

The Programmable City



Programmable City

Working Paper

Big Data Series: Critical Analysis of the Irish Big Data Skills Report

Tracey P. Lauriault, Programmable City Project
National Institute for Regional and Spatial Analysis (NIRSA)
National University of Ireland Maynooth
County Kildare, Republic of Ireland

Tracey.Lauriault@NUIM.ie, <http://www.nuim.ie/progcity/>

24 May 2014

1. Introduction to a series of on big data:

My area of research on the Programmable City Project examines *how digital data materially and discursively support and process cities and their citizens*. This work falls within an emerging field called critical data studies. A recent working paper entitled Small Data, Data Infrastructures and Big Data, is an example of critically thinking about data, as is the forthcoming book The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences and my earlier work on data preservation, access, and data infrastructures. The new Big Data and Society journal also provides some theoretical orientation. Dublin and Boston, are the Programmable City's formal study sites and I will add a little bit of Canada in that mix from time to time.

The first part of my research consists of constructing 'data landscapes' for Dublin and Boston, by which I mean mapping out the different organisations involved in producing, analyzing, and selling data and analytic services, of which big data are an important component. A key element is reading a number of reports, papers, and consultations on the topic, as well as conducting interviews and attending industry and academic events, I will share questions and observations here in blog posts. At first, these will mainly be descriptive and comparative. However, this reading combined with empirical case studies will lead to a more critical examination of the data assemblages of big data, open data, indicators, maps and data infrastructures. I will be guided by two theoretical frameworks, Rob Kitchin's 'Data Assemblages' and a modified Ian Hacking Making Up People and Spaces framework.

Attributes	Elements
Systems of thought	Modes of thinking, philosophies, theories, models, ideologies, rationalities, etc.
Forms of knowledge	Research texts, manuals, magazines, websites, experience, word of mouth, chat forums, etc.
Finance	Business models, investment, venture capital, grants, philanthropy, profit, etc.
Political economy	Policy, tax regimes, public and political opinion, ethical considerations, etc.
Govern-mentalities / Legalities	Data standards, file formats, system requirements, protocols, regulations, laws, licensing, intellectual property regimes, etc.
Materialities & infrastructures	Paper/pens, computers, digital devices, sensors, scanners, databases, networks, servers, etc.
Practices	Techniques, ways of doing, learned behaviours, scientific conventions, etc.
Organisations & institutions	Archives, corporations, consultants, manufacturers, retailers, government agencies, universities, conferences, clubs and societies, committees and boards, communities of practice, etc.
Subjectivities & communities	Of data producers, curators, managers, analysts, scientists, politicians, users, citizens, etc.
Places	Labs, offices, field sites, data centres, server farms, business parks, etc, and their agglomerations
Marketplace	For data, its derivatives (e.g., text, tables, graphs, maps), analysts, analytic software, interpretations, etc.

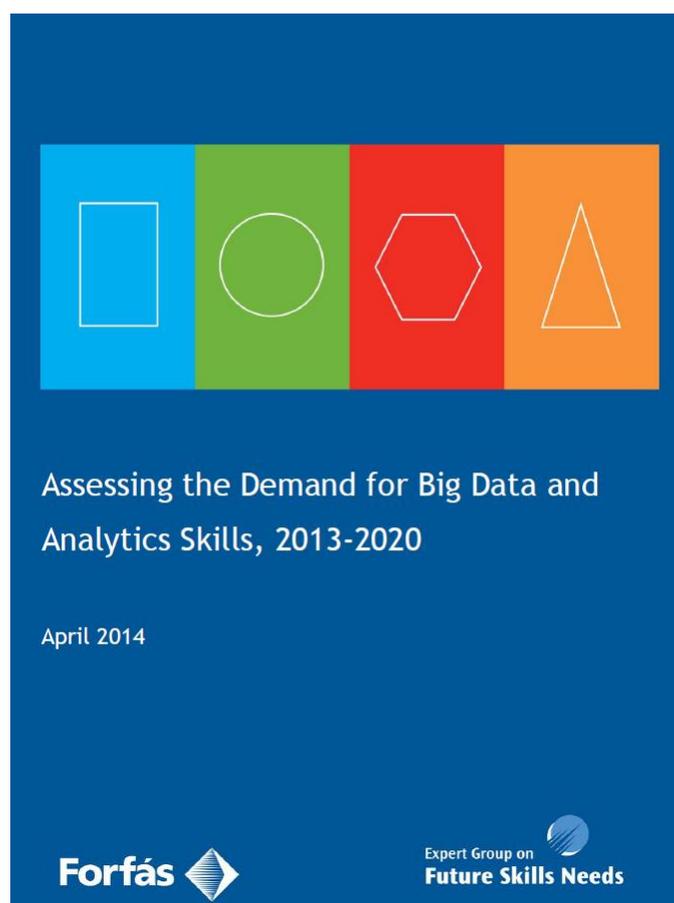
Table 1 Source: Kitchin Data Assemblages Framework, The Data Revolution (2014)

2. A Critical Reading of Big Data Reports

This first post concerns the newly published Republic of Ireland report entitled **Assessing the Demand for Big Data and Analytics Skills, 2013 – 2020** (herein referred to as the Big Data Skills Report). I will follow-up in later posts with a critical reading of The White House Big Data Seizing the Opportunities, Preserving Values report, MassTech’s Massachusetts Big Data Report as well as MassTLC’s Big Data & Analytics: A Major Market Opportunity for Massachusetts report. In addition, I will share thoughts on the Canadian Internet Public Policy Interest Clinic (CIPPIC) study of the data brokerage industry entitled On the Data Trail, The Canadian Standing Committee on Access to Information, Privacy and Ethics Privacy report on Social Media in the Age of Big Data, and Privacy and Big Data: The Players, Regulators, and Stakeholders by O’Reilly Media.

3. Backgrounded on the Big Data Skills Report:

The *Assessing the Demand for Big Data and Analytics Skills, 2013 – 2020* report was published by Forfas and the Expert Group on Future Skills Needs (EGFSN). Forfas is the Republic of Ireland’s policy advisory board for enterprise, trade, science, technology and innovation while the EFSG reports to the Minister for Education and Skills and the Minister for Jobs, Enterprise and Innovation.



The report was commissioned “to advance recommendations on measures to build up the big data and analytics talent pool through domestic graduate output, continuing professional development within industry, and, where necessary, attraction of talent from abroad including expatriate talent”. The Government’s ambition is for the Republic to become a leading country in the field of big data in Europe, and the report is designed to guide the creation of an Irish big data ecosystem based on public and private collaboration.

The research for this study was driven by the Joint Industry/Government task force on Big Data which was formed in June 2013. Big data is also considered to be a disruptive reform in the Action Plan for Jobs 2013, which is a reform considered to have the “potential to have a significant impact on job creation, to support enterprises or where Ireland can profit from a natural advantage or opportunity that presents itself in the economy” and a research priority Data Analytics, Management, Security

and Privacy 2013. Forfas commissioned the research in a Request for Tender for the Provision of Research for the Study Assessing the Demand for Big Data / Data Analytics Skills in Ireland 2013-2020 (LINK TO PDF) in July 2013.

Ernst & Young (E & Y) and Oxford Economics were jointly commissioned to blend their different fields of expertise to conduct the international trend analysis, undertake the consultations with companies and key informants and to model big data and analytics skills demand scenarios. Their work was overseen by a 17 member Steering Group chaired by the EGFSN.

The Big Data Skills Report was overseen by a Joint Industry/Government task force on Big Data and a Steering Group with representation primarily from the private sector (IBM, Accenture, Glanbia, CPL, Datahug, Nanthean Technologies, Deloitte, Vidiro Analytics, Enterprise Ireland, IDA Ireland and the Quinn School of Business) and 4 government offices (Revenue Commissioners, Department of Education and Skills, Higher Education Authority and Forfás).

3.1. Provenance:



The report is the outcome of a number of fast tracked high level strategies for the Government of Ireland as follows:

1. The Department of Jobs, Enterprise and Innovation (DJEI) considers big data to be Priority Area B in the Research Prioritization Steering Group (Mar. 2012) report after Priority Area A which is Future Networks and Communications.
2. This is followed by the DJEI's Action Plan for Jobs 2013 (Feb. 2013) that repositions big data as Disruptive Reform 1 which is to *"Build on our existing enterprise strengths to make Ireland a leading country in Europe in 'Big Data'"*.
3. Meanwhile outside of government SAS UK and Ireland commissions the Centre for Economics and Business Research (Cebr) to assess the big data sector and releases Data equity – Ireland Unlocking the value of big data (June 2013) coining the term 'data equity' from the idea of brand equity. Companies will be forming data equity with the power of analytics.
4. The Joint Industry/Government task force on Big Data was formed in June 2013 to *"drive the development of this high-growth sector in Ireland"*.

5. Big data is then considered to be a *Global Technology and Service Trends Influencing Irish ICT Skills Demand* in the Forfas Expert Group on Future Skill Needs (EGFSN) report Addressing Future Demand for High-Level ICT Skills.(Nov. 2013).
6. As part of the actions plan, Forfas launches a Request for Tender for The Provision of Research for the Study Assessing the Demand for Big Data / Data Analytics Skills in Ireland 2013-2020 which was awarded to E & Y and Oxford Economics.
7. Big data is positioned again as disruptive and integrated with ICT skills, it is a “*disruptive growth and innovation phase. This includes the adoption of cloud computing, the penetration of mobile devices and technologies and the Internet of things, the emergence of Big Data analytics, IT security, micro- and nanoelectronics and the adoption of social technologies in both the personal and business environment*” in the EGFSN ICT Skills Action Plan (Mar. 2014) also an outcome of the Action Plan for Jobs 2013.
8. In April of 2014, the Assessing the Demand for Big Data and Analytics Skills, 2013 – 2020(Big Data Skills Report) is released.

4. How are big data analytical skills categorized in the Big Data Skills Report?

The objectives of the Report were to forecast the future annual demand across the economy between 2013 – 2020, for big data & data analytics and related roles which included a mapping of qualifications, skillsets and competency requirements. Skill categorization was pre-determined in the Request for Tender as follows:

1. Deep Analytical Talent skills requiring a combination of

- advanced statistical, analytical, and machine learning skills,
- business skills to assess the meaning of data and to derive business insights &
- communications skills to explain/persuade other executives.

These roles can originate from a range of backgrounds such as mathematicians, statisticians, actuaries, operational research analysts, economists and include the newly termed role of the ‘data scientist’.

2. Big Data Savvy; comprising of “data savvy”

- managers,
- CIO’s,
- market research analysts,
- business and functional managers.

These professionals require an understanding of the value and use of big data & data analytics to enable them to interpret and utilise the insights from the data and take appropriate decisions to advance their company strategy and drive business performance.

3. **Supporting technology roles;** these personnel have the skills to develop, implement and maintain the hardware and software tools required to make use of Big Data/Data Analytics software and hardware. The core technologies of those implementing Big Data solutions tend to be focused Hadoop and a growing range of SQL databases. These roles include:

- programmers and software development professionals and
- IT business analysts, architects & system designers.

These 3 categories were originally created by the McKinsey Global Institute (MGI) in their 2011 Big data: The next frontier for innovation, competition, and productivity report. The original MGI categorization is as follows. Note that these categories are restricted to the mention of ‘hard skills’ types of occupations. Therefore big data is considered to be primarily a technical field.

	Deep Analytical Talent	Big Data Savvy	Supporting Technology
Definitions	People with advanced training in statistics and/or machine learning and who conduct data analysis	People who have basic knowledge of statistics and/or machine learning and define key questions data can answer	People who service as database administrators and programmers
Occupations	<ul style="list-style-type: none"> ▪ Actuaries ▪ Mathematicians ▪ Operations research analysts ▪ Mathematical technicians ▪ Mathematical scientists ▪ Industrial engineers ▪ Epidemiologists ▪ Economists 	<ul style="list-style-type: none"> ▪ Business and functional managers ▪ Budget, credit and financial analysts ▪ Engineers ▪ Life scientists ▪ Market research analysts ▪ Survey researchers ▪ Industrial organisation psychologists ▪ Sociologists 	<ul style="list-style-type: none"> ▪ Computer and information scientists ▪ Computer programmers ▪ Computer software engineers for applications ▪ Computer software engineers for system software ▪ Computer system analysts ▪ Database administrators

Source: MGI

Table 2: Role Categories and Occupation Types (p.38)

E & Y expanded the MGI categorization in the Big Data Skills Report as seen in the skills, competences and qualifications requirements in the table below. In this case, under the ‘Big Data and Analytics Savvy’ category, some public policy, regulatory and governance roles ‘hard skills’ such as data protection and IP Knowledge are taken into consideration, while Ethics is considered to be a ‘soft skill’ for both ‘Deep Analytical Talent’ and ‘Big Data Savvy’.

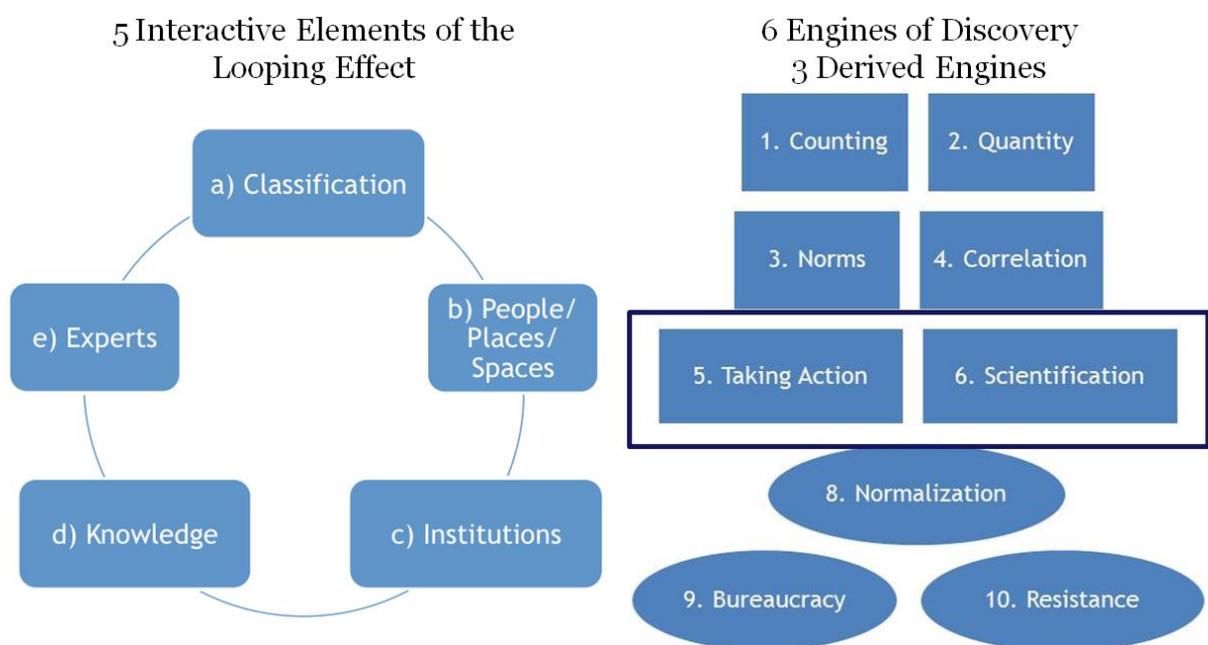
	Deep Analytical Talent	Big Data and Analytics Savvy	Supporting Technology
Technical and analytical	<ul style="list-style-type: none"> Advanced mathematical, statistical and analytical ability NoSQL Data visualisation Web analytics R and other statistical analysis packages Data mining Social media analytics Ability to deal with structured and semi/unstructured information. 	<ul style="list-style-type: none"> Conceptual knowledge, quantitative and analytical skills Data protection, governance, and IP knowledge Enterprise Data management Specific user tools (i.e. dashboards/KPI data/market analysis) 	<ul style="list-style-type: none"> Various programming languages, data related applications and processes Hadoop, Java , C++ , Oracle Building, implementing and managing Hadoop environments Mapreduce Database management and administration - SQL, MySQL, NoSQL
Business and soft skills	<ul style="list-style-type: none"> Communications skills Business acumen Problem solving, creativity , ethics and teamworking skills Domain and sectoral knowledge 	<ul style="list-style-type: none"> Detailed domain and sectoral knowledge Knowledge of social media Business Strategy Project Management Business intelligence Ethics 	<ul style="list-style-type: none"> Communications. problem solving, ethics and teamworking skills Domain and sectoral knowledge Social media technologies Design / user experience skills

Table 3: Types of Skills and Competencies Required Across Categories (p.50).

This is the last we see of these types of ‘soft’ and ‘hard’ skills in the report. Which is rather unfortunate, as leadership and innovation in any field requires good governance, and this is especially the case for big data. Furthermore, once a categorization gets amplified, institutionalized and authorized it becomes real and begins to ‘make up kinds of people’ as per the Ian Hacking’s framework and as illustrated below in Lauriault’s Making up People and Places framework based on Hacking’s work. Classifications describe people and places, institutions begin to use that class, it becomes part of expert knowledge, in this case MGI, E & Y and Oxford Economics. Once set, experts begin to count, quantify, normalize and correlate things, in this case skills, academic qualifications, and employment data. Once these are reported, an institution or groups of experts take some action, in this case they develop a strategy and some action plans. Once actions are taken, the class really becomes true, scientific or real as it becomes a thing. It becomes a normal part of

doing business, a part of a bureaucratic process, in this case, where public sector moneys will be directed toward increasing a set of skills to meet a demand determined by the data analytics private sector in Ireland. In terms of resistance, it may be, that with time we will see those labelled as part of the ‘Deep Analytical Talent’ pool may want to be called data scientists, the ‘Big Data Savvy’ group may want to be big data governors while the supporting Technology Group may want to be considered as big data infrastructure providers.

In this case the DJEI in its Request for Tender has firmed up these categories by making them the focus of the research, and these have now become real categories in the Republic by E & Y and Oxford Economics who structured their analysis accordingly. It is expected that these will appear in many future analyses.



Tracey P. Lauriault, 2012

Figure 1. Hacking Framework

5. Expert knowledge informing the skills demand model in the Big Data Skills Report:

The development of baseline estimates and projections for big data employment in Ireland were informed by the following reports:

- McKinsey Global Institute (MGI) (2011) *Big data: The next frontier for innovation, competition, and productivity* (focusing in the main on the US).
- Accenture Institute for High Performance (2013) *Crunch Time: How to overcome the looming global analytics talent mismatch* (focusing on key sectors in the US, UK, Singapore, Japan, Brazil, India and China).
- e-skills UK (SAS) (2013) *Big Data Analytics – Adoption and Employment Trends, 2012-2017*.

- Council for Economics and Business Research (Cebr) (2013) *Data equity – Ireland. Unlocking the Value of big data* (applying international findings to Ireland).

In addition, the **Big Data Skills Report** adopts the most common definition of big data developed by Gartner in 2001, namely a field that works with datasets that are big in volume, velocity and variety, or the 3Vs. The definitions of big data have however evolved and below is a comparative table developed by Kitchin (2014). The Big Data Skills Report does not however discuss data.

- huge in *volume*, consisting of terrabytes or petabytes of data;
- high in *velocity*, being created in or near real-time;
- diverse in *variety*, being structured and unstructured in nature, and often temporally and spatially referenced;
- *exhaustive* in scope, striving to capture entire populations or systems (n=all) within a given domain such as a nation state or a platform such as Twitter users, or at least much larger sample sizes than would be employed in traditional, small data studies;
- fine-grained in *resolution*, aiming to be as detailed as possible, and uniquely *indexical* in identification;
- *relational* in nature, containing common fields that enable the conjoining of different data sets;
- *flexible*, holding the traits of extensionality (can add new fields easily) and *scalability* (can expand in size rapidly).

(Boyd and Crawford 2012; Dodge and Kitchin 2005; Marz and Warren 2012; Mayer-Schonberger and Cukier 2013).

The definition of big data, the Big Data Skills Report, skills demand models, skill categories, resulting scenarios and recommendations were all authored and created by private sector entities. Concurrently, the Steering Group overseeing this study mostly represents private sector interests. And the direction of the strategy falls under the Joint Industry/Government task force on Big Data. It can then be concluded that those who stand the most to gain from the recommendations made in the report also authored and oversaw them. In other words, there is no critical assessment of the findings, and it would seem that no one is looking out for the public interest.

A more objective approach would have included broader membership of the Steering Group and the EGFSN by including academic experts from the social sciences, information studies, communications or media studies, members of legal and financial professions, ethicists, public sector big data producers or independent think tanks. Also, what if the process was overseen by the ‘Joint Industry / Government **Public Interest** task force on Big Data’? Had that been the case, there may have been a focus on some of the more pernicious effects of big data & data analytics, such as privacy, procurement, security and ethics. It would have been preferable to first have a national big data research and development strategy for the Republic of Ireland with a focus on how big data can help resolve national issues

(e.g., environment, housing, financial, energy, agriculture), identify where expertise in those areas lies or is lacking, and then direct resources toward the resolution of those. In addition, a national strategy would have included an inventory of all the private sector entities involved in big data on the Island and a survey of the skills sets in house and sought to develop a skills demand baseline from that, combined with an inventory of skills in the public and academic sectors. Finally, an analysis of the public policy issues related to big data, including threats to the public interest would have been assessed, and along with a big data private sector, R & D strategy, there would have been a big data governance process in place that would include a regulatory framework and some checks and balances to guide the sector toward ethical conduct.

6. What did the Big Data Skills Report do?

6.1. The drivers for a big data were defined.

Big data is a disruptive reform in the *Action Plan for Jobs 2013* related to a growing industrial sector and subsequently job growth in the country, big data is also considered to be a sector from which significant value can be derived and therefore worth investing in. Interestingly, open government and open data are considered as drivers. This is unusual since most data found in open data portals are characterized as small data. It is true that the public sector and publicly funded research produce and analyze datasets that would be characterized as big data, but these have been being produced for quite some time, are quite specialized and have been disseminated under a PSI Licence prior to the advent of open data. Open data as a driver may be attributed to the fact that the private sector can better commercialize public sector data, both big and small, by combining these with theirs or using public sector data as a baseline. All sectors stand to gain if public sector data are coordinated, centrally discoverable, well described in metadata, free and under an open licence.

The new Post Code system to be deployed in Ireland by 2015 was considered to be a “late mover adoption” in the Report. Paradoxically, this will not be an open data framework dataset available to the public as it is being produced and licenced under a cost recovery program. The “Postcode Management Licence Holder” (PMLH) contract is to Capita Business Support Services Ireland (supported by BearingPoint and Autoaddress) and the PMLH contract will be fulfilled by the Capita-owned business, Eircode. The new post code will be particularly helpful to big data companies, marketers and direct mailing companies wishing to compete with the postal service, but will not necessarily be useful for civil society organizations as it will be too costly. The average citizen may also be paying for these data multiple times:

- the public paid for the creation of the dataset in the first place,
- it will become a cost included in the price of products,

- as part of taxation because government agencies will have to acquire these under a cost recovery regime (e.g., each local authority, all national departments, and etc.)
- tuition fees as academic institutions will also have to pay for this dataset.

There are a number of other important framework datasets not mentioned in the report, most notably the small area file produced by the National Centre for Geocomputation at NUIM developed for the CSO. This allowed for the first time, in the 2011 census, fine grained analysis of neighbourhoods in Ireland and would be of benefit to the big data sector as data big and small need to be aggregated into common and small geographies. Along with open data, and framework datasets, the report did not discuss the formation of a spatial data infrastructure which is how many other nations deal with national and sub-national geospatial datasets such as road network files, small area files, post code files, demographic base files, environmental and natural resources data to name a few. SDIs are the means by which datasets of this nature, are standardized, made interoperable and accessible. This would also be of assistance to big data interests, but also for social, economic and environmental management for the country. These are more robust than open data initiatives as they are based on interoperable policies, standards and technologies that are adopted enterprise wide. The Canadian Geospatial Data Infrastructure for instance is now becoming a platform and open data will be integrated into it while departments will be geo-enabling their administrative datasets accordingly.

6.2. Baseline employment demand was determined

The MGI, Accenture, Cebr and SAS reports mentioned earlier informed the skills demand model. The following explains how it was done:

- E & Y and Oxford Economics determined that reports consistently estimated that existing employment in this sector consisted of 1.5% to 2% of total employment, although categorizations differed.
- Eurostat data were used to compare shares of employment in Ireland and the UK by sector and CSO Data from the Quarterly National Household Survey (QNHS) were used (See figure 3.2 p. 42).
- Based on estimates and proportions for the US and UK, the **Deep Analytical Talent** employment demand for Ireland was estimated.
- Also, the proportion of those in established roles was considered to be more than half of that demand while the remainder are emerging analytical roles.
- The **Big Data Savvy** demand was determined by applying the proportion of total employment in this category in the US to the Irish employment data.
- The **Supporting Technology** demand was derived from structured interviews with the 45 enterprises and organizations in Ireland. A ratio of 1:4 was determined of **Deep Analytical Talent** to **Supporting Technology** professionals.
- To understand the sectoral distribution, the analysis in the 2011 MGI report was used. In the MGI analysis low, medium and high data intensity based on data storage capital stock per firm groups were devised. CSO data were used

here to “understand the data intensity of industries as the capital stock in computer software per employee” (p.44).

Below is the table of the Results:

Category	Employment	% Total Employment
Deep analytical talent	3,300	0.18
Of which <i>emerging analytics roles</i>	1,500	0.08
<i>established analytical roles</i>	1,800	0.10
Big data savvy	25,780	1.38
Supporting technology professionals	6,000	0.32
Total	35,080	1.88

Table 4: Estimate of Baseline Demand in Big Data and Analytics in Ireland (p.45).

The above was the extent of the methodological information provided explaining how baseline demand numbers were derived. No metadata, tabular data or dates were provided, nor details about which CSO and Eurostat data were used, nor was there any analysis provided on how the classes were adjusted to the 3 categories of skills defined in this Big Data Skills Report. Ireland is not that big a country and a survey of all the companies doing big data analytical work could have yielded a much more robust baseline. Estimates of growth were based on the UK and the US. The assumption that the composition of the firms here differs as a result of the favourable tax regime and that Ireland is a much smaller country suggest that those estimates would be an overstatement for Ireland.

6.3. Structured consultations were conducted

A Structured Interview Survey about the employment levels across each of the three categories was conducted by telephone with 45 Forfás selected Irish based enterprises and organisations that employ big data & data analytics talent. This included 35 companies (foreign owned and indigenous), 7 government bodies, 2 data analytic research centres and 10 key informants. Some workshops were also held.

Demographics, vacancies, the difficulty in filling vacancies were determined. In addition future demands, predicted growth, the difficulty in acquiring and keeping talent, sources of future skills, where sources of skill could be sought within the most common disciplines, training opportunities, relationships with academia, turnover, and attrition were estimated. The results of which are presented in the Information box below (p. 59).

Box 4.1 Proposals advanced by survey respondents

Education

A number of proposals were made relating to learning and education. These include that computer skills should be introduced at an earlier stage in the education system and there should be more focus on the teaching of maths at primary level. Regarding the teaching of maths, many respondents advanced the idea that a problem-solving approach should be adopted in second level maths curricula. Greater female participation in analytics should be targeted from an early age.

At third level, cooperation between industry and academia is crucial to ensure that education reflects skills needs. Proposals included involving industry experts in devising curricula and teaching analytics on university programmes. Work placements, both in industry and in research centres, should form an integral part of third level courses for students of big data and analytics and collaboration with industry is vital in this regard. In particular the Canadian University of Waterloo model was noted, whereby students are equipped with 24 months of experience upon graduation. Facilitating students on relevant courses to meet people already working in the area to discuss career options is seen as an effective way to attract them to the field.

As a means of raising standards it was suggested that universities with the strongest records in maths and statistics should come together in order to offer specialised courses. Increased funding should be provided for research to support career options for analytics in academia and to drive innovation. Respondents expressed the view that analytical modules should also feature across less technical disciplines, such as business and social sciences.

Steps should also be taken to allow transition into analytics from other disciplines and up-skilling and retraining into the area should be encouraged. To this end, short courses for individuals with technical backgrounds should be provided and be VAT reduced or exempt.

Promoting big data and data analytics

Companies were of the view that arriving at clear definitions of 'big data' is important to improve the appeal of the sector. This would facilitate the development of analytics-related career paths, making a career in this area more attractive. Recruitment agencies' understanding of analytics should also be improved; business associations and representative groups can play more active roles in fostering the area by holding workshops, seminars and awareness campaigns to demonstrate the potential value of data analytics to firms and as a career path.

In lots of ways this part of the report left more questions than answers. Which companies and institutions were surveyed? How many companies do this type of work in Ireland? What were the questions asked in the survey? How were vacancies and the difficulty to fill positions determined? Did companies classify the skill sets they needed in the same way that MGI, E & Y and Oxford Economics did? If not how were differences rectified? As discussed, the big data categories and classifications did not focus on public policy issues which resulted in the social sciences and law faculties being unmentioned in the lists of university programs, and most notably absent from the report are the pools of both 'Deep Analytical Talent' and 'Big Data Savvy' skills found in geodemographics, geocomputation, geomatics, geophysics as well as the earth sciences. While humanities is mentioned in the report, the digital humanities was not.

This section, like the previous lacked a methodological explanation and the bread crumbs required to assess results. The focus is heavily oriented toward the hard skills, and the recommendations are generally quite sensible. However, of concern is the recommendation that “*third level, cooperation between industry and academia is crucial to ensure that education reflects skills needs. Proposals included involving industry experts in devising curricula and teaching analytics on university programs*” (p.59). This is problematic as universities could simply become instruments to serve corporate interests, and in order to cut costs research may become too instrumentally focussed into a particular sector at the expense of others, and in both cases this may thwart critical reflection in this area for fear of budget losses. Finally there is a concern that this may also lead to an overemphasis on quantitative research and undermine programming in other areas such as qualitative research, and the social sciences.

6.4. Big data future demand scenarios were produced

Three scenarios were offered which included expansion, replacement and skills demand projections as well as up-skilling demand. These were derived by examining the growth projections and demand baselines discussed earlier, the information collected during the consultations and E & Y and Oxford Economics domestic economy sectoral employment forecasts.

The scenarios factored in the following global drivers for big data:

- Increase in the creation and availability of data
- Growing recognition of economic returns from the use of big data
- Analytics driving performance
- Finding and accelerating growth performance
- Managing risk and regulatory pressures
- Exploiting emerging technologies

Global growth projections were based on the 4 reports discussed earlier, a Gartner (2012) [news release](#), and the US Bureau of Labour Statistics (2012) [Occupational Handbook Study of Computer Support Specialists](#).

A few global risks were described, none of which included privacy, security, intellectual property, procurement and ethics, although the fact that along with a lack of relevant data there may also be an abundance of high quality data.

The scenarios for Ireland were based on a number of assumptions, drivers and supporting conditions as seen in the table below:

		Scenario 1	Scenario 2	Scenario 3
Drivers	Wider economy baseline sector projections	■	■	■
	Increase in existing firm data intensity to forecast UK and US levels		■	
	Marked increase in data intensity across the whole economy, including the public sector and within a broad range of private firms			■
	Expansion of existing firms' big data activities		■	■
	New, additional greenfield mobile big data FDI attracted to Ireland			■
	Full implementation of measures relating to Open Government		■	■
	Full implementation of measures set out in the <i>Action Plan for Jobs</i>			■
Supporting conditions	No special effort, public or private, to influence the demand-side or supply-side	■		
	Modest boost in domestic big data talent supply and modest improvement in environment for attracting overseas talent		■	
	Major boost in domestic big data talent supply and improvement in environment for attracting external talent			■
	Removal of barriers to domestic big data growth such as greater access to open data		■	■
	Widespread adoption of data analytics in the public sector			■
	Significant improvement in Ireland's big data international competitiveness offering, especially versus European competitors			■
Outcomes	No change in economy-wide data intensity from today's level	■		
	Decline in Ireland's global big data market share	■		
	Ireland's global big data market share converges with other countries		■	
	Exploitation of domestic big data opportunities amongst existing organisations, public and private			■
	Rise in Ireland's global big data and analytics market share			■
	Ireland becomes a leading country in Europe for big data and analytics			■

Table 5: Summary of Scenario Assumptions, Drivers and Supporting Conditions (p.71)

Resulting in the following predictions as seen in the table below. Scenario 3 would Position Ireland as the leading country in Europe in big data, Scenario 2 would be the forecast if Ireland caught up to other countries such as the UK, while Scenario 1 represents low growth.

	Scenario 1 (low growth)	Scenario 2 (delayed catch-up)	Scenario 3 (a leading country in Europe)
Expansion demand	3,230	9,200	17,930
Deep analytical talent	410	1,050	2,380
<i>of which emerging analytics roles</i>	190	830	2,160
<i>of which established analytical roles</i>	220	220	220
Big data savvy	2,040	4,810	8,450
Supporting technology	780	3,340	7,100
Upkilling demand	2,130	5,160	9,200
Deep analytical talent	20	70	170
Big data savvy	2,050	4,820	8,460
Supporting technology	60	270	570
Replacement demand	11,950	11,950	11,950
Deep analytical talent	1,250	1,250	1,250
<i>of which emerging analytics roles</i>	370	370	370
<i>of which established analytical roles</i>	880	880	880
Big data savvy	9,020	9,020	9,020
Supporting technology	1,680	1,680	1,680
Total demand	17,310	26,310	39,080
Deep analytical talent	1,680	2,370	3,800
<i>of which emerging analytics roles</i>	580	1,270	2,700
<i>of which established analytical roles</i>	1,100	1,100	1,100
Big data savvy	13,110	18,650	25,930
Supporting technology	2,520	5,290	9,350

Source: EY, Oxford Economics

**Table 6: Summary of Future Demand Change Projections, 2013-2020
(p.82)**

It is difficult to say if any of these projections are reasonable as there are no supporting data, data sources, cross classification mapping or algorithms provided. Furthermore, these scenarios are also based on the problematic baseline demand numbers discussed earlier. The implications of Scenario 3 are however enormous. Should Scenario 3 be adopted, significant investment in education and training in quantitative and computational areas would be required in universities, potentially at the expense of non-quantitative fields, such as those related to governance and public interest issues discussed earlier, some fields would be overlooked while inducements would be required to keep existing and new talent here. The drivers for Scenario 3 are in play, and the outcomes are entirely focussed on private sector big data R & D, and not directed at applying big data resources toward addressing issues of national importance such as environment, energy, quality of life, housing or

climate change. If my observations are correct, the strategy is for public sector funding to be directed toward the alleviation of the shortage in skills in this sector, for R & D to answer questions developed by the private sector (INSIGHT and CeADAR), and for inducements to keep skilled personnel here, while also continuing to support a favourable tax regime and FDI inducements. Some Irish citizens may benefit from some of the new jobs that may come on offer, but how Irish Society benefits as a whole and what is in the general public interest in any of these scenarios is uncertain.

6.5. *The big data skills supply was determined*

The MGI 2001 and the Cebr 2013 reports informed the model used to determine the current supply of big data skills. MGI conducted an analysis of US education data in fields renowned for having advanced quantitative training. Ratios were derived of the total graduates and then applied internationally. It is only here do we find the physical sciences and social sciences listed. The Cebr report considers supply and demand but its analysis factored in a different set of academic programs where the physical and social sciences, are not factored in. The analysis in Ireland examined a series of courses and programs considered most relevant to the production of the Deep Analytical Talent cohort such as:

- Dedicated big data & analytics programmes
- Programmes that include significant training/elements in data analytics
- Maths, Statistics and Science
- Engineering programmes
- Physics programmes
- Data analytics programmes in Northern Ireland
- Private data analytics programmes
- Online education in data analytics

This part of the report was more comprehensive as the types of programs, courses and numbers were provided in the analysis and it was estimated that approximately 2000 graduates per year in courses most directly related to emerging deep analytics. Again, graduates related to the 'Big Data Savvy' governance and law types of programs were not focussed on, this is a missed opportunity for Ireland and a very large oversight of this report.

6.6. *Policy measures here and abroad were examined*

Policy was defined loosely to encompass a range of methods to grow the skills supply discussed throughout. A list of international measures to grow talent was discussed, such as post secondary program measures, but also collaboration between industry and higher education institutions that might entail the following:

- Offering 'real world' work problems and large datasets to mine;
- Providing data analytics software and hardware;
- Providing relevant work experience opportunities;

- Shaping specialisms or electives within programmes (including the actual provision of the taught modules);and
- Promoting analytics as a career path for students.

These are relatively straight forward approaches, but I would add the cautionary note of ensuring that students should be able to study on multiple platforms and there should be no exclusive deals or monopolies on platforms and software; that work experience be tied with some sort of fair remuneration, that there be the allowance for critical analysis. The offering of real world data sets from the private sector would also be most exciting. Because of a lack of focus on big data governance issues, no recommendations in this area were made.

There are also many other initiatives already in place in Ireland which form part of the *Action Plans for Jobs* report. Significant investment in research institutes such as INSIGHT which partners with industry to address private sector problems within a number of important themes such a health, economy, journalism and humanities, energy and the environment, sport and wellness, telecom & network and media processing. There is also R & D investment into CeADAR which is another major research institute that partners with the private sector to deliver quantitative and computational skills. There has also been investment in infrastructure such as the Irish Centre for High-End Computing and the Telecommunications Software and Systems Group.

6.7. Recommendations in the Big Data Skills Report

Many of the recommendations in the report were previously commented on. Chapter 8 provides a number of tables with short, medium and long term suggestions. There were however a number of items scattered in this chapter that are new and notable:

- “The overarching recommendation is that a consultative group comprising representatives of industry, academia and relevant agencies should be established” to oversee the implementation of recommendations over a 6 month period.

This is refreshing, and it is hoped here that more than just business school and technical fields from academia will be reflected in the composition of this consultative group and that there will be room for critical debate.

- “For Ireland to become a leading country for big data a higher level of skills supply will also be required. This will involve an enhanced supply of Masters and PhD graduates along with an expansion of places on courses designed for those in advance degrees in STEM to transition into high-end analytics.”
- “Introduce targeted competitive funding available for post graduate specialist analytics programmes to reduce tuition fees, incentivise participation and increase places available.”

Beyond what was previously mentioned earlier in terms the focus on ‘hard skills’, there is currently a hiring cap in universities in Ireland, a wage cap, and a

devaluation of the ‘soft skills’ required to develop the governance and legal issues related to a growing big data industry in Ireland. Careful consideration here is required to ensure that the non STEM departments do not suffer losses in order to build up the ‘hard skills’ demand, and that a set of ‘soft skills’ be developed. The following are some notable examples in other jurisdictions: The Centre for Law Technology and Society, Canadian Internet Public Policy Interest Clinic, The Munk School of Global Affairs, The Citizen Lab, and The Berkman School for Internet and Society.

- “Business communication skills, critical thinking and project management skills should be taught across all STEM disciplines”.
- This is a welcome addition. A bonus would be up-skilling ethics, law and public policy departments toward an understanding of big data issues.
- “Firms should adopt an enterprise-wide approach to managing their analytical capabilities, including the up-skilling of staff for data protection and governance”.

Beyond up-skilling staff, there needs to be a recognition that this sector needs to be regulated, and a number of public policy issues need to be addressed, which implies that resources toward academic disciplines such as political science, public policy, law, communications, contract law, media studies, and social-sciences.

- “...Career guidance in schools should communicate the availability of career opportunities in analytics to students (particularly females) and their parents.”

This is also a refreshing recommendation. Women Invent Tomorrow, Digi Women, Coder Dojo’s and Women in Technology and Science (WITS) in Ireland have done great work in this area, and a number of important studies have been published on the topic. Female STEM experts often complain that they are invited to boy’s schools but not girl’s schools. The Big Data Consultative Group should include experts from these organizations and refer to existing studies to develop this recommendation.

- “Industry and State Agencies should work with the CSO and Revenue Commissioners to explore the possibility of further developing official measurements of big data and analytics employment”.

This is very important, and the recommendation I would make is for the creation of an open and comprehensive business register that would include a profile of the skills in all firms, including big data, combined with an open and accessible inventory of all businesses involved in big data in Ireland. The greatest concerns about this report, is the fact that the numbers are based on very loose definitions, combined with the absence of an inventory of what is currently ongoing in Ireland, along with the imposition of categories that may not apply.

- The public service recommendations are seen in the table below (p.111):

8. Unlock the potential of big data and analytics in the public service

- Government bodies should undertake a review of data sources held and make open as much data as feasible. This would also act to encourage the development of groups around the use of open source data and free technologies and stimulate interest in data analysis.

Time frame: Short to Long-term

Lead: Chief Information Officer, Government Departments

- Increase efforts to build up the data infrastructure including the development of a unique business identifier.

Time frame: Short to Medium-term

Lead: Central Statistics Office

- Enterprises with significant data analytics capacity should be invited to assist Government bodies in carrying out demonstration projects to address specific challenges faced in the public service and illustrate the potential value of big data and analytics.

Time frame: Medium-term

Lead: Chief Information Officer, Government Departments, Companies, Employer Bodies

- Publish examples of successful deployment of analytics in the public services throughout the Government system in order to illustrate the potential benefits.

Time frame: Medium-term

Lead: Chief Information Officer

- Consider tailored recruitment to analytics jobs in the public service including academic positions specialising in teaching analytics skills, and the development of a Government Analytics Service to work on a cross-departmental basis to provide solutions to public service problems. An inventory of analytics skills and applications in the public sector should be compiled.

Time frame: Medium-term

Lead: Government Departments

These are all welcome recommendations, and as discussed making the post code file open should also be a part of this, and this procurement process should be referred to as an example of how to acquire a service not for the public interest. There should also be serious consideration for the creation a national spatial data infrastructure. Companies can help the public sector overcome its data analytical short comings; however this needs to be accompanied by good procurement practices and agreements whereby government own the data and can look under the hood of any software or analytical output provided, and that data be interoperable with multiple analytical platforms and not locked into proprietary systems. In addition, in the

science based departments, there is significant data analytical capabilities that have been overlooked here, especially in the EPA, Marine, ordinance survey, the Commissioner for Irish Lights, the CSO, population health and others. Perhaps along with an inventory of data assets an inventory of in house skills and the development of a cross departmental and jurisdictional public sector big data analytical working group should be created.

Finally, the report ends with “the economic and social benefits available from enhanced adoption of big data and analytics are potentially transformative. The spectrum of benefits on offer spans from improved health and environmental outcomes to better efficiency”. This is fantastic; however, the body of the Report did not discuss any targeted efforts at these issues. And it is this along with other issues discussed throughout this analysis that is part of the Report’s major shortcomings.

7. Final Observations

The report is to be lauded for attempting to address in a relatively tight time frame the criteria as set in the Request for Tender. It sets out a picture of the big data landscape in Ireland, before providing a number of recommendations as to how to nurture and expand Ireland’s role as the international site for big data research and development and commercial enterprise.

The report, however, also has some shortcomings and limitations. It is, for example, narrowly focused on developing strategies to build the human resources capacity for private sector companies in Ireland through the use of public education and research funds, with only a small focus on the private sector funding the training of its own staff. Moreover, the focus is almost exclusively on the ‘hard skills’ in areas of product and service development. In fairness, it is part of a larger Action Plan for Jobs and an outcome of this is the government’s research strategy of Data Analytics, Management, Security and Privacy 2013. Although, management, security and privacy in that strategy are primarily discussed in instrumental and technological terms and not in public policy terms, nor are these discussed in the Big Data Skills Report, most notably there is no mention of law and regulation.

Unfortunately the report was not created to inform and complement a national big data strategy. The strategy therefore is not R & D in general, or for the support of thematic areas that could benefit from big data analytics such as molecular biology, marine research, pharmaceutical and agribusiness, addressing climate change or energy production and efficiency in Ireland. Instead, the strategy is about jobs and skills for existing industry already involved in big data analytics, and to some extent the public sector, and not about a plan for the nation, nor for what might be considered the public good. For example, there is no systematic mapping out of the big data sector in Ireland; there is no inventory of the companies engaged and what they do; there is no analysis of the infrastructure required, albeit there is discussion elsewhere about cloud and high end computing; there is no understanding of the software used; nor do we know if there are clusters of excellence or knowledge where big data analytics are concentrated except for INSIGHT and CeADar which are

excellent R & D instrumental programs. There was an excellent investigation on the skills supply side in Chapter 6, although digital humanities, geosciences, geocomputation, geomatics and the earth sciences were left out as major contributors to developing a pool 'Deep Analytical Talent'.

Furthermore, the report does not include a focus on building public policy capacity related to issues such as law, privacy, data protection, regulation, copyright, procurement, patents, intellectual property (IP) and data resale, to name a few. The creation of public internet interest legal clinics, developing centres of excellence or think tanks in these areas as well as the creation of internet law and policy research chairs would be a good way to complement the demand for 'hard skills' focus. These 'soft skills', competencies and qualifications are subsumed within the *Big Data Savvy* category, although there is no explicit mapping of these against what is offered in academic institutions. The solution is envisaged as a hollowing out of the state with the private sector undertaking public sector work: *'Public sector employers are less ambitious about their future employment levels, in part due to the current restrictions on recruitment. Part of their demand is likely to be addressed through outsourcing to the private sector'* (p.13).

The policy implications of government outsourcing data analytical functions to the private sector are troublesome to say the least. Questions of who own the data? Can those data be shared with the public in an open fashion? What was the model used to do the analysis? What algorithms were deployed and are they propriety? There are issues of ownership and control that must be addressed. The Post Code exemplifies these problems, it is a lauded example in the Report, yet this is a significant public sector big data investment will primarily benefit the private sector and not the public. If evidence based decision making is something the Irish Government is truly invested in, and it would seem that it is with its open government and open data strategies, then those legal, policy, methodological, data life-cycle management and procurement issues need to be addressed. This is especially the case if issues of public expenditure and forecasting are based on the output of private sector companies who may stand to gain from the analysis, as is the case of the outcomes of this report. Uncritical predictive governance is also of growing concern, necessitating, along with investment in the 'hard skills'. Investment in the social sciences and critical thinking are also required, and these skills need to be found in non STEM academic departments. Political economy, communications, media studies, political science, socio-technological studies, public policy and public affairs, should also be part of the big data ecosystem in Ireland. I think this is one of the big missed opportunities of this Report.

If Ireland truly wants to become an innovator in this field and the leading country in the field of big data in Europe then part of that leadership should include law, regulation and policy. These are conspicuously absent and perhaps this is the result of the drivers. It is overseen by a Joint Industry/Government task force on Big Data and a Steering Group with representation primarily from the private sector. The EGFSN also does not include lawyers, auditors, actuaries and public policy experts,

but it does include an excellent group of subject matter experts and dedicated public servants in a number of areas. Industry needs to be tempered by solid public policy and regulation, while the public interest needs to be factored into any government strategy, industry is very well represented here, the public interest beyond jobs and FDI is not.

The Report, curiously mentions the Government's joining of the Open Government Partnership as well as the development of its open data strategy, as these are considered as enablers of big data. This is unusual, as most of the data public administrations produce are small data, and that includes the Census and mapping agencies such as the Ordnance Survey. Making government data easier to find and under less restrictive licencing would make everyone's job easier, and that is a good thing. Although, a spatial data infrastructure would be better in the long term. In addition, this would allow for greater commercialization of public sector data, as small data, and the big datasets produced by some departments and publicly funded research could be combined with private sector data and also be used to develop baselines. The Report does not discuss small data in a meaningful way except to mention that they exist. The Big Data Skills Report is, overly reliant on the analysis of the McKinsey Global Institute 2011 report entitled Big data: The next frontier for innovation, competition, and productivity, and it is here that open data is mentioned as a driver. This aligns with earlier work on the Programmable City that points to open data increasingly being of interest for commercialization purposes. Opening government is a good thing, as are open data in general, but these are a small part of the big data puzzle, and procurement needs to take an open by design approach, and avoid situations such as the post code and the Ordnance Survey.

The report offers an analysis of existing employment demand, produced three future demand scenarios, assesses Ireland's skill supply, and makes some useful recommendations and these were discussed in detail earlier on. There are a number of big data strategies found in a number of government reports introduced at the beginning of this post, in addition to this one, but what is sorely missing is an overall strategy that would be of benefit to the Republic, that would tackle issues of importance to Irish society, be in the public interest, while also encouraging and supportive of private sector investment.

Box 4.1 Proposals advanced by survey respondents

Education

A number of proposals were made relating to learning and education. These include that computer skills should be introduced at an earlier stage in the education system and there should be more focus on the teaching of maths at primary level. Regarding the teaching of maths, many respondents advanced the idea that a problem-solving approach should be adopted in second level maths curricula. Greater female participation in analytics should be targeted from an early age.

At third level, cooperation between industry and academia is crucial to ensure that education reflects skills needs. Proposals included involving industry experts in devising curricula and teaching analytics on university programmes. Work placements, both in industry and in research centres, should form an integral part of third level courses for students of big data and analytics and collaboration with industry is vital in this regard. In particular the Canadian University of Waterloo model was noted, whereby students are equipped with 24 months of experience upon graduation. Facilitating students on relevant courses to meet people already working in the area to discuss career options is seen as an effective way to attract them to the field.

As a means of raising standards it was suggested that universities with the strongest records in maths and statistics should come together in order to offer specialised courses. Increased funding should be provided for research to support career options for analytics in academia and to drive innovation. Respondents expressed the view that analytical modules should also feature across less technical disciplines, such as business and social sciences.

Steps should also be taken to allow transition into analytics from other disciplines and up-skilling and retraining into the area should be encouraged. To this end, short courses for individuals with technical backgrounds should be provided and be VAT reduced or exempt.

Promoting big data and data analytics

Companies were of the view that arriving at clear definitions of 'big data' is important to improve the appeal of the sector. This would facilitate the development of analytics-related career paths, making a career in this area more attractive. Recruitment agencies' understanding of analytics should also be improved; business associations and representative groups can play more active roles in fostering the area by holding workshops, seminars and awareness campaigns to demonstrate the potential value of data analytics to firms and as a career path.

8. Key words

Big Data, Skills, Gartner,

9. Acknowledgements

The research for this submission is funded by a European Research Council Advanced Investigator award (ERC-2012-AdG-323636-SOFTCITY).