OPEN DATA FOR ECONOMIC GROWTH

June 25, 2014

Transport & ICT Global Practice

Table of Contents

Acknowledgments:

Standard Disclaimer:

Copyright Statement:

Summary

1.   Many governments are pursuing Open Data policies. One of the key policy drivers has been to use Open Data to drive economic growth and business innovation. This paper examines the evidence for the economic potential of Open Data and concludes that, despite a variation in published estimates and some methodological difficulties, the potential is very large indeed. It reviews the latest data about companies using Open Data, and highlights four companies which did not exist ten years ago, which are driven by Open Data, and which are each now valued at around $1 billion or more. It discusses the five archetypical types of businesses using Open Data, and cites concrete examples of each, and discusses the types of data which are proving most likely to lead to widespread business adoption and innovation. It suggests that governments should see themselves not only as a Supplier of Open Data but also as a Leader, Catalyst and User of it. Finally it makes some recommendations for policies and actions which governments could take to maximise the economic growth possible from their data.

Introduction

2. Governments collect or generate a lot of data for their own purposes or from their own operations. *A priori* the collection and generation of this data is justified by the benefits of the use of this data within governments for the purposes for which it was collected. The benefits may be financial, for instance in terms of policy making, resource allocation or operational management and optimisation; or they may be non-financial, for instance in meeting legislative and other requirements for transparency and accountability. In principle if governments did not gain a net benefit from collecting and using the data then they would not collect it in the first place.

3. Governments may also collect data which is not of direct benefit to government itself but in order to provide wider economic and social benefits. For instance cadastral information contributes to an authoritative system of land ownership and proof of title, which has important economic benefits. In some cases the provision of information is a natural monopoly and the government collects it as the best-placed actor to do so - for instance the Danish government intervened to create a single, accurate, national address register for all sectors of the economy - including the public services themselves[1].

4. However because use of data is "non-rivalrous" the fact that governments (or others) have used the data for the purpose for which it was originally collected does not prevent that data being used for other purposes - by others or, indeed, by other

---

[1] http://www.adresse-info.dk/Portals/2/Benefit/Value_Assessment_Danish_Address_Data_UK_2010-07-07b.pdf

parts of the government itself. Economic theory suggests[2] that benefits are maximised when access to the information is priced at the marginal cost of distribution - and the internet has made the marginal cost of distribution of digitised data by download from the web effectively zero. While there may be distribution costs such as setting up an Open Data portal and policy and technical work to implement the extraction data from internal databases, these are mainly one-off expenditures which are usually infeasible or uneconomic to amortise in recurrent charges for data access - the amounts involved are typically small compared to the on-going costs payment collection and administration that would be necessary to recover them.

5. So many governments have now launched "Open Data" programmes to make their digital, machine-readable data available to business and citizens to use and reuse free of charge for any lawful purpose. For most governments the policy objectives have been a mixture of four distinct drivers: economic growth, including business innovation and the creation of companies and jobs; closer and more inclusive citizen engagement in improving public services, particularly by giving citizens information on standards of provision and service performance which equips them to be more engaged "consumers" of public services; increased transparency and accountability; and improvement in the efficiency and operations of public services themselves, for instance through better decision-making through access to data from other agencies.

6. For the citizen engagement, transparency and efficiency objectives Open Data has assisted and enhanced civic and governmental processes already underway. However the release for the first time of government data has created completely new opportunities for the opportunities for economic growth, innovation and jobs - both in the creation of economic activity in new, "data-rich" services themselves and in the wider economic benefits of the use of these services by others in the business and personal sectors.

The Size of the Potential Prize

7. Open Data is a relatively new concept in its current, recognisable, form, dating from 2006-07 - although in some sectors and in some countries there were earlier developments which would now be seen as Open Data[3]. Even in countries which started relatively early to release government information as Open Data the benefits still appear to be developing. Like other innovations, it is likely to be many years until there is full adoption and the implications for downstream users have been fully understood. Indeed, as discussed later in this paper, the usage of Open Data

---

[2] See for instance Pollock et al "Models of Public Sector Information Provision" 2008
http://www.berr.gov.uk/files/file45136.pdf

[3] See for instance Weiss 2003 http://www.codata.org/archives/2003/03march/papers/Weiss.pdf

depends not only on the availability of the data itself but also on the ideas of how to use it, and the tools to use it effectively.

8.   However various policy-makers have commissioned studies of the potential economic benefits of Open Data as an evidence base for decisions which need to be taken now in order to take the initial step of releasing data on open terms.

9. The European Commission commissioned two studies[4] as the evidence base for their proposals to amend the EU Reuse of Public Sector Information Directive. The POPSIS study looked at some case studies where previously restricted data had been made Open, and showed a consistent pattern of much greater usage as a result. The second study, by Graham Vickery of Information Economics, reviewed the evidence from studies on public sector information re-use as a whole and other market developments related to data. The Vickery study found that the size of the narrowly defined EU direct, "business as usual", public sector information reuse market was of the order of 28 billion Euro in 2008 with an annual growth rate of around 7%. However this excluded domains where re-use was not a principal activity and the value of government activities. In addition it was expected that open public sector information policies, easier access to information and the abolition of charging would further stimulate the market, with direct public sector information use and re-use activities increasing to 40 billion Euro a year.

10. Importantly the Vickery study also investigated the indirect benefits of open data - for instance the additional benefit to the personal or business customers of using the services provided by the "direct" re-users. In total, and with easier access, improved infrastructure and lower barriers, aggregate direct and indirect economic benefits for the whole EU economy could have been of the order of 200 billion Euro in 2008 - 1.7% of the GDP of the EU as a whole.

11. More recently McKinsey[5] have claimed that, globally, seven sectors[6] alone could generate more than $3 trillion a year - and perhaps as much as $5 trillion a year - in additional value as a result of open data. While they recognised that open data is in its early days, they showed that it is already giving rise to hundreds of entrepreneurial businesses and helping established companies to segment markets, define new products and services, and improve the efficiency and effectiveness of operations. McKinsey saw a clear potential to unlock significant economic value by

---

[4] Vickery: http://ec.europa.eu/digital-agenda/en/news/review-recent-studies-psi-reuse-and-related-market-developments

Deloitte/POPSIS: http://ec.europa.eu/digital-agenda/en/news/economic-analysis-psi-impacts

[5] "Open data: Unlocking innovation and performance with liquid information" 2013

http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information

[6] Education, Transportation, Consumer Products, Electricity, Oil & Gas, Health Care and Consumer Finance

applying advanced analytics to both open and proprietary knowledge to help uncover consumer preferences and allowing companies to improve new products.

12. In the United Kingdom, one of the leading governments in the field of Open Data and one where the economic benefit has always been strongly on the agenda, an independent review[7] was commissioned as the evidence base for the next phase of the Open Data program and published in 2013. It "conservatively" estimated the direct economic benefits of public sector information at around £1.8bn a year, with an overall impact including direct and indirect benefits (such as time saved by access to real-time travel data) of around £6.8bn.

13. In Spain, a study found that the "infomediary" sector (companies that sell services on top of Open Data) has at least 150 companies and employs around 4,000 people and generates 330-550 million Euros annually that can be directly attributed to Open Data reuse[8].

14. Other, sectorial, evidence suggests similar potential benefits. A study[9] by Oxera for Google estimated the Gross Value Added (GVA) by the Geospatial Services Sector as $113 billion a year, 0.2% of global GVA and approximately half the GVA of the global airline industry. Oxera also pointed to additional indirect benefits including $17 billion in time savings, $5 billion in fuel savings and $13 billion in education. They also pointed out that the industry was growing rapidly - at a rate of 30% per annum globally. A study of the importance of the Global Positioning System to the US Economy[10] found that it was difficult to estimate the total economic value of GPS because it had become so pervasive and because it was still increasing at a high rate. However the best estimate as of June 2011 was that there were $96 billion in annual direct economic benefits attributable to GPS and 3.3 million US jobs were dependent on it: the agriculture sector alone had productivity and cost-saving gains of $19.9 billion a year due to GPS. In the Netherlands, the geo-sector accounted for 15,000 full time employees in 2008[11]. Another study reported that the weather derivatives market, driven by the free availability of weather data in the US, had grown from $4.6 billion in 2004 to a value of $19.2 billion in 2007[12].

---

[7] Shakespeare review: an independent review of public sector information
https://www.gov.uk/government/publications/shakespeare-review-of-public-sector-information

[8] Spanish Open Data Portal Annual Report, "Characterization Study of the Infomediary Sector", July 2012 via Cap-Gemini http://www.capgemini.com/resource-file-access/resource/pdf/the_open_data_economy_unlocking_economic_value_by_opening_government_and_public_data.pdf

[9] http://www.oxera.com/Oxera/media/Oxera/downloads/reports/What-is-the-economic-impact-of-Geo-services_1.pdf

[10] http://www.gpsalliance.org/docs/Economic_Study_Backgrounder.pdf

[11] http://www.capgemini.com/resource-file-access/resource/pdf/the_open_data_economy_unlocking_economic_value_by_opening_government_and_public_data.pdf

[12] http://usatoday30.usatoday.com/weather/forecast/2008-06-09-weather-derivative_N.htm

15. These studies do not converge on a single, consistent, measure of the economic potential of Open Data. However all studies and all the analytical approaches that they have adopted appear to point in the same direction. It seems safe to conclude that the economic potential of Open Data is significant - the remaining issue is whether it is just significant or extremely significant.

Unique Characteristics and Challenges

16. The studies also highlight a number of other characteristics of the economic potential.

17. First, a common finding is that although the "direct" economic benefit of the business providing data-rich services is significant, the majority of the overall benefit is "indirect" in the sense that it accrues to the users of data-rich services, in both the business and the personal sectors, rather than the providers of those services. Many data-rich services are themselves delivered over the internet and so have very low marginal costs of distribution and customer management so, freed of upstream data charges or restrictions, end-user prices can be driven down so as to reach a very large market. This in turn increases the consumer surplus.

18. Second, there is some evidence[13] that open data attracts new <u>types</u> of re-users, in particular SMEs, and new business models such as advertiser-pays rather than end-user-pays. For instance as a consequence of the price cuts by the Dutch Meteorological Office KNMI, a new SME re-user launched an innovative service which allows anyone to find out whether it is going to rain in the current location over the next few hours. It generated around 300 million hits per year throughout Europe in 2010. This service is provided completely free of charge to end-users, and is paid for through advertising revenues.

19. Third, it is also apparent that the value is not in the data by itself. It is the combination of the data with one or more other factors such as an innovative idea, the inadequacy of existing services, the availability of new techniques to process data (such as "big data analytics" methods) or new technologies for the delivery services, including the use of smartphones or even the "internet of things" (for instance in-car navigation systems).

20. Fourth, it is increasingly difficult to attribute the total benefits to the individual factors. There is a blurring between "Open Data" and "Big Data", not least because much Government Open Data is inherently also Government Big Data due to the size of government operations and of the population and economy they serve and regulate. McKinsey admit that the $3 trillion figure encompasses total benefits. However they rightly argue that it is unlikely that those benefits would occur at all in

---

[13] For example the POPSIS study by Deloitte

the absence of freely available data. Any attribution of benefits to the various inputs - data, technology, innovation - would inevitably be arbitrary.

21. Finally, the measurement of benefits from Open Data is still imprecise. In a recent seminar hosted by McKinsey and the UK Advisory Panel on Public Sector Information a range of holistic and sectorial studies were reviewed and some common difficulties noted, including the difficulty already discussed of separating the value of the data from the value of the innovation.

22. One other difficulty was that in many cases benefits took time to emerge; innovators and entrepreneurs were more likely to act as a result of spotting a market opportunity, and then seek the data needed to build a service to address it, rather than observe the availability of data and try and invent ways of using it immediately. Moreover even where entrepreneurs acted quickly it took some time for products and services to come to market and to be widely adopted. Simple applications for presenting essentially the original information in more attractive and accessible ways can develop quickly, as has been seen in the number of public transport applications in cities such as New York (68 for the subway alone), London (500 in total) and St Petersburg (30 for buses); more analytical data services, particularly where the data is enriched, can take longer to design, develop and market widely.

23. Another difficulty is that it can be harder to measure the indirect benefits to consumers than to measure the direct economic activity of data-rich service providers themselves. This is unfortunate since the majority of the benefits are usually in those indirect benefits. However in some sectors progress has been made on measuring indirect benefits. For instance in the transport sector the existing approach to cost-benefit analysis usually measures the value of <u>any</u> investment - for instance in a new road or a new public transit scheme - principally by the benefit to end users in terms of the economic value of time saved. The transport sector has developed robust methodologies to estimate this.

24. A final difficulty is that it has so far proved hard to measure the public sector benefits of open data. Like other business consumers, public institutions are purchasers of data-rich services. Indeed, in some cases they buy <u>back</u> their own data after it has been aggregated or enriched by data-rich service providers. The amount of use of Open Data within government has been one of the unexpected and surprising observations of the last five years: for instance, one third of the data downloads from the Open Data portal of the province of British Columbia in Canada have been observed to be coming from the province's own internet addresses; and in the Catalonia Region of Spain the cost savings and efficiencies to public institutions themselves of open metadata on geospatial datasets mandated by the EU INSPIRE Directive recovered four years of development costs in just six months[14].

---

[14] http://inspire.ec.europa.eu/reports/Study_reports/catalonia_impact_study_report.pdf

What data is most important for economic growth

25. There is also strengthening evidence about some of the types of data which are most important for economic growth.

26. The highly pervasive and leveraged nature of geospatial reference data in the economy is already well recognised. The data from the Ordnance Survey, the UK's National Mapping Agency, underpins around £100 <u>billion</u> a year of economic activity for a production cost of around £100 <u>million</u> a year[15]. The single Address Register for Denmark, available as Open Data, has an annual return of economic benefit of 70 times its maintenance cost[16].

27. Other government-held information is also widely used - weather data, road and other transport data, official registers including company registers and cadastres, and so on. Together with geospatial data these provide core reference data for the economy as a whole - both through the utility of the data itself and through the role of the government providing basic, common, reference data including identifiers to which other actors can add their own, possibly private, data. For instance access to comprehensive and up to date company data is important to credit reference agencies, and so to the financing of businesses. This is sometimes called the National Information Infrastructure.

28. Both the GovLab Open Data 500 and the Deloitte-Open Data Institute work is giving evidence on which datasets are most important for companies building services with Open Data. The GovLab study shows particular importance in the US to data from the Department of Commerce (used by 120 out of the 500 companies in the Open 500), the Department of Health and Human Services (75) and the Security and Exchange Commission. In the UK six categories of data - geospatial, environment, economic, transport, energy & resources, and demographic are used in at least 10 of the 20 defined sectors of the economy.

Open Data Companies

29. In North America and Western Europe there are a growing number of companies set up to use government open data to provide data-rich services and a growing number of other companies adapting their business models to accommodate open data into their services.

30. In the United States GovLab - the Governance Lab at New York University - have produced an "Open Data 500", a catalogue of 500 US-based companies providing services using Open Data. Of the 500 in the index, more than 180

---

[15]

http://webarchive.nationalarchives.gov.uk/20090323145858/http://ordnancesurvey.co.uk/oswebsite/aboutus/reports/oxera/oxera.pdf

[16] http://www.adresse-info.dk/Portals/2/Benefit/Value_Assessment_Danish_Address_Data_UK_2010-07-07b.pdf

responded to the GovLab's survey. The responses showed that these were typically new companies, with two thirds of them founded in the last five years and half had more than ten employees.

31. In the United Kingdom an early report from an ongoing study by Deloitte in conjunction with the Open Data Institute[17] reported that it was studying over 230 companies and about 15 distinct business models had been identified.

32. While many of the companies using Open Data to provide services are (for the moment) small, there are already a number of companies which have grown quickly and have significant valuations, including:

* Zillow with a market capitalisation of over $3 billion, provides a home and real estate on-line marketplace to help homeowners, home buyers, sellers, renters, real estate agents, mortgage professionals, landlords and property managers find and share vital information about homes, real estate, mortgages and home improvement. It is driven by a database of more than 110 million U.S. homes – including homes for sale, homes for rent and homes not currently on the market.

* Zoopla, a UK-based analogue of Zillow using house sales data from the UK Land Registry, has annual sales of £76m and profit of £25m. 86% of its sales come in subscriptions from over 19,000 estate agents. It is expected to float in summer 2014 with a total value estimated between £1.0bn and £1.3bn[18].

* Waze, acquired for a reported $1.3 billion by Google in June 2013, is a GPS-based geographical navigation application for smartphones with GPS support and display screens which provides turn-by-turn information and user-submitted travel times and route details, including crowd-sourced information. It won the Best Overall Mobile App award at the 2013 Mobile World Congress.

* Climate Corp, acquired for $930 million by Monsanto in October 2013, uses 60 years of detailed crop yield data, weather observations from one million locations in the United States and 14 terabytes of soil quality data - all free from the US Government - to provide applications that help farmers improve their profits by making better informed operating and financing decisions. A key product is "Total Weather Insurance", an insurance offering that pays farmers automatically and without proof of loss for bad weather that may impact their profits.

---

[17] http://www.deloitte.com/view/en_GB/uk/market-insights/deloitte-analytics/bfb570a79416b310VgnVCM1000003256f70aRCRD.htm

[18] http://www.ft.com/cms/s/0/eb5a8b76-e17d-11e3-9999-00144feabdc0.html?siteedition=uk#axzz344i4HNjO

> Founded by two ex-Google engineers in 2006, Climate Corp uses three
> million new data points a day from 22 datasets using advanced
> analysis techniques. The data comes from a range of third-party
> providers such as the US National Weather Service, which publishes
> its data free for re-use.

33. Venture capital has followed.   A recent estimate[19] puts VC investment in open data-related areas in the first half 2013 alone at $1.28 billion across 127 deals. There is already some exit activity too: the same period saw a total of 22 M&A transactions and 3 IPOs.

Types of Businesses

34. The range of companies now actively using Open Data has also given some insight into typical business models.   The examples above give some specific examples, and the Deloitte-Open Data Institute work[20] has identified about 15 distinct models which it clustered into five main archetypes:

Suppliers

35. These are organisations that publish their data as Open Data to allow others to use and reuse it. They include not only public sector bodies but also some private sector companies, such as the Association of Train Operating Companies in the UK (where most of the railway operators are in the private sector). Revenues from the publication of data are nil, but the data may lead to increased levels of customer engagement and loyalty, with revenue gains. There is also the opportunity to sell value-added services based on the supplier's intimate understanding of the data: for instance after Statistics Germany started to publish its data free of charge in October 2008 the income from premium and advisory services <u>rose</u> by 95% in two years. Some data suppliers, such as the UK Weather Bureau ("The Met Office"), have two-tier models, where data is available for free on a "best endeavours" basis and where the <u>same</u> data on the <u>same</u> licence terms is available with a guaranteed level of availability for a service fee.

Aggregators

36. These are organisations that collect and aggregate open data and, sometimes, other proprietary data. This is typically on a sectorial or geographic theme. Examples include: Data Publica, which is a one-stop shop for data about France; Placr/Transport API which aims to be the UK first comprehensive open platform for transport solutions; and Open Corporates, which holds data on 66,665,508

---

[19] http://www.cbinsights.com/blog/trends/big-data-funding-venture-capital-2013

[20] http://www.deloitte.com/view/en_GB/uk/market-insights/deloitte-analytics/bfb570a79416b310VgnVCM1000003256f70aRCRD.htm

companies taken from 81 national or state company registers. Business models range from remuneration for aggregation itself through added-value data access services such as APIs to tiered models (for instance free on a "share-alike" basis and charged-for if re-sold with closed additional data).

Developers

37. These are organisations and individual developers who design, build and sell web or smartphone applications to deliver government open data to customers (normally in the personal sector) in attractive and informative ways, sometimes in competition with "official" applications. There are over 68 such applications for the New York City subway system and around 500 applications using Transport for London's 25 Open Data feeds (employing around 5000 people in development, marketing and support).

Enrichers

38. These are organisations which use open data to gain new or better insights that they can deliver in services or products to their customers - often completely new services which could not exist before Open Data. This is Climate Corp's business model. It is also largely that of Zillow in the US and Zoopla in the UK. Enrichers are probably the highest value-added type of business, and this seems to be recognised in valuations.

Enablers

39. These are organisations which provide platform and technologies that other businesses and individuals use. They are a vital part of the Open Data "ecosystem" - while being revenue generating themselves, they also provide cost-effective and easy-to-access services for both data suppliers and data consumers. Examples include Socrata, which provides a Software-As-A-Service Open Data portal for governments and municipalities (and which received $25m of venture capital funding in 2008 and 2013) and its more recent France-based competitor OpenDataSoft[21], and MuSigma and Cloudera which provide data analytics services and platforms.

What about the ICT sector?

40. No doubt some companies in the "traditional" ICT sector of hardware supply, software development and systems integration are providing services to companies working to create data-rich services. Some are also providing essentially enabling services, such as the Microsoft Azure Datamarket[22] (it does not do active aggregation). However the drive to exploit Open Data is coming from outside the traditional ICT sector, driven by entrepreneurs and innovators seeing market

---

[21] http://www.opendatasoft.com/

[22] http://datamarket.azure.com/

opportunities for data-rich services rather than through "pushing" technological solutions. Belatedly some of the traditional ICT sector firms are responding, particular in the area of big data analytics, where they can leverage their existing engagements with large corporations.

41. That is not to say that a strong ICT sector is irrelevant for achieving economic growth and business innovation from Open Data. It is instead that the data companies are bringing together the latest ICT skills, open source software, and increasingly commoditised ICT services such as cloud computing to provide a reliable and cost-effective technological base for their business.

Policy Implications

42. Achieving the economic potential of Open Data is more complex than a government simply putting some data on a website. International experience is that the delivery of the initial promise of Open Data is only the first step on the Open Data road. Some jurisdictions have done nothing more - and have not delivered their objectives. Conversely even the most successful initiatives have found that sustaining Open Data has been a hard road to follow; and even now in the first countries to work on Open Data the full requirements of successful and sustainable use of Open Data and its supporting ecosystem are still being discovered. The "gestation period" of the leading international examples quoted above has been several years from the first release of data, and has required continued focus and support from the governments concerned. Therefore supplementing the provision of data with more engagement and participatory approaches is a significant key to realizing the potential of open data.

43. In doing so, governments need to recognise that they will have four distinct but related roles:

*   Supplier Governments need to release the data they hold which is needed for economic growth and business innovation, to do so publicly and regularly, and to steadily improve quality and access.

*   Leader In addition to the data which governments themselves hold, they will need to provide both policies and active leadership and encouragement to other institutions to release data important to economic growth and business innovation. This includes public institutions at regional and city level, state owned enterprises, and private sector companies providing important public services.

*   Catalyst Governments should serve as catalysts for the use of open data by nurturing a thriving ecosystem of data users, coders, and application developers and incubating new, data-driven businesses.

*   <u>User</u> Open Data programmes often enable government agencies to overcome their own internal barriers to using data. Governments need to promote the use of public data within public institutions at national, regional and city level. This will also mean investment in skills and tools. It should also mean that governments should be leading, and proactive, customers for innovative private sector products and services using open data, including advanced analytic services to improve internal decision making and to help create new services. In addition using one's own data can give greater understanding of how the data could be made more usable and useful.

44. The following specific policies and practices are recommended to do this:

Supplier

(1)   **Release data which businesses and others request and need**

Government's future program of data release need to be driven not only by the knowledge of the officials of different ministries on what could be released but also by a public system by which businesses (and others) can request, discuss and prioritize the data that they want. Making this work effectively requires three supporting steps. First, government institutions need to make publicly visible more details of their overall data holdings - including those datasets which are <u>not</u> yet available as Open Data[23]. Using a national Open Data portal effectively is an important part of this. Second, when data is requested, businesses must get a quick answer. Third, because of the importance to the wider economy, individual ministries should not be permitted to refuse data without a wider and fast review by the government as a whole of the overall arguments for and against data release.

(2)   **Prioritize the release of "core reference data"**

As outlined above, experience suggests that there is a conceptual subset of government data that can be termed "core reference data" or "national information infrastructure". This subset is government data is widely used across the economy to underpin the efficient operation of businesses of all kinds. Typical examples of such data include maps, address databases, demographic data from the Census, data about roads and other transport links, official data about registered companies and other businesses and data about public procurement. In consultation with business, governments should identify the datasets which would comprise the "core reference data" and should then ensure that the release of this data is prioritized.

---

[23] In the United Kingdom data.gov.uk contains metadata not only about 10,700 "Open Data datasets" but also about around 4,000 other geospatial datasets and 4,000 non-geospatial datasets which are not yet open. The publication of metadata about government-held geospatial datasets is required by the EU INSPIRE Directive.

(3) **Ensure that data can be found**

It can be a challenge for potential users to find the data that they need within the structures of government. In addition to division into separate ministries at national level, data also resides at regional and municipal levels. A national Open Data portal can help address this issue if it has a collection of the richer metadata on each dataset needed to assist locatability. It will also mean developing national policies on others portals and ensuring that they all can collaborate and share their metadata[24]. It also means making the data more visible on third party search services such as Google, Yahoo and Bing.

(4) **Ensure continuity of supply of data**

To give businesses the confidence to invest, the government will need to give continuing assurances that the data that businesses use will continue to be released, maintained and updated regularly. This will give business users the reassurance that they have continuity and that their investment is protected. It will also be important to ensure that ministries do not unilaterally withdraw data which has previously been published without adequate consultation and notice. Governments should establish policies and actively monitor compliance with them, escalating to ministerial level quickly any failure in data supply which is endangering businesses.

(5) **Release fine-grained and disaggregated data**

The Open Data initially released by many governments is often statistics or aggregate data. This is because such datasets are relatively easy to release - and often they are re-usable and "open" forms of information already available in a closed or printed form, and thus do not raise policy issues. However most of the data useful to business - particularly for the "developer" and "enricher" business types where most of the economic benefit is - is more detailed and less aggregated data. This is because it can be used in the context of individual business transactions or it can be analyzed in different, innovative, ways using big data analytics and other techniques. So it is important to ensure that the right level of detailed data is released while protecting national security and personal privacy. Each public institution should have a concrete plan for releasing specific datasets in a more disaggregated form.

---

[24] The EU have created a prototype pan-European data portal harvesting metadata from national and other portals http://publicdata.eu/ - see http://ec.europa.eu/digital-agenda/en/open-data-portals

(6) **Extend the release of data beyond government ministries**

Some of the most valuable and sought-after data may not be owned by a government's ministries themselves, but by state owned enterprises, private operators of public services, academic institutions or by publicly-funded researchers. This would, for instance, include data about fares, routes, timetables and real-time running of transport services and scientific data generated as part of government-funded research. It is therefore important that national ministries see their role not only as opening up their own data but also as a leader and influencer to their sector, encouraging and supporting regions, municipalities, state owned enterprises and private operators to open their data as well. Each ministry should have a broader strategy for Open Data in its sector but which it does not itself control.

(7) **Actively participate and promote the use of open data**

Individual ministries should not only give leadership to other data suppliers in their sector; they should also be seen to give leadership to the uses of data in their sector too. Individual ministries should be given a remit and a target to promote the use of Open Data both from the ministry itself and from other data suppliers within the Sector.

Catalyst

(8) **Ensure that Open Data portals are more collaborative and demand-driven.**

Leading governments in Open Data are not only focused on sustaining a supply of high-value data. They are actively encouraging and enabling businesses and citizens to help lead the evolution of their Open Data portal(s). This means seeing the Open Data portal not only as a place where people can find data but also as a place where they can discuss the use of the data, find applications using the data and place links to related datasets or other assets. By doing so governments can help accelerate the development of "communities" of data users, and also build the links between officials responsible for the supply of data and the developers responsible for the innovative use of those data.

(9) **Ensure that government data is properly explained, and that issues can be raised with the relevant expert officials**

Even with good metadata a government dataset can be hard to understand and to use. Developer-activists may be prepared to gain understanding by trial and error over a period of time out of personal interest, but potential

business users may be more easily discouraged. In addition, the lack of supporting information makes business use of data even more difficult because reliable and predictable results are essential. Therefore the government experts responsible for each dataset need to be responsible for engaging with the community of people using their data and to be responsible for supporting and helping people make good use of their data. This should involve not only answering questions and supplying documentation about standards and coding, but also actively reaching out to developers by contributing to online communities (e.g., on the Open Data portal) and through seminars and workshops.

(10) **Reach out not just to developers but to innovators and entrepreneurs in specific sectors**

It is important for governments to see the use of Open Data as an issue of business innovation of all types, and not solely or primarily an issue for the ICT sector. Software developers or ICT service companies are not the sole - or even necessarily the best - source of ideas. The most successful drivers will come from a business problem which the innovator seeks to address. Promotional activities for the use of Open Data can focus not only on Open Data in general but additionally on specific sectors - or indeed on specific issues on which the government would like more innovation. For instance, in the United States, there is an annual "Health Datapalooza" to showcase innovation using health data. The first event, in 2010, had 45 attendees and 17 submitted applications; the third in 2012 had nearly 1,600 attendees and 240 submitted applications; the fourth in 2013 had 2,000 attendees - and would have had more had a cap on number not had to be imposed. Each ministry should have a strategy and a target for generating economic growth and business innovation in its sector from the data it and other public institutions in its sector hold.

(11) **Actively support and incubate innovation using Open Data and create institutional structures to do that on a sustainable basis**

Many governments have tried to support innovation using "hackathons" and competitions. These certainly have their place, particularly in the very early days of an Open Data initiative to generate ideas and some proof of concepts - as well as to help build the community of Open Data developers itself. However the trend is towards establishing something more institutional and sustainable. For instance the UK have created an Open Data Institute to "convene world-class experts to collaborate, incubate, nurture and mentor new ideas, and promote innovation"; this is part-funded by the UK Government's Technology Strategy Board which provides funding for innovation and applied research in partnership with business. In Germany the Fraunhofer Institute for Open Communication Systems is similarly working on Open Data, with the mission to emphasize "the high potential of open data for

businesses and to challenge the view that open data is not only a matter for administration, but also for enterprises, NGOs and science". A 2013 conference funded by the Knight Foundation about Open Government and Open Data in the United States identified the need for an Open Data Institute in the United States, and this is now well on the way to being established[25].

Governments should therefore consider how they could establish similar "centres of excellence" for Open Data to engage and bring together both the data suppliers from within public institutions and the data users in the private sector. Such a centre could specifically provide incubation facilities for start-up businesses in Open Data and have a role to promote national expertise and capability in Open Data and to assist local businesses in competing globally in the supply of data-rich services. It may also provide a sustainable mechanism to provide leadership and management to the programme as a whole and to the other recommendations.

(12) **Leverage existing government support programmes to ensure that they also contribute to Open Data objectives where possible**

Governments should ensure that other sources of innovation investment, ICT industry stimulation and start-up incubation are playing their part in supporting the growth of innovative uses of Open Data and of the supporting ecosystem of ICT and other services. It may be possible to leverage the substantial existing funding in these areas to accelerate the exploitation of Open Data. Once explained, the relatively high potential returns on Open Data and the synergies with other related technologies such as Big Data should be attractive to the managers of the existing funding.

User

(13) **Develop Open Data skills within the government institutions, regions and municipalities**

The re-use of Open Data has opportunities for the efficient and collaborative operations of government itself. Other jurisdictions have found that once data is freely available as Open Data there is a surprising and serendipitous reuse of data within government itself. However, fully exploiting this potential requires the development of "Open Data skills" among the relevant officials. These skills should include a sound understanding of the laws in relation to Open Data including their interpretation and limits, together with more detailed explanation, guidance and clarification; how to extract and prepare data for publication in machine-readable formats and how to generate rich metadata and other supporting materials to help ensure the reuse of the data; how to release datasets at a level of fine granularity and disaggregation without

---

[25] http://usodi.org/

compromising personal privacy (eg anonymization techniques); how to assist businesses and other data users make the best use of the data; and where supporting technical, legal and other resources can be found. Success in Open Data represents a new way of working for officials that will require sustained leadership and change management.

(14) **Ensure that the Government is using data services and products from the private sector.**

Governments should be leading, and proactive, customers for innovative private sector products and services using open data, including advanced analytic services to improve internal decision making and to help create new services. This public procurement demand will help stimulate early investment.

Conclusion

45. While sources differ in their precise estimates of the economic potential of Open Data, all are agreed that it is potentially very large. In countries which were early movers in Open Data there is already evidence of significant businesses having developed to exploit that potential. Leading governments have recognised that their role is not simply to publish data - they are supporting the whole value chain of the use of data through four distinct though interlinked roles. Other governments should consider how to use their Open Data to enhance economic growth, and should put in place strategies to promote and support the use of data in this way.