader objective is to implement fundamental change in the way the '*http protocol*' of the Internet is used for data sharing. The suggestion is that this “can be extended from all the documents posted online to the things that the documents are about – products, places, events, concepts and so on”, thereby connecting “otherwise disparate data from many sources”.<sup>123</sup>

The concept of '*Linked Data*' is about connecting at the level of the data rather than the document in which a *universal resource identifier* (URI) is used as a means of uniquely identifying the 'Thing' or 'Resource'.<sup>124</sup> Clicking at the data level “creates a powerful way to view hidden relationships between things and answer more informed questions” and in so doing

---

<sup>122</sup> Ibid.

<sup>123</sup> *UK Location Programme – Linking information and location – A guide to the benefits of Linked Data and the UK Location Strategy*, uk location, April 2010, p. 2.

<sup>124</sup> Chief Technology Officer Council, *Designing URI Sets for the UK Public Sector – A report from the Public Sector Information Domain of the CTO Council's cross-Government Enterprise Architecture, Interim paper*, Version 1.0, October 2009, para 7.

begin to “dismantle data silos and enable cross-organisational data sharing [and] in turn ... support better decision making and the delivery of more effective, as well as innovative, services”.<sup>125</sup>

The Government’s *Chief Technology Officer Council* has stated that:

“URI sets can be published by the UK public sector to provide comprehensive and reliable identifiers for ‘things’ such as schools, roads, legislation, locations, projects, events and so on. Where the quality of these sets can be described consistently, other data owners will have the confidence to re-use them in their own data, leading to a web of data that can be linked, queried, and aggregated”.<sup>126</sup>

Linked Data is defined as a key element of the White Paper ‘*Putting the Frontline First: Smarter Government*’ and therefore of ‘*data.gov.uk*’. The Government indicated that it planned to make “a number of important technical improvements to public data” with the “aim for the majority of government-published information to be re-usable, Linked Data by June 2011”.<sup>127</sup> In the interim came the General Election, although the policy with regard to *data.gov.uk* has continued.

To underpin the Government’s ‘*Making Public Data Public*’ objectives, in respect of geographic information, the UK’s national map-maker, Ordnance Survey (OS), has begun to publish some of its projects as Linked Data via *data.gov.uk*, as well as in alternative formats from its own ‘*OS OpenData™*’ website. This has followed a Consultation<sup>128</sup> to assess the implications, given that up to 90% of OS revenues have been derived from product sales. Although OS will continue to charge for some products, it is also recognised that it needs “to make it easier for customers and re-users to access its data and services”.<sup>129</sup>

The coming together of these elements of *location policy* is likely to have had a strong impact on the decision to proceed with the long awaited ‘*National Address Gazetteer*’.<sup>130</sup> In 2012, this is expected to bring together the spatial address databases of both OS (*MasterMap Address Layer 2*) and the *Local Government Improvement and Development Agency* (LGID) into a single more accurate, INSPIRE and UKLS compliant, set of maintained geo-referenced addresses.<sup>131</sup>

---

<sup>125</sup> Op. cit. note 123, ante.

<sup>126</sup> Ibid. para 9.

<sup>127</sup> *Putting the Frontline First: smarter government*, Cm 7753, Presented to Parliament by the Chief Secretary to the Treasury, December 2009, p.28.

<sup>128</sup> *Policy options for geographic information from Ordnance Survey, Consultation*, Department for Communities and Local Government, December 2009. See also, Impact Assessment, and Government Response, March 2010.

<sup>129</sup> Op. cit. note 27, ante para 2.18. One of the new services is *OS OnDemand* – “a web service for directly licenced customers delivering mapping over the web directly into organisations”. See *Ordnance Survey, OS OnDemand – User guide and technical specification* 11/2011 and *Ordnance Survey, Annual Reports and Accounts 2010-11* HC1188, TSO September 2011.

<sup>130</sup> For further comment see: Saxby S, *Three years in the life of UK national information policy – the politics and process of policy development*, Int. J. Private Law, Vol. 4, No 1, 2011 pp.1-32.

<sup>131</sup> This ends a dispute of more than a decade in which the owners of the data could not agree on the way forward. In February 2011, (*OFT Press Release 18/11*, 15 February) the Office of Fair Trading announced that it had decided not to refer the issue to the Competition Commission. Its Chief Economist, Amelia Fletcher commented: “Comprehensive and accurate spatial addressing information is important in delivering frontline public services, as well as for certain private sector customers, so any competition

The concept of Linked Data has also been grafted into UKLP where it is perceived as likely to contribute significantly if “barriers to connectivity” are removed, such as “local data stores or databases which act as isolated silos, since there is no common way of joining them up.”<sup>132</sup>

UK location comments:

“Many organisations may keep raw, unadulterated data back from publication, fearing the costs of making it fit for use by third parties. Other barriers include proprietary data structures, different data formats and resolutions, different coordinate systems and little classification. All of this presents immense problems for users who wish to integrate data from several sources from across the public sector. The concept of Linked Data can help address these issues. Imagine you have information about a spatial object and you want to know what it is connected to. In a Linked Data world you could click on that object and get back links to everything that is connected to it on the web. Alternatively, you can perform further analysis across the web using database querying”.<sup>133</sup>

At first glance, one might conclude that Linked Data would not sit comfortably within the structured environment of INSPIRE. However, it is suggested that both have rules<sup>134</sup> based on best practice and both depend on persistent identifiers and ontologies. UKLC believes that “wrapping core INSPIRE services and data provision with wider (cross domain) open standards”, of the kind found in Linked Data, can “provide public service information in a location context beyond the typical geospatial communities”.<sup>135</sup>

This approach also has the support of those working in the DNF community and within OS. While location is important “as a means to reference other information to data representing real world objects” that information is diverse. It goes beyond location and “impacts on all information domains whether health, statistics, transport and so on. The alignment with the original DNF concepts is very strong”.<sup>136</sup>

In the interim, the European Commission has initiated investigation of Linked Data and related issues through its *Linked Open Data* (LOD2) project<sup>137</sup> that runs to September 2014. This is looking at “exploitation of the web as a platform for data and information integration and the use

---

concerns resulting from the joint venture needed careful consideration. A merger to monopoly would normally warrant further investigation. However, the Government’s buying power, combined with expected benefits from combining these two databases, made reference to the Competition Commission disproportionate”. The database will be managed by Richard Duffield, of GeoPlace, a public sector limited liability partnership between Local Government Association and OS.

<sup>132</sup> Op. cit. note 123, ante, p. 3.

<sup>133</sup> Ibid.

<sup>134</sup> Linked Data proposes a five star rating for data on the web to provide some metrics for evaluating particular datasets: \* available on the web (whatever format), but with an open licence; \*\* As (one star) plus available as machine-readable structured data (e.g. Excel instead of image scan of a table); \*\*\* As (two star) plus use of a non-proprietary format (e.g. CSV and XML); \*\*\*\* All the above plus use of open standards from the World Wide Web Consortium (W3C) such as RDF (Resource Description Framework – W3C specification) and SPARQL21 to identify things, so that people can point at your stuff; \*\*\*\*\* All the above, plus link your data to other people’s data to provide context. Op. cit. note 51, ante para 8.9.

<sup>135</sup> *UK Location Council Response to Consultations on Data Policy for a Public Data Corporation and Making Open Data Real*, uk location Version 10, October 2011, p. 18. See further note 51, ante.

<sup>136</sup> Comment reported on 16.2.11 by DNF Expert Group Chair, Keith Murray or Ordnance Survey.

<sup>137</sup> See: <http://lod2.eu/>.

of semantic technologies to make government data more useable". The *OpenAIRE project*<sup>138</sup> comprising partners from 25 EU Member States and several associated countries has, since December 2009, been building a "participatory infrastructure for the EC Pilot for *Open Access to Research Information*". Work has also been undertaken, via the ISA Action on Semantic Interoperability (SEMIC.EU),<sup>139</sup> to promote the idea of Open Government Metadata as a "first step towards metadata alignment at both national and European level."<sup>140</sup>

## 8. Conclusion

For regular observers of PSI policy this is a fascinating time. After what would seem to be years of comparative inertia, with policy edging forward towards a more open and flexible approach, the position has changed to the point when one might legitimately question what exactly the public sector, indeed the government as a whole is for? Progress has been made so that some of the basic issues connected with PSI, such as the opening up of data for public access and re-use, are no longer partisan. What is less clear are the respective roles of UK public and private sectors regarding the "supply chain" for PSI. Different partnership models now exist in the exploitation of Open Data and in the distribution of information products. Proposals for a *Public Data Corporation* are rooted in this assumption.

Government is no longer perceived as simply the provider of data it happens to hold, but an entity with obligations, perhaps more a "platform... than producer of products";<sup>141</sup> and as much a consumer as creator. To the extent that government is adding value it is doing so, more often than not, by "'co-mingling' its own data with that from private sector sources, muddying the intellectual property rights involved".<sup>142</sup> It is also, to the extent that this is legally imposed, going to be subject to obligations of data capture, storage, preparation, description and supply applicable to specific categories of PSI such as spatial data. UKLS will take this strategy forward, but of equal importance will be the accretion of expertise, particularly within public sector administration, that will be necessary to deliver that task.

It remains to be seen how far the INSPIRE model will expand. The policy is certainly an intensive consumer of resources, but with that comes an information rich resource that is immediately usable across jurisdictions, meets minimum standards of form and content, supported by high standards of provenance in the integrity of the data.

Alongside such a model comes the 'share data now' philosophy of Linked Data as an open, modular and scalable resource<sup>143</sup> in which the value of what is produced is likely to be measured more in the actual results of the process than in reliance upon the provenance of the data.

---

<sup>138</sup> <http://www.openaire.eu/>.

<sup>139</sup> <http://www.semic.eu>.

<sup>140</sup> Reported in: *Communication from the Commission ... Open data – An engine for innovation, growth and transparent government*, COM(2011) 882 final, Brussels, 12.12.11, para. 3.2.2.

<sup>141</sup> *APPSI response to the Open Data Consultation*, Advisory Panel on Public Sector Information, December 2011, p. 5.

<sup>142</sup> *Ibid.*

<sup>143</sup> Tim Berners Lee, *Putting Government Data online*, 30 June 2009 at: <http://www.w3.org/DesignIssues/GovData.html>. Open: Linked Data accessible through a variety of applications; Modular: Linked Data combined (mashed-up) with any other piece of Linked Data such as health care expenditure combined with population characteristics of the area to assess the effectiveness of government programmes; Scalable: ability to add more Linked Data to what is already there, even when the terms and definitions that are used change over time.

Provenance was never really the issue in respect of Linked Data: it is more about establishing the connections to data and then seeing what this produces. Because of its simplicity, such a model may well favour the disenfranchised, in terms both of access and influence.

But perhaps of most immediate concern is what government will do with its ‘core reference datasets’.<sup>144</sup> It is in the national interest that these be continuously maintained and their integrity secured. These are the datasets which, in the words of APPSI:

“carry a “moral right for the public to have access (e.g. details of all UK laws) or where there are large safety, public service, efficiency and cost benefits if everyone uses the same definitive and regularly updated sources of data.”<sup>145</sup>

Just as government has a *National Infrastructure Plan* for its “financial, human, intellectual, natural and physical capital”<sup>146</sup> so too does it need to define a national information infrastructure plan for data so that it can maintain its core material. Adoption of such a plan would define the ‘*public task*’ for its core reference datasets and ensure that pursuit of data policy for PSI, in what are clearly challenging and fast moving times, is conducted from a sound and secure base.

**17,563 words**

---

<sup>144</sup> An obvious example would be products produced from the National Address Gazetteer database. This will be available to all customers through Ordnance Survey, and to the public sector under the terms of the new Public Sector Mapping Agreement which, for the first time, allows free access to Ordnance Survey data under single arrangements from April 2011.

<sup>145</sup> *APPSI response to the Public Data Corporation (PDC) Consultation*, p. 11.

<sup>146</sup> *National Infrastructure Plan 2010*, HM Treasury & Infrastructure UK, October 2010, para. 1.1.

