Project acronym:	BYTE
Project title:	Big data roadmap and cross-disciplinarY community for addressing socieTal Externalities
Grant number:	619551
Programme:	Seventh Framework Programme for ICT
Objective:	ICT-2013.4.2 Scalable data analytics
Contract type:	Co-ordination and Support Action
Start date of project:	01 March 2014
Duration:	36 months
Website:	www.byte-project.eu

Deliverable D2.3:
Open Access to Data

Author(s):	Anna Donovan, Rachel Finn and Kush Wadhwa, Trilateral Research and Consulting Lorenzo Bigagli, CNR -IIA Guillermo Vega Gorgojo and Martin Georg Skjæveland, University of Oslo
Dissemination level:	Public
Deliverable type:	Final
Version:	1
Submission date:	30 September 2014

Table of Contents

EXECUTIVE SUMMARY

In this deliverable, we utilise deskwork to review the literature associated with open access to data policies and initiatives, and to discover the meaning of open access policies and initiatives in the public sector as compared with similar policies and initiatives driven by the private sector. We also explore how effective these policies and initiatives have been, to identify barriers associated with open access to big data, and in particular, explore good practices. This work was undertaken as part of the EU FP7-funded project "Big data roadmap and cross disciplinary community for addressing societal externalities" (BYTE), within work Package 2 (WP2).

In chapter 2, we examine policies for open access to big data in the public and private sectors. Publicly funded data such as research data and government data in the general sense includes statistics about environmental and geographical data, meteorological data, business information, and legal information. The public sector leads the charge with open access policies relating to big data. Governments have formal roles to play in this process such as facilitating open access to large data sets, especially by facilitating interoperability of datasets, as well as regulating to diminish the negative externalities arising out of new risks associated with making big data open. However, the accessibility of big data sets is no longer synonymous with only government held or other publicly funded research data, as powerful private sector profiling and data mining technologies are increasingly supporting open access initiatives for commercial purposes. However, open access to big data policies are not yet implemented in the private sector as widely as they have been in the public sector. This is despite the great potential for industry when business models include open access elements.

In chapter 3, we evaluate open access initiatives in relation to a number of case study areas including: health data; crisis data; energy data; environmental data; transport data; cultural data; and smart cities/ utilities. Open access policies can be developed in relation to any types of large data sets to produce a multitude of benefits as seen in the case studies examined in this section. For each case study, we identify open access policies, initiatives and/ or models that have access to the data, the extent to which that data is made open, and the associated benefits or issues raised by these policies and initiatives.

Health data
First, big health data may be made open by a number of stakeholders in the industry including patients, providers, insurers, and governments. However, there are more examples of access initiatives in the public sector than the private sectors at this time. Some of those initiatives examined in this report include the GOSgene health initiative, Gen Bank and Protein Data Banks (in Europe and internationally), Teralab initiative and the Yale University Open Data Access initiative. An examination of open access initiatives in health care reveal that open health data produces a number of positive outcomes for public and private sector stakeholders. However, it appears that open access to large volumes of health data are provided by the public sector organisations and being utilised for commercial gains in the private sector, whilst private health care companies drive fewer open access initiatives. We reveal that collaborations between the public and private sectors may yield greater benefits for both.

Crisis data
Second, big crisis data is creating value for communities and organisations when employed in open access initiatives. Crisis mapping and social media networks that disseminate crisis data

produce a number of positive outcomes for users and society, as well as holding a number of potential commercial benefits for the companies implementing them. Crisis data is increasingly being made open via crisis mapping and through social media networks. These examples serve not only as a public service, but additionally as initiatives capable of generating profit by acquiring personal data, including behavioural data, from users who access these services during peak times. Examples of crisis data creating value for communities and organisations through open access initiatives include the Ushahidi crisis mapping platform and Twitter. However, the move towards developing open access initiatives such as criss mapping and social media to disseminate crisis data present challenges as well as opportunities. Some negative implications include when data used in crisis mapping or disseminated via social media is inaccurate, comprising personal security and privacy or hampering emergency efforts. However, negative implications of the use of crisis data ar often overlooked because such initiatives are produced largely as a public benefit and as part of humanitarian efforts. Relevantly, it is difficult to identify examples of European open access initiatives involving crisis data, which represents a discrepancy in the market and also an opening for the fostering of such initiatives.

Energy data

Fourth, releasing open energy data produces a number of opportunities for innovation. This is illustrated by the Norwegian Petroleum Directorate ("NPD") initiative called FactPages. The NPD is a governmental specialist directorate and administrative body in the energy sector that promotes open access. To that end, the NPD manages the NPD FactPages, which contain data about petroleum activities on the Norwegian continental shelf (NCS) ranging from operating companies and production licences, fields and discoveries, facilities and pipelines, to wellbores and stratigraphic data. Some of this data dates back to the start of oil production on the NCS in the early 1970s. The data in the FactPages is collected from companies that operate on the NCS, and this information forms the basis for the authorities' planning of future activity and their judgement of existing activity. Additionally, an important purpose of the FactPages is to secure efficient sharing of information between the companies, and to provide sufficient information to the public. This case study example highlights that releasing both the raw data and schema of a database to the public can be a terrifying thought for database administers, and requires properly implemented change management processes since internal changes are also exposed to the public. However, these are procedures that should be in place for any important dataset in any event, and the benefits of having truly open data outweighs the costs of producing and maintaining open data sets.

Environmental data

Fourth, the Digital Agenda for Europe recognises that the big data revolution brings about novel ways of understanding and addressing environmental challenges, and two examples that reflect the sentiment of this statement are the following cases of open access to big data in the environmental sector: the GEO initiative and the Copernicus Programme. These examples illustrate several implications of accessing a big amount of data in conjunction with open policies, in particular, as regards data stemming from Remote Sensing and Earth Observation, where a "big data revolution" is predicated. The increasing availability of multidisciplinary data available from new observation platforms is expected to empower scientists and society with unprecedented resources to understand our planet and better control or mitigate the environmental dynamics. The examples briefly discussed above highlight that the sector is witnessing a general push to abandon the traditional model of data protection, in favour of full and open exchange of data, believed to lead to new applications, additional jobs and more open competition. However, a number of negative implications of open environmental data

sharing include: interoperability issues due to the large heterogeneity of technologies, applications, languages, and legal frameworks characterizing the context; and financial issues given the investments required for Earth Observation, and the strong interest of the industrial sector to protect their investments and competitiveness. Addressing these problems requires mutual policies on the exchange, sharing, access and use of interoperable data and services across the various levels of public authority and the different sectors of society, at a global level.

Transport/ logistics data
Fifth, the Digital Agenda for Europe recognises that the transport sector can clearly benefit from big data collected through sensors, GPS data and social media in particular. In that regard, we examine smart transport systems, in particular, the UK National Public Transport Data Repository. This example reveals that applications in transport and logistics can clearly benefit from the open availability of increasing quantities of data. Sharing, analyses and cross-combination of transport data related to people and goods support both the public and the private sector in optimising multimodal transport, managing traffic flows, and reducing delivery expenses (e.g. time, fuel and human resources) through the use of route planning support systems. However, negative implications are also perceived, mostly by the private sector, regarding the costs and benefits of a generalized uptaking of the open data approach in the transport domain, and the resulting business model. The reluctance of commercial stakeholders does not seem to be easily addressable in a policy-driven way, that is via regulatory actions.

Cultural data
Lastly, cultural data provides another example of open data initiatives that display both benefits and negative externalities. Several activities are already stimulating the re-use of cultural heritage in order to demonstrate the social and economic value of cultural content. An example of this is the Europeana Creative, which aims at facilitating the re-use of digital objects by the creative industries. Europeana provides a framework for many other current initiatives promoting open access to big data in the cultural sector, such as the OpenGLAM initiative, and the European Library. The European Library is the aggregator of digital content from national libraries for Europeana and delivers digital content from national libraries on a monthly basis to Europeana. This section reveals that considerable financial and operational benefits could arise from better use of data in the cultural sector. However, the process of value creation in the cultural sector, sometimes characterized by a certain lack of interest in market success (if not by relying on subsidies), may also impact negatively on the perception of the social and economic value of cultural content.

Smart cities/ utilities
Smart cities are essentially based on the idea that open access to information can improve decision-making, resource management, and ultimately the living conditions of citizens. The European Commission has clearly expressed its vision of "smart cities" that provide public services to their citizens in a more convenient way, that are more responsive and citizens-centred, that provide the right information in real-time to allow for better everyday and business decision-making, and that achieve all this in an economically viable way so as to improve environmental sustainability. This is, in part, achieved because of access to open data from other sectors including transport and energy sectors, and represents collaboration across sectors. We review Jakarta and Florence as two notable examples of smart cities. These examples illuminate that open access to big data is seen as having positive implications in the specific context of smart cities, for instance on decision-making, resource management, and

the living conditions of the citizens, through resource optimization and planning, support to economical growth, innovation and value-added services (e.g. for tourism), transparency and citizen engagement, safety and control, creation and promotion of cultural value. However, fragmentation of initiatives, scarcely synergetic (e.g., through joint uses of physical and digital infrastructures), and sometimes insufficiently clear goals other than infrastructure deployment, and too little attention to commercial viability are the main barriers to realising the potential positive externalities that can flow from smart cities.

Overall, these case studies represent the relationship between open access policies across a variety of sectors, and they combine a number of public and private sector open access policies and initiatives. However, these examples also highlight the varying degree of openness that is provided by public and private sector organisations, as well as the potential for collaboration between sectors.

In chapter 4, we provide our recommendations concerning good practice lessons for open access to big data. This report summarises good practice examples, and culminates in an identification of good practice lessons be translated across sectors, as well as promoting the collaboration between private and public sectors in the development of 'open' access projects. This is important as it might result in diversity of the information available through open access models. Ultimately, good practice policies support maximising the value to be derived by open access policies to big data, as well as open access business models based on big data. Best practice policies can address a number of the following issues or include, but are not limited to, the following aspects that have been identified in the above case study examples:

- Focus on the development of e-infrastructure and interoperability
- Support information and education on the benefits of open access policies and initiatives in both public and private sector
- Encourage cooperation, collaboration and partnership between the private and public sector to produce integrated services
- Promote regulation governing open access
- Promote regulation of legal and ethical issues as they arise in relation to open access
- Promote technical security and safety aspects such as privacy by design

Overall, this report aims to illuminate the burgeoning relationship between big data and open access policies. It also recognises open access initiatives as a great benefit to society. Thus, open data is heralded as providing an abundance of opportunities for Europe, despite the presence of negative externalities, some of which are raised in this report. In the digital economy, big data represents a significant tangible asset. Encouraging asset holders to provide free and open access to that asset requires both voluntary and proscribed policies and initiatives so that the socio-economic benefits of big data can be fully realised.

1 INTRODUCTION

There are a number of definitions describing open access that reflect the primary reference of this concept to the context of publicly-funded scientific research publications, scholarly communications and government data. More recently, open access has come to refer to a number of different types of data practices with the advent of big data and its collection and utilisation in the private sector. This has particularly occurred through the provision of services and the development of products that draw on data obtained through open access initiatives and with a shift towards models that incorporate aspects of the open access philosophy. Thus, the relationship between open access and big data is evidenced by open access policies and initiatives in the public and private sectors. Such policies and initiatives also present possibilities for collaboration and partnerships between stakeholders in both sectors through the sharing of large datasets and by jointly contributing to the technological measures required in the facilitation of open access. The relationship between open access and big data is important because it produces a number of positive externalities by presenting opportunities for economic growth, innovative business models, improved health care and education, disseminating real time updates about crisis situations, and in making society more energy efficient. These positive externalities have been witnessed in the public sector where there is a longer history of open access policies and initiatives. Open access also presents a number of risks such as threats to data protection, consumer welfare and intellectual property rights as well as differential access to data among public and private organisations. However, there is increasing potential for all organisations to capture the benefits of open access to big data, so long as negative externalities are diminished.

This report will examine the accessibility of large data sets in public and private sectors by providing examples relating to health data, crisis data, energy data, environmental data, transport and logistics data, cultural data, and utilities/ smart cities data. These examples illuminate instances where the public and private sector stakeholders develop open access policies, or facilitate the implementation of open access initiatives and business models. By looking at these examples, we ultimately aim to explore and assess how well governments, companies and institutions have fared in making big data accessible to the public. The report reveals that the motivations behind the provision of open access to big data differs depending upon whether the policy is developed, or the initiative is undertaken, in the public or private sector. Further, there is much opportunity for the private sector to generate profit by developing business models that deliver products and services based on open access to the big data they hold. However, collaboration between both sectors may result in an optimal model that captures the value of big data.

1.1 OVERVIEW

Big data is an umbrella term referring to the large amounts of digital data continually generated by the global population. The speed and frequency by which data is produced and collected—by an increasing number of sources—is responsible for today's data deluge: the amount of available digital data is projected to increase by an annual 40%.[1] Big data has long been viewed as having three defining properties: volume, variety and velocity[2], and that

[1] United Nations Global Pulse, *Big Data for Development: A primer*, June 2013, p.1.
[2] Laney, D., 2001. 3D *Data Management. Controlling Data Volume, Velocity, and Variety in Application Delivery Strategy*, META Group, Stamford, CT, February 2001. http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

definition, put forward in 2001, has been accepted as providing the universal understanding of what distinguishes big data from other data sets.

Big data is different from open access data. The European Commission defines open access data as data that is freely accessible over the Internet.[3] Open access to government data (known as "open data") and/ or scientific research data are data that have been more commonly associated with data provided through open access policies and initiatives. Whilst, open access is not a defining characteristic of big data, which can be privately owned or have varying levels of access control[4], fostering the development of a relationship between open access and big data in public and private sectors, and possibly in collaboration, can result in an array of potential socio-economic benefits for the governments, businesses and citizens. Nevertheless, these benefits must be balanced against a number of risks that are associated with open access to big data.

Much of the data that is open access is public sector data or data resulting from publicly funded research primarily because these organisations are subject to a number of laws and mandates about opening access to data.[5] This includes data underlying scientific publications such as curated data and raw data[6]. In addition to these public sector initiatives, private sector organisations are increasingly collecting and utilising open access data through business models and initiatives that incorporate open access elements. There is evidence of these approaches involving the following types of big data: health data; crisis data; energy data; environmental data; transport and logistics data; and utilities/ smart city industries spanning the public and private sectors in Europe. However, despite a few solid private sector business models, it appears that the public sector is still the leader in providing open access to data. Furthermore, when these open data sets are collated with other datasets, they may amount to big data. However, open data and big data are not interchangeable. A report by McKinsey and Company clarifies:

> Big data refers to data sets that are voluminous, diverse, and timely. Open data is often big data, but "small" data sets can also be open. We view open and big data as distinct concepts. "Open" describes how liquid and transferable data are, and "big" describes size and complexity of data sets. The degree to which big data is liquid indicates whether or not the data are open.[7]

---

[3] European Commission, Commission Recommendation on access to and preservation of scientific information, C(2012) 4890 final, Brussels, 17 July 2012, p.13. http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf

[4] United Nations Global Pulse, *Big Data for Development: A primer,* 2013, p.2. http://www.unglobalpulse.org/sites/default/files/Primer%202013_FINAL%20FOR%20PRINT.pdf

[5] See for example, the European Commission's Directive on Public Sector Information (Get ref from RECODE WP3 Report).

[6] Chan, Leslie, Darius Cuplinskas, Michael Eisen, Fred Friend, Yana Genova, Jean-Claude Guédon, Melissa Hagemann, Stevan Harnad, Rick Johnson, Rima Kupryte, Manfredi La Manna, István Rév, Monika Segbert, Sidnei de Souza, Peter Suber, Jan Velterop, *Budapest Open Access Initiative,* Budapest, 2002. http://www.budapestopenaccessinitiative.org/readPatrick O. Brown, Diane Cabell, Aravinda Chakravarti, Barbara Cohen, Tony Delamothe, Michael Eisen, Les Grivell, Jean-Claude Guédon, R. Scott Hawley, Richard K. Johnson, Marc W. Kirschner, David Lipman, Arnold P. Lutzker, Elizabeth Marincola, Richard J. Roberts, Gerald M. Rubin, Robert Schloegl, Vivian Siegel, Anthony D. So, Peter Suber, Harold E. Varmus, Jan Velterop, Mark J. Walport, and Linda Watson, "Bethesda Statement on Open Access Publishing", June 2003. http://www.earlham.edu/~peters/fos/bethesda.htm; Max Planck Society,"Berlin Declaration on Open Access in Sciences and Humanities", Berlin, 2003. http://openaccess.mpg.de/286432/Berlin-Declaration

[7] Manyika, James, Michael Chui, Peter Groves, Diana Farrell, Steve van Kuiken and Elizabeth Almasi Doshi, "Open Data: Unlocking Innovation and Performance with Liquid Information", *McKinsey & Company,* October 2013, p. 14.

Nevertheless, the concept of open access is broadening to reflect an increasing understanding of the benefits associated with providing open access to big data across industry and sectors. For example, open data can generally be understood across many sectors as being data that can be freely used, reused and redistributed by anyone, subject only, at most, to the requirement to attribute and share alike.[8] What this means is that data may be made available as a whole in an easily accessible format, the data must be capable of re-use and re-distribution, and be universally available. To that end the interoperability of large, different data sets is required to enable effective open access to big data.[9] This is because "Open means ensuring that there is little or no barrier to access for anyone who can, or wants to, contribute to a particular development or use its output."[10] Thus, key features of openness ought to include: availability and access; reuse and redistribution; and universal participation.[11]

Big data sets can be local, national or global and they can be obtained from both government and commercial sources. The Article 29 Working Party opines that big data "refers to gigantic digital datasets held by corporations, governments and other large organisations, which are then extensively analysed using computer algorithms".[12] The business of both the public and commercial sectors is increasingly driven by the gathering progressively more sophisticated analysis of data from a range of sources. It has been estimated that by 2020, 35 zeta bytes of digital data will be created each year.[13] This creates enormous opportunities for developments and discoveries based on that data, especially in the event that it is openly accessible. For example, the European Union policy on big data is entwined with its policy on open data, both of which are expected to foster significant innovations, public-private partnerships and increase European competitive advantage. This is because open access to the widest variety of information is thought to help achieve the highest impact, and because open access is believed to be the key to the linking, sharing and re-use of data, which also means the highest possible efficiency of utilisation. The European Commission recently highlighted the following four key reasons to encourage open access policies:

1. Untapped business and economic opportunities including a predicted 140billion profit across the EU27;
2. Better governance and citizen empowerment as a result of open data increasing transparency, citizen participation, and administrative efficiency and accountability;
3. Addressing societal challenges by enhancing the sustainability of health care systems, and it is considered essential for tracking environmental changes; and
4. Accelerating scientific progress particularly with regards to scientific learning and development.[14]

Thus, the development of big data research and policy goals should be aligned and combined

[8] Open Definition, "The Open Definition", no date. http://opendefinition.org/
[9] *Open Data Handbook,* no date. http://opendatahandbook.org/
[10] e-InfraNet Consortium, *e-InfraNet: 'Open" as the Default Modus Operandi for Research and Higher Education,* European Commission, Brussels, 2013*,* p.6.Whilst the focus of this report is research and higher-education sectors, the breadth of this definition enables the definition to be applied more generally and across sectors.
[11] Open Knowledge Foundation, *open data,* no date. https://okfn.org/opendata/
[12] Article 29 Working Party, Opinion 03/2013 on purpose limitation, 00569/13/EN, WP 203, Brussels, 2 April 2013, p. 35. http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf
[13] Working Group on Expanding Access to Published Research Findings, *Accessibility, Sustainability, Excellence: How to Expand Access to Research Publications ("The Finch Report")*, Publisher?, 2012, p.26.
[14] Nagy-Rothengass, Marta, 'Public Sector information at European Commission', *European Data Forum Presentations,* 2014. http://2014.data-forum.eu/edf2014-presentations

with open data practices to serve decision-making and meet scalability and performance challenges. Challenges include developing interoperable systems, services and content to facilitate open access to big data.

Data has been described as "the new gold"[15], and there is now widespread access to a broader variety of increasingly large data sets such as health data, crisis data, energy data, transport and logistics data, cultural data, and utilities and smart cities data. This is because the technologies of collection and analysis that fuel big data are being used in every sector of society and the economy. Open access policies and initiatives implemented in relation to these big data sets held by public and private sector stakeholders beget a number of positive externalities that culminate in purporting to enhance how people live and work in integral areas of life and society, as well as capture profit generated by the implementation of open access policies and initiatives relating to big data. Open access to big data initiatives create social and economic value, as well as decreasing the discrimination previously inherent in access being directly linked to an individual, an organisation's or a institution's ability to fund its data collection. Practical benefits of "openness" include the enhanced ability to combine different datasets together, thereby developing a greater number and better quality products and services. This, in turn, provides organisations, particularly those in the private sector, with new revenue streams.

However, whilst open access policies and initiatives are heralded as producing this benefits in the public interest, they also raise a number of important legal and ethical issues that may negatively impact society and its citizens. Open access polices for big data may compromise rights such as privacy, data protection and intellectual property. This means that the development of open access policies and the implementation of such open access initiatives must accord with applicable laws in these areas, as well as respecting the ethical values that make up the fabric of society. Furthermore, while the large-scale collection, storage and analysis of data has always been a discomfort for some scholars, the mandate to provide open access to this data by default augments these concerns.

This deliverable examines open access policies, initiatives and business models relating to health data, crisis data, energy data, environmental data, transport and logistics data, cultural data, and utilities and smart cities data. This information is used to identify the positive and negative effects associated with fostering the relationship between open access and big data and to provide good practice suggestions to support the continued growth of open access to big data.

1.2 STAKEHOLDERS TO BE EXAMINED

The information gathered in this report considers a number of stakeholders relevant to open access and big data. The stakeholders examined for this report include commercial bodies, public sector bodies, academic institutions, other institutions, with specific reference to the following sectors:

- Health data
- Crisis data
- Energy data
- Environmental data
- Transport data

[15] Nagy-Rothengass, Marta, 'Public Sector Information at European Commission', *European Data Forum Presentations,* 2014. http://2014.data-forum.eu/edf2014-presentations

- Cultural data
- Utilities/ smart cities

Open data, by way of the release of information by governments and private institutions and the sharing of private data in the aforementioned sectors, enable insights across these industries. Manyika et al. argue that this trend has "profound implications for companies, governments, and individuals."[16] However, publicly funded organisations and institutions that are subject to open access mandates drive more open access initiatives than stakeholders in the private sector. This is because private sector organisations tend to make data open only to the extent that it will enable them to grow their proprietary data stocks or to attract advertisers as a means of generating revenue. The effect of this is that private sector organisations are placed at a competitive advantage because they have access to their own privately held data, as well as other types of data from organisations subject to open access mandates, especially research data[17] and government data[18].

1.3 METHODOLOGY

Consortium partners undertook desk-based research to review the literature associated with open access to data policies and to discover the meaning of open access policies and initiatives in the public sector as compared with similar policies and initiatives driven by the private sector. We also explored how effective these policies and initiatives (including business models based on the provision of access to proprietary data) have been, to identify barriers associated with open access to big data and explore good practices. In order to identify the burgeoning relationship between big data and open access policies, we conduced a review and analysis of examples relating to environmental data, transport data, health data, energy data, cultural data, commercial data, and utilities/ smart cities data. This analysis also illuminates the difference between the open access and big data relationship in the public and private sectors. Ultimately, this study presents a number of good practice examples that may be implemented with the goal to encouraging access to big data in the industry sectors examined, but that may also be helpful in other sectors of society. They may be helpful to the extent that they increase the effectiveness of open access policies by identifying the benefits produced by these policies.

[16] Manyika, James, Michael Chui, Peter Groves, Diana Farrell, Steve van Kuiken and Elizabeth Almasi Doshi, "Open Data: Unlocking Innovation and Performance with Liquid Information", *McKinsey & Company,* October 2013, p.11.

[17] For an overview of open access mandates for research see Schmidt, Birgit and Iryna Kuchma, Implementing Open access Mandates in Europe: OpenAIRE Study on the Development of Open Access Repository Communities in Europe, Universität Göttingen, Germany, 2012.

[18] The revised PSI Directive 2013/37 EU. Also see Commission Decision 2011/833/EU on the reuse of Commission documents.

2 POLICIES FOR OPEN ACCESS TO BIG DATA

2.1 OVERVIEW

Open access policies in relation to big data carry the potential to increase transparency in scientific, commercial and government decision-making[19], as well as generating growth and profit. The difference in the varying degrees to which stakeholders are making big data open in the public and private sectors reflects the motivations behind open access policies and initiatives. Although the extent to which big data is being made open differs between public and private sectors, open access policies continue to improve the flows of information and knowledge, and increases innovations and key developments. However, the increase in socio-economic benefits resulting from open access to big data can be accompanied by negative impacts. These negative externalities present challenges for governments and private organisations alike. Nevertheless, it is recognised that "Open data—from both public and private sources—are adding a new dimension to big data analytics and giving rise to novel, data-driven innovations".[20] Irrespective of the source of the big data, these open data sets share the same characteristics: accessibility (a wide range of users are permitted to access the data); machine readability (the data can be processed automatically); cost (data can be accessed free or at a negligible cost); and rights (limitations of use, transformation, and distribution are minimal).

Open access policies are premised on the provision of free, immediate and unrestricted availability of content, and in some case, unrestricted re-use of the data or literature.[21] Thus, the development of open access polices, and their implementation through related initiatives or as an aspect of innovative business or humanitarian efforts, are vital to extracting the potential value and benefits of the increasingly vast array of combined large data sets. However, the degree of openness varies subject to whether they are developed and implemented within the public or the private sector. This is related to the development of these policies with resect to government data or scientific research and publications, and other scholarly communications[22] as a way of achieving a public good. That tradition in the public interest is in contrast with recent private sector initiatives that may incorporate open access aspects into business models and strategies as a means of achieving commercial gain, in addition to producing societal benefits. Corporations are harnessing the benefits of open access to their materials, which enables collaboration to produce developments in their work, and in turn, provide benefits for consumers that result in further commercial opportunity for the company concerned. This transition from mandated open access in the public and research sectors to voluntary open access in the private sector, when commercial operators have a proprietary interest in the output, is occurring because of the positive externalities produced by this activity, despite a number of potentially negative impacts, such as legal and ethical

[19] Royal Society, *Science as an Open Enterprise*, June 2012. http://royalsociety.org/policy/projects/science-public-enterprise/report/

[20] Manyika, James, Michael Chui, Peter Groves, Diana Farrell, Steve van Kuiken and Elizabeth Almasi Doshi, "Open Data: Unlocking Innovation and Performance with Liquid Information", *McKinsey & Company,* October 2013, p. 4.

[21] Pinfield, Stephen, "A Mandate to Self-Archive? The role of Open Access institutional Repositories", *Serials,* Vol. 18(1), March 2005, p.31.

[22] As far back as 1998, the idea of electronic open access was being pursued: "Free access to traditional journals is affordable and achievable. It is the right thing to do for those who pay for the research and for those who do it: Organisation for Economic Cooperation and Development Committee for Information, Computer and Communications Policy Working Party on the Information Economy, 'Digital Broadband Content' DSTI/ICCP/IE(2004)11/FINAL, September 2005, p.72 - 73.

concerns associated with privacy and data protection. These negative externalities reflect the sentiment that "just because content is publicly accessible doesn't mean that it was meant to be consumed by just anyone (Boyd & Marwick, 2011)".[23] Further, open big data can expose individuals and businesses to several risks, especially reputational ones. Opening information such as electricity use or school performance to create aggregated views of population behavior raises serious concerns among consumers who fear that their data will be tied to them and could harm their economic or social standing. Conceivably, credit card companies could raise interest rates on households that waste electricity or the inadvertent release of information about a particular student could lead to bullying which may impact his/her future educational options or lead to discrimination and bullying. For businesses, open data released by third parties could expose poor environmental or labour practices or show that their products or services compare poorly for price and quality. Companies can also put consumers off by using open data to create online advertisements or marketing offers that show that the company knows too much about the consumer. Another risk arises in sharing benchmarking data among businesses, if the pooled data inadvertently reveal confidential information.[24]

Regardless of whether open access policies are developed by public or private sector stakeholders, the sustainability of such policies and initiatives are largely dependent on funding. What this means is that they open access to big data will continue subject to partisan funding in the public sector and the ability of such initiatives to generate revenue for companies in the private sector. Funding is also vital to develop technical support measures enabling interoperability.

Overall, whether open access policies are developed by public or private sector stakeholders, or as a result of the collaboration between stakeholders from both sectors, access to big data has

> become a critical element for breaking down information gaps across industries, sharing insights that can raise productivity, enable innovation, and replace traditional and intuitive approaches with data driven processes. Analytics powered by open data can also help uncover consumer preferences, anomalies in costs, and variations in performance—all of which can inform new products.[25]

Thus, the proliferation of open access policies and the exponential growth in the amount of data that is expected to be stored and accessible may have significant impacts. Ultimately, this is why a number of open access policies have been developed in the public and private sectors.

2.2 PUBLIC SECTOR OPEN ACCESS DATA POLICIES

Publicly funded data such as research data and government data may include statistics about environmental and geographical data, meteorological data, business information, and legal information as well as others. The public sector leads the charge with open access policies relating to big data. Governments have formal roles to play in this process such as facilitating

[23] Cited in Boyd, Danah and Kate Crawford, "six Provocations of Big Data', *A decade in Internet Time: Symposium on the dynamics of the Internet and Society*, September 2011, p.2. SSRN:http://ssrn.com/abstract=1926431 or http://dx.doi.org/10.2139/ssrn.1926431

[24] Manyika, James, Michael Chui, Peter Groves, Diana Farrell, Steve van Kuiken and Elizabeth Almasi Doshi, "Open Data: Unlocking Innovation and Performance with Liquid Information", *McKinsey & Company,* October 2013, p. 18.

[25] Ibid., p. 11.

open access to large data sets, especially by facilitating interoperability of datasets, as well as regulating to diminish the negative externalities arising out of new risks associated with making big data open. Despite these potential negative impacts, recent European policy has displayed a firm commitment to providing open access to data resulting from publicly funded activities, and this is increasingly also drawing in large data sets.

Government-held big data technologies are purchased with public funds and result in initiatives that justify the open access infrastructure. For example,

> More than 40 countries—from every region of the world and at every stage of development—have established open data initiatives. Thus, national governments are facilitating the opening up of all kinds of data sets to promote economic development, spark innovation, and find ways to make government work better.[26]

In particular, the European Commission refers to open data as "an engine for innovation, growth and transparent governance.[27] However, despite these benefits, open access to big data presents negative externalities such as legal barriers including privacy issues and intellectual property issues (addressed in Deliverable 2.1). Barriers are also presented by the lack of interoperability between open access platforms.[28]

Developing open access policies and initiatives across the public sector (in relation to research data and government data) still reflects, to some extent, the open access philosophy for publicly funded research. That philosophy was articulated in the 2002 Budapest Open Access Initiative[29], which recognises that "[…] Research results are best utilised when others are permitted to build upon them, provided credits are duly given."[30] Directive 2003/98/EC on the re-use of public sector information ("PSI"), and Commission Decision 2011/833/EU on the reuse of Commission documents reiterate the sentiment of the Budapest Open Access Initiative. On 23 June 2013, the Commission adopted a revised PSI Directive 2013/37 EU to promote the implementation of PSI policy across Europe by ensuring compliance and the development of soft-law instruments such as guidelines on high value data, licensing, re-use and charging, with its key points of revision to be transposed by July 2015. This embodies the European Commission's "open by default" policy" to facilitate exploitation and reduce transactions costs associated with re-use of data.[31] Other EU initiatives include open access publications resulting from FP7 such as the OpenAIRE[32] initiative which links 150

[26] Manyika, James, Michael Chui, Peter Groves, Diana Farrell, Steve van Kuiken and Elizabeth Almasi Doshi, "Open Data: Unlocking Innovation and Performance with Liquid Information", *McKinsey & Company,* October 2013, p.15

[27] European Commission, Open Data, an engine for innovation, growth and transparent governance, COM (2011) 882 final, Brussels, 12 December 2011.
http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0882:FIN:EN:PDF

[28] Nagy-Rothengass, Marta, 'Collaborating on Interoperability to Achieve a Single Digital Market', *European Data Forum Presentations,* 2014. http://2014.data-forum.eu/edf2014-presentations

[29] Budapest Open Access Initiative, 2002, p.3. The Budapest Open Access Initiative 2002, was followed by the Bethesda Statement on Open Access Publishing, and later the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities in 2003. Also see recognition of the general approach to open access in this context in OECD Report 2007, and the Royal Society Report, op. cit., 2007.

[30] See Chan, Leslie, "Supporting and Enhancing Scholarship in the Digital Age: the Role of Open Access Institutional Repositories in the Digital Age", Canadian *Journal of Communication,* Vol. 29, 2004,p.280.

[31] European Commission, *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions Towards a Thriving Data-Driven Economy,* COM (2014) 442 final, Brussels, 2.7.2014, p.5.

[32] www.openaire.eu

repositories, as well as the European Commission-led Open Access Pilot in FP7.[33] In addition, in December 2013, the European Commission launched a pilot to open up data resulting from publicly funded projects. The pilot asks researchers to make the underlying data needed to validate the results presented in scientific publications and other scientific information available for use by other researchers, innovative industries and citizens.[34] New requirements for open access to research data continue to be a focus of EU policy because of the following perceived benefits:

- Closer linkages between research and innovation with benefits for public policy and services, and for economic growth;
- Improved efficiency in the research process itself, through increases in the amount of information that is readily accessible, reductions in the time spent in finding it, and greater use of the latest tools and services to organise, manipulate and analyse it; and
- Increased returns on the investments made in research, especially the investments from public funds.[35]

European policy continues to encourage and facilitate open access to research data in the public sector.[36] More recently, open access policies and initiatives have developed into "A mainstream movement that is receiving worldwide attention from researchers, institutional leaders, policymakers, and funding bodies, as well as commercial publishers."[37] These are increasingly involving big data. Further, improving the flows of the information and knowledge in the public sector through open access to government data (and not just research data) continues to promote enhanced transparency, openness and accountability. This is perhaps why big data initiatives are not a new concept:

> Governments since have gathered and shared (to varying degrees) vital information: GPS data, weather data, and census information are examples of information sets that are collected by public agencies in the course of their work and then made freely available for use by citizens, businesses, and academics.[38]

Other open government data initiatives involving big data stores include e-infrastructure developments such as European Data Infrastructure (EUDAT), which brings together data service providers and users. In fact, there are currently more than 150 open data portals in Europe and include the European Commission Open Data Portal for the re-use of the Commission's data which includes 6500 data sets from 40 providers (2012), the Pan-European Open Data Portal Pilot (LOD2 Project) with 46,000 referenced data sets from 14 countries (2013), and the European Digital Service Infrastructure for Open Data (2014 – 2020). The LOD2 Project is a significant example as it is committed to providing one single gateway to re-usable information with the aim of enabling combination of information held

[33] European Commission, "Fact Sheet: Open Access in Horizon 2020", 9 December 2013.
[34] European Commission, "European Commission Launches Pilot to Open Up Publicly Funded Research Data", *Press Release,* Brussels, 16 December 2013. http://europa.eu/rapid/press-release_IP-13-1257_en.htm
[35] Finch, Janet, *Report of the Working Group on Expanding Access to Published Research Findings, Accessibility, Sustainability, Excellence: How to Expand Access to Research Publications*, June 2012, p. 5. http://www.researchinfonet.org/wp-content/uploads/2012/06/Finch-Group-report-FINAL-VERSION.pdf
[36] See European Commission, "Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions Towards a Thriving Data-Driven Economy", COM (2014) 442 final, Brussels, 2.7.2014.
[37] See Chan, Leslie, "Supporting and Enhancing Scholarship in the Digital Age: the Role of Open Access Institutional Repositories in the Digital Age", Canadian *Journal of Communication,* Vol. 29, 2004,p.280.
[38] Manyika, James, Michael Chui, Peter Groves, Diana Farrell, Steve van Kuiken and Elizabeth Almasi Doshi, "Open Data: Unlocking Innovation and Performance with Liquid Information", *McKinsey & Company,* October 2013, p.11

by various open data portals held at various levels throughout the EU, providing open data related services including a dedicated language service infrastructure for language resources to enable multi-lingual sharing of data. The Connecting Europe Facility ("CEF") is also another good example of the public agency commitment to open access to big data. CEF will initiate the deployment of a comprehensive open data core platform to, among other things, make data accessible and downloadable through links to open data providers' resources, and to employ agreed standards for metadata and data.[39] Further, the EU has committed to implementing the G8 Open Data Charter through the publication of core data sets held at EU level (budgets, elections etc.), the publication of high value data sets held at EU level, publishing data on the EU Open Data Portal, promoting the application of the principles of the G8 Open Data Charter in all 28 Member States, supporting open data activities, and sharing experiences of open data work.[40] The? G8 principles are the open data by default principle, the quality and quantity data principle, the usable by all principle, releasing data for innovation principle, and the principle relating to releasing data for improved governance.[41] NESSI – European Technology Platform for the new digital information society and economy and is powered by software and services and data. The BIG Project and NESSI have aligned interests in fostering big data value creation in Europe. Open data repositories (e.g., the UK data service and the European Bio-Informatics Institute).

The number of open access policies that involve big data are increasing, and they also potentially create big data from linking and joining small data sets. It has thus become necessary to concentrate on supporting interoperability as part of developing open access initiatives and principles for big data. For example, the European Commission adopted a Communication "Towards Interoperability for European Public Services" in December 2010. The ultimate objective of that communication is a common data ecosystem. This approach takes into account legal, organisational, semantic and technical layers of interoperability. This is based on the view that interoperability framework is important for data driven innovation in the EU digital market.[42] This is especially important to the facilitation of open access to big data.

Therefore, the public sector continues to facilitate open access policies and is the source of a number of examples of the successful implementation of open data policies both in relation to research data and publications and government data. These policies increasingly include big data and or lead to the creation of big data by enabling the linking of a number of data sets through interoperability measures. In particular, the European open government data movement has the potential to produce successes in the form of greater accountability within government and transparency for the governed on the premise that citizens are now armed with information that was previously inaccessible to them. Generally, in the public sector, restrictions on access to research and government data are viewed as barriers against realising the full potential of that data, and the subsequent benefits and innovations that emerge as a result of that access. For example, without levels of access to, and use of, research and technical information by knowledge-based, technology SMEs in Denmark, 27% of products and 19% of processes developed or introduced during the last three years would have been

[39] Nagy-Rothengass, Marta, 'Public Sector information at European Commission', *European Data Forum Presentations,* 2014. http://2014.data-forum.eu/edf2014-presentations
[40] Ibid.
[41] Nagy-Rothengass, Marta, op. cit., 2014.
[42] Ibid.

delayed or abandoned without access to academic research.[43] These new products contributed an average of 46% of annual sales and on this basis they calculated that the value of academic research to sales was equivalent to £1.8 million per firm per year and the average value of cost savings was £58,000.[44] However, whilst big data are made available by governments and public sector agencies and organisations, it is no longer considered that "open data to be synonymous only with data released by governments"[45], or other publicly funded institutions, as the concept of open access is being adopted in the private sector within a commercial framework. However, the degree of openness varies subject to whether it is granted by a public or private organization, the motivation behind the open access policy and the whether the policy is voluntary or mandated.

2.3 PRIVATE SECTOR OPEN ACCESS DATA POLICIES

Powerful private sector profiling and data mining technologies can be utilised in a way that supports open access initiatives for commercial purposes, however only few open access initiatives have been developed in the private sector. Private sector organisations tend to rely more so on the data available through open access initiatives in the public sector. Although businesses can take advantage of existing open access initiatives to develop their own products and services by utilising available data, they too can develop open access initiatives on their own or in collaboration with public sector organisations. There is great potential for private sector open access initiatives considering that 57% of businesses that responded to a 2013 survey conducted by the Data Warehousing Institute consider themselves to be "managing big data", in the sense of "very large datasets" which can include "streaming data from machines, sensors, web applications and social media".[46] Big data in the private sector can shape products and services available to consumers and businesses and ultimately, spark innovation, productivity and value. However, in the absence of any mandated provision of open access in the private sector, private organisations may be reluctant to grant free and open access to their asset data as that data is still viewed as a competitive asset that requires protection for the generation of profit. Whilst this produces benefit such as trade secrets that support European industry driven competitiveness, it may be that greater advantages exist when the data is made open access, especially as industry in other sectors may re-use data in a different way entirely to its original use. This scenario does not pose any threat to business that made the data open after it had finished with it. These benefits are identified below in our examination of open access initiatives and policies relating to environmental, energy, transport, health, culture, commercial ad utilities/ smart cities data.

Manovich[47] describes three classes of people in the realm of big data and their interaction with big data: "'those who create data (both consciously and by leaving digital footprints),

[43] Parsons, David DR., Dick Willis and Dr. Jane Holland, *Benefits to the Private Sector of Open Access to Higher Education and Scholarly Research: A Research Report to JISC from Host Policy Research,* Host Policy Research, UK, 2011, p.8.
http://open-access.org.uk/wp-content/uploads/2011/10/OAIG_Benefits_OA_PrivateSector.pdf
[44] Ibid., p.8.
[45] Manyika, James, Michael Chui, Peter Groves, Diana Farrell, Steve van Kuiken and Elizabeth Almasi Doshi, "Open Data: Unlocking Innovation and Performance with Liquid Information", *McKinsey & Company,* October 2013, p.14.
[46] Russom, Phillip, "TWDI Best Practices Report: Managing Big Data", *TDWI Research*, Fourth Quarter 2013, p.5.http://www.pentaho.com/sites/default/files/uploads/resources/tdwi_best_practices_report-_managing_big_data.pdf
[47] Manovich, L.,"Trending: The Promises and the Challenges of Big Social Data, Debates in the Digital Humanities", ed M.K.Gold. The University of Minnesota Press, Minneapolis, 2011. http://www.manovich.net/DOCS/Manovich_trending_paper.pdf

those who have the means to collect it, and those who have expertise to analyze it." In that regard, "We know that the last group is the smallest, and the most privileged: they are also the ones who get to determine the rules about how Big Data will be used, and who gets to participate."[48] This effectively produces a gap in the market for new businesses that focus on this activity. Whether the private sector open access initiatives involve privately obtained data or publicly funded data, the sustainability of these models is dependent on the relationship between open access to big data and profit. A study conducted by Mckinsey & Company in 2013 reveals that:

> Making data more "liquid" (open, widely available, and in shareable formats) has the potential to unlock large amounts of economic value, by improving the efficiency and effectiveness of existing processes; making possible new products, services, and markets; and creating value for individual consumers and citizens. Realizing this potential will involve creating safeguards for personal privacy and business confidentiality, investments in technology, and changes in mindsets and work processes.[49]

Thus, open access to big data approaches in the private sector can enable efficient use of expensive resources, shared approaches, reduce duplication of effort and can save time, as well as support collaboration. These commercial gains are of major interest to private sector organisations as initiatives and business models and strategy open incorporating access to big data will likely present new avenues of competitiveness and profit generation for businesses, especially those operating in the digital marketplace. To date, a number of open access models (including open source) have been sharing information on the web. Open approaches in e-Infrastructures such as regional grids, high performance computing and cloud computing can provide a much more powerful system in a much more cost effective way than any "closed" localised system could offer.[50] This alone may be considered a good basis for the commercial decision to enter the open access arena, especially as open access to privately held data might enable commercial organisations (especially SMEs) to find new business opportunities and innovate. In the US, open meteorological data has enabled a range of local commercial weather services, and AppliSci, a specialist SME supporting big pharmaceutical and healthcare companies with leading edge research services, found directly relevant "process" evidence via openly accessible material thought capable of opening the way for an entirely new medical application for treatment of a rare infectious disease, with a market potential estimated at tens of millions of dollars.[51] Hence, providing access to data in the private sector can be both open and profitable.

---

[48] cited in Boyd, Danah and Kate Crawford, "Six Provocations of Big Data", *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*, September 2011, p.2. SSRN:http://ssrn.com/abstract=1926431 or http://dx.doi.org/10.2139/ssrn.1926431

[49] Manyika, James, Michael Chui, Peter Groves, Diana Farrell, Steve van Kuiken and Elizabeth Almasi Doshi, "Open Data: Unlocking Innovation and Performance with Liquid Information", *McKinsey & Company,* October 2013, p.11.

[50] e-InfraNet Consortium, *e-InfraNet: 'Open" as the Default Modus Operandi for Research and Higher Education,* European Commission, Brussels, 2013*,* p.16

[51] Ibid., For similar examples where businesses, especially SME's have benefitted from open access, see Parsons, David DR., Dick Willis and Dr. Jane Holland, *Benefits to the Private Sector of Open Access to Higher Education and Scholarly Research: A Research Report to JISC from Host Policy Research,* Host Policy Research, UK, 2011, pp.23-25. http://open-access.org.uk/wp-content/uploads/2011/10/OAIG_Benefits_OA_PrivateSector.pdf

There is a growing sentiment that "the market sees Big Data as pure opportunity"[52], and as such, granting free and open access to data forms the basis of some commercial business models, particularly by social media companies. However, open access initiatives in the private sector are threatened by the perception that opportunities flowing with big data are only realised if the data is kept secret or utilised by the data holder first and fullest before that data are released. Here, public benefits produced by granting open access to data are not important to the decision to publish data as open. Such products, such as social media networks, or location or sports fitness devices that rely on GPS, generate revenue for companies, whilst delivering benefits to consumers. Companies have been able to utilise data from open access sources to improve the productivity of current business processes, as well as develop new products, services, and entire lines of business for both established companies and entrepreneurs. However, private organisations are not themselves making their data open for others to re-use. This signifies an imbalance in the market. This includes businesses that are exploiting the need for better discoverability of data, which may include open access data.[53] Further,

> businesses are finding new ways of segmenting markets by blending open data with proprietary data and discovering new ways to raise productivity by using open and proprietary data to benchmark operations. There are also opportunities related to open access policies in the private sector such as opportunities for companies that aggregate and sell data and advise or consult companies on open data use.[54]

Therefore, the extent to which data is made open by private sector organisations is determined by the commercial value to be gained by doing so. This is despite private sector organisations realizing the potential of open access initiatives from the public sector. Private sector organisations rely on open and free access to other data sources to create their own big data stores. This can foster the growth of a few standout digital players who continue to build their resources and their data banks. This puts smaller, newer businesses and SME's at a competitive disadvantage. Chairman of the UK Competition and Markets Authority, David Currie, expressed this sentiment:

> The rapidly expanding online market or markets… increasingly touch all aspects of business. Making sure competition works effectively in these markets will be a major priority… the growing collection, processing and use of consumer transaction data for commercial ends …is proving an increasingly important source of competitive advantage [which could be] an increasing source of consumer detriment.[55]

Further, the European Commission has recognised the need for all stakeholders to have access to the possibilities presented by big data, and that the complexity of the current legal environment together with the insufficient access to large datasets and enabling infrastructure

[52] Boyd, Danah and Kate Crawford, "Six Provocations of Big Data", *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*, September 2011, p.2. SSRN:http://ssrn.com/abstract=1926431 or http://dx.doi.org/10.2139/ssrn.1926431

[53] Parsons, et al., op. cit., 2011, p.7.

[54] Manyika, James, Michael Chui, Peter Groves, Diana Farrell, Steve van Kuiken and Elizabeth Almasi Doshi, "Open Data: Unlocking Innovation and Performance with Liquid Information", *McKinsey & Company,* October 2013, p. 17.

[55] Cited in EDPS, Preliminary Opinion of the European Data Protection Supervisor: on Privacy and Competitiveness in the Age of Big Data: The Interplay between Data Protection, Competition Law and Consumer Protection in the Digital Economy", March 2014, p.32. https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2014/14-03-26_competitition_law_big_data_EN.pdf

create entry barriers to SMEs and stifle innovation.[56] What this means is that open access policies and initiatives ought be adopted as much in the private sector as they are in the public sector. Furthermore, although "large data companies have no responsibility to make their data available, and they have total control over who gets to see it"[57], there exists an imbalance between the private and public sectors to the extent that private sector organisations have the significant advantage of their own data as well as open data provided by the public sector.[58] On the other hand, public sector agencies and organisations cannot innovate through access to big data held within the private sector. "In fact, open data encourages the commercialization of 'big data'".[59]

Moreover, a collaboration or partnership between both sectors has been recognised as vital to harnessing the benefits of the digital economy and in particularly stimulating growth and producing jobs in the EU digital market. For example, the Commission has proposed a Contractual Public-Private Partnership that "should develop incentives to share datasets between partners"[60] Thus, collaboration between the sectors may enable both parties to harness the benefits of open access policies when related to big data to produce even greater productivity and revenue. Thus, the potential of open access policies and initiatives in the private sector has not yet been realised.

Therefore, open access policies provide a unique opportunity for industry to explore. An integration of open principles such as transparency, community building and sustainable development promotes the premise that everyone should be able to use, adapt or re-distribute products and information, including those who provide and use it for commercial purposes. This subsequent use supports the contention that that data can be both open and profitable. However, the private sector is not yet engaging in big data on this level. This may be because of the perception of data as a valuable asset that must be protected, or because of the perceived risks of open access initiatives involving big data. These risks include risk of privacy invasive practices, potential intellectual property right infringements, and the risk that companies already equipped to deal in big data continue to amass greater amounts of data and also profit from big data related practices. The benefits and risks are discussed in detail in the case study examples below.

2.4 SUMMARY

A number of open access policies and initiatives have been implemented in the public sector and are motivated, to a large extent, by achieving a public good. On the other hand, there are fewer open access initiatives in the private sector, although companies and businesses are accessing public stores of data without providing access to their private data stores, which

[56] European Commission, Communication from the Commission to the European Parliament, the council, the European Economic and Social Committee and the Committee of the regions Towards a Thriving Data-Driven Economy, COM (2014) 442 final, Brussels, 2.7.2014, p. 3.
[57] Boyd and Crawford, op. cit., 2011, p.2.
[58] For a discussion of the benefits to businesses of publicly funded research see Parsons, David DR., Dick Willis and Dr. Jane Holland, *Benefits to the Private Sector of Open Access to Higher Education and Scholarly Research: A Research Report to JISC from Host Policy Research,* Host Policy Research, UK, 2011, PP.5-6. http://open-access.org.uk/wp-content/uploads/2011/10/OAIG_Benefits_OA_PrivateSector.pdf
[59] Holmwood, John, "Commercial Enclosure: Whatever Happened to Open Access?", *Radical Philosophy,* Sept/Oct 2013. http://www.radicalphilosophy.com/commentary/commercial-enclosure
[60] European Commission, Communication from the Commission to the European Parliament, the council, the European Economic and Social Committee and the Committee of the regions Towards a Thriving Data-Driven Economy, COM (2014) 442 final, Brussels, 2.7.2014, p. 7.

indicates an imbalance in the digital market. However, the potential for equal opportunity access to knowledge and information that can be built upon, developed, and that may subsequently lead to innovation and business opportunities in either the public or private sector. These benefits are accompanied by risks such as ethical and legal risks of invasions of personal information privacy. This can lead to discriminatory practices being levied against users. Further, smaller players in the market such as SMEs are at risk of being unable to develop big data policies simply because they lack the resources to keep up with organisations already implementing initiatives and capturing dominant market shares. However, should private sector stakeholders only make data available so long as it serves their perceived business and commercial aspirations, the true potential of open data in the private sector, and the opportunity of collaboration between public and private sector stakeholders, may not be fully realised.

3 OPEN ACCESS CASE STUDY EXAMPLES

3.1 OVERVIEW

Open access policies can be developed in relation to any types of large data sets to produce a multitude of benefits as seen in the following case study examples of health data, crisis data, energy data, environmental data, transport and logistics data, cultural data, and smart cities/utilities. For each case study below, we identify open access policies initiatives and business models, as well as who has access to the data, the extent to which that data is made open and the associated benefits or challenges raised by these policies and initiatives. These case studies represent relationships between open access policies across a variety of sectors, and they combine a number of public and private sector open access policies. However, these examples also highlight the varying degree of openness that is provided by public and private sector organisations, as well as the potential for collaboration between sectors.

3.2 HEALTH DATA

3.2.1 The relevance and importance of big health data to open access

Big health data may be made open by a number of stakeholders in the industry including patients, providers, insurers, and governments. These sources may store volumes of raw data about the health of patients, diseases and injuries, treatments used, and fees charged. As such, the amount of data being collected, analysed, and shared among health-care stakeholders is believed to have reached a great mass, with growing volumes of digitised medical records, aggregated research and development data, as well as data that governments have been collecting over the years. These big data sets are yielding critical insights into effective therapies for specific types of patients, enabling hospitals to isolate common causes of costly hospital readmissions and allowing insurers and other payers to identify variations in care delivery that add needless costs.[61] Health data is voluminous, varied and can be processed at high speed, and open access policies and initiatives in relation to these data are not new. Open access to health data undoubtedly yields a number of benefits for society at large, as well as individual stakeholders within society. However, at this stage, there are more examples of open access initiatives in the public sector that are providing commercial opportunities for private sector companies, than private sector companies making the data they hold open. This imbalance can hinder the potential for greater uses and re-use of big data in the health care sector.

3.2.2 Examples of policies and initiatives relating to health data

Big health data is recognised as the single most valuable asset for healthcare organisations in terms of providing the basis for decision-making. It is suggested that the insights that can be gained from health care data may serve as one of the most compelling drivers for big data and analytics, even in its unstructured form.[62] This is because open access policies for big health data produce a number of benefits such as the identification of clinical treatments,

[61] Manyika, James, Michael Chui, Peter Groves, Diana Farrell, Steve van Kuiken and Elizabeth Almasi Doshi, "Open Data: Unlocking Innovation and Performance with Liquid Information", *McKinsey & Company,* October 2013, p. 89.

[62] Burghard, Cynthia, "Big Data and Analytics Key to Accountable Care Success", *IDC Health Insights Industry Brief*, October 2012.
http://public.dhe.ibm.com/common/ssi/ecm/en/iml14338usen/IML14338USEN.PDF

pharmaceutical responses, and public health interventions that may not be produced by traditional research methods, or at least, not at the same speed. In fact, predictive medicine is becoming reliant upon bio-repositories that link genomic data to health care data. Although, information relating to health remains a private part of our lives, big data enables more powerful discoveries. These examples also highlight how negative externalities such as the potential compromise of privacy rights and data protection may be mitigated.

An example of one such beneficial initiative is the GOSgene health initiative. Whilst GOSgene is a focus of Work Package 3, it requires brief mention here as it embodies the positive outcomes of open access initiatives in the area of health data. GOSgene initiative involves working with congenital disorders to enable the discovery of new genes, the identification of disease and innovation in health care. At the core of this open access initiative is the altruistic view that health care will be improved by developing the information made available through the open access initiatives. As the National Institute funds the GOSgene project for Health Research (NIHR), it is undertaken in accordance with the NIHR Open Access Policy, which has been in place since 2006:

> The Department of Health (DH) and the National Institute for Health Research (NIHR) agrees with the other main biomedical research funders with the principal of Open Access to the outputs of its research. In doing so the DH recognises that open access to the outputs of its research can offer both social and economic benefits as well as aiding the development of new research and stimulating wider economic growth of the UK economy.[63]
>
> The Government, in line with its overarching commitment to transparency and open data as well as part of its economic growth strategy, is committed to ensuring that published research findings should be freely accessible. As a publically funded research body, the NIHR is subject to the requirement to make the outputs from its research publicly available – not just to other researchers, but also to potential users in business, charitable and public sectors, and to the general tax-paying public.[64]

This open access initiative has produced undeniable benefits such as the discovery of a gene related to blindness in children.[65] The initiative helps improve diagnostic testing, supports genetic counselling, and will guide further functional analysis aimed at understanding the pathogenesis of disease and improvements in patients' management of disease or developing novel therapies. Open access to this material and the benefits can only encourage further discoveries and treatments.

Two other pertinent examples in the life sciences field are the Genbank and the Protein Data Bank. The publicly funded Protein Data Bank is "a repository for the three-dimensional structural data of large biological molecules, such as proteins and nucleic acids".[66] The Protein Data Bank usage policy provides: "Data files contained in the PDB archive are free of

[63] NIHR, "Open Access policy Statement", *NIHR*, *no date.* http://www.nihr.ac.uk/research/Pages/Research_Open_Access_Policy_Statement.aspx

[64] Ibid.

[65] Kelberman, Daniel, Lily Islam, Jorn Lakowski, Chiara Bacchelli, Estelle Chanudet, Francesco Lescai, Aara Patel, Elia Stupka, Anja Buck, Stepehn Wold, Philip L. Beales, Thomas S Jacques, Maria Bitner-Glindzicz, Alki Liasis, Ordan J. Lehmann, Jurgen Kohlhase, Ken K. Nischal and Jane C Snowden, "Mutation of SALL2 causes recessive ocular coloboma in humans and mice", *Human Molecular Genetics,* Oxford Journals, 17 April 2014. http://hmg.oxfordjournals.org/content/early/2014/01/12/hmg.ddt643

[66] RCSB Protein Data Bank, "Policies and References", *RCSB*, no date. http://www.rcsb.org/pdb/static.do?p=general_information/about_pdb/policies_references.html

all copyright restrictions and made fully and freely available for both non-commercial and commercial use. Users of the data should attribute the original authors of that structural data."[67] This unrestricted access to large data sets in health sciences are recognised as producing beneficial results, such as enabling the Genome project:

> The success of the genome project, which is generally considered to be one of the great scientific achievements of recent times, is due in no small part to the fact that the world's entire library of published DNA sequences has been an open-access public resource for the past 20 years. If the sequences could be obtained only in the way that traditionally published work can be obtained, that is, one article at a time under conditions set by the publisher, there would be no genome project. The great value of genome sequences would be enormously diminished.[68]

The data are submitted by biologists and biochemists from around the world and are freely accessible on the Internet via the websites of its member organisations such as the Protein Data Bank in Europe ("PDBe").[69] Further, EMBL-EBI's PDBe OpenEye Scientific Software have partnered to integrate innovative cheminformatics solutions into PDBe's publicly available resource for studying 3-D cellular and macromolecular structures.[70] Thus, open access policies may result in commercial ventures that would not otherwise be possible such as collaborations between data software companies in the private sector and public sector funded research institutes.

A French initiative, TeraLab, is an open access initiative that combines technical infrastructure and software tools for utilisation in the e-health sector. TeraLab's focus on a secure platform means that sensitive data such as health data may be processed through this solution and in accordance with the relevant French privacy laws. Thus, the particular risks associated with open access to health data can be effectively mitigated so that positive externalities resulting from open access to this type of data may be captured. For example, TeraLab enables data matching, and longitudinal studies[71], and supports the storage of all available health data relating to each individual under a unique identifier in a central database. Currently, the health database produces approximately 250 terabytes of data annually, and includes more than 1.2 billion entries with more than a thousand variables. This data is supported to produce real-time analysis in an "ultra-high security" framework.[72] Thus, this initiative and others like it, provide a number of opportunities for public and private stakeholders not only in terms of diagnostics but also as a source of innovation.

There are a number of open access initiatives in the public sector. For example, In the United Kingdom, data made public through the National Health Services (NHS) Choices program allows patients to compare hospital ratings and to review qualitative reviews of doctors. Another example of an open access policy to health data is found in the Danish Basic Data

---

[67] Ibid.

[68] Organisation for Economic Cooperation and Development Committee for Information, Computer and Communications Policy Working Party on the Information Economy, "Digital Broadband Content" DSTI/ICCP/IE(2004)11/FINAL, September 2005, p.60 citing Brown, P., in Committee on Electronic Scientific, Technical, and Medical Journal Publishing (CESTMJP) (2004), Electronic Scientific, Technical, and Medical Journal Publishing and Its Implications: Proceedings of a Symposium, National Research Council, National Academies Press, Washington DC, p30.

[69] "Protein Data Bank", *Wikipedia no date. http://en.wikipedia.org/wiki/Protein_Data_Bank*

[70] Ibid.

[71] Gadouche, Kamel, 'TeraLab, A Secure Big Data Platform: Description and Use Cases', *European Data Forum Presentations,* 2014. http://2014.data-forum.eu/edf2014-presentations

[72] Ibid.

Programme, which provides a study in open access to health care data. This programme is currently delivered by the Danish Agency for Digitisation, in association with the Danish Ministry for Finance. The programme is scheduled to conclude in 2017 and to date, the Danish government has invested €125 million in the programme in anticipation of resultant benefits such as efficiency gains and cost savings.[73] This programme means that all common health care basic data will be distributed by the "'Data Distributor." This data is the subject of a '[…] Worldwide, free, non-exclusive, and otherwise unrestricted licence to use the data."[74] This licence required to be exercised in accordance with all Danish law. Although this health data is made open as a result of public sector open access policies and initiatives, both private and public sector actors, despite an unequal contribution from both sectors, capture the benefits. However, it is likely that open big data health initiatives can produce greater benefits when more stakeholders are involved, such as intergovernmental initiatives. The European Molecular Biology Laboratory "is an intergovernmental organisation specialising in basic research in the life sciences with 21 member states, one prospect and two associate member states."[75] EMBL is funded by its members.[76] Thus, such initiatives are common in the public sector and reflect the motivation of achieving a public good through the provision of open access. However, similar open access initiatives in the private sector are viable, and the opportunities for industry, privately, and or collaboration with the public sector.

In the private sector, Swedish software company, ESTeam AB, considers that a Europe-wide E-Heath information Infrastructure initiative is viable. This initiative would make medical data such as meanings, actions and patients, accessible.[77] This would be achieved through merging multiple technologies in the same language into one unified source. This information infrastructure would open the domain to governments and software developers. This infrastructure is predicted to benefit SMEs by enabling them to build targeted solutions that reach a bigger audience.[78] This infrastructure fosters the development of the EU single market by providing data interoperability.[79]

Relevantly, for over a decade, policymakers have sought to expand public access to information about planned, on-going and completed clinical trials.[80] "Recently, several initiatives have broadened the focus from the registering of trials to the sharing of protocol details, enrolment opportunities, and study results, and now to providing access to participant-level data."[81] A number of these initiatives have taken place in the US. One important example of access to large raw data sets in the area of clinical trials is the Yale University Open Data Access ("YODA") Project.[82] The YODA project has developed a model to facilitate access to participant-level clinical research data. This process also includes making participant-level clinical research data available for analysis for external investigators in

[73] Horst, Nicolas Lemcke, 'Danish Basic Data', *European Data Forum Presentations,* 2014. http://2014.data-forum.eu/edf2014-presentations
[74] Ibid.
[75] European Molecular Biology Laboratory. *EMBL,* no date. http://www.embl.de/aboutus/index.html
[76] Ibid.
[77]Magnusdottir, Gudrun, 'Interoperability Solutions', *European Data Forum Presentations,* 2014. http://2014.data-forum.eu/edf2014-presentations
[78]Ibid.
[79]Op. Cit., 77.
[80] Mello, Michelle M, J.D., Jeffrey. K. Francer, J.D., m.p.p., Marc Wilenzick, J.D., Patricia Teden, M.B.A., Barbara E. Bierer, M.D., and Mark Barnes, J.D., LL.M, "preparing for Responsible Sharing of Clinical Trial Data" *The New England Journal of Medicine,* Vol. 369;17, October 2013, p.1651. http://www.nejm.org/doi/pdf/10.1056/NEJMhle1309073
[81] Ibid.
[82] Yale School of Medicine, "YODA Project", no date. http://medicine.yale.edu/core/projects/yodap/index.aspx

pursuit of the overall mission of the project to promote open science by promoting the sharing of clinical research to advance science and improve public health and healthcare; promote the responsible conduct of research; ensure good stewardship of clinical research data; and protect the rights of research participants.[83] YODA Project is an effort by a group of academically-based clinical researchers to facilitate access to participant-level clinical research data and/ or comprehensive reports of clinical research, such as full Clinical Study Reports. This model is based on the facilitation of open access as an intermediary but does not provide access to data owned or held directly by Yale. For example, in 2011, YODA reached an agreement with medical device maker Medtronic to act as an intermediary for releasing all data on clinical trials of a controversial bone-growth protein whose safety had been questioned.[84] In an effort to defend its reputation, the company gave up any right to decide who would get the information. YODA then commissioned two systematic reviews of the protein, which conveyed mixed results that were then published. Following Medtronic's example, Johnson & Johnson pledged in January to make all its clinical trial data available for perusal by outsiders through YODA.[85] This directly responds to the practise of Pharmaceutical companies frequently withholding the results of negative or inconclusive trials.[86] The effect of this type of open access model means that doctors can provide patients with more complete information pertaining to their illnesses and treatments, and regulators are not at risk of approving medications that have hidden health hazards, and prevents drugs from going to market with critical safety data kept secret.[87] The benefits associated with open access to clinical trial results include: "Independent scientists could re-analyze data to verify the accuracy of reports prepared by trial sponsors, which might deter sponsors from mischaracterising or suppressing findings. Data sharing would also allow analysts both within and outside drug companies to pool data from multiple studies, creating a powerful database for exploring new questions that can't be addressed within any given trial because the sample is too small to support such analyses."[88]

The European Medicines Agency has also considered the openness of clinical trial data and has released a draft policy for the provision of public access to some of the data.[89] However, this has resulted in commercial litigation brought by objecting companies who claim their data is confidential information.[90] The EMA's updated policy has expanded its policy require disclosure of certain raw data, clinical study reports and individual case-report forms. The EMA confirmed that in general, clinical trial information is not confidential information. In

[83] Yale University Centre for Outcomes Research and Evaluation, *Yale university Open Data Access (YODA) Project Policy to guide External Investigator Access to Clinical Trial Data,* July 2014, p.1. http://medicine.yale.edu/core/projects/yodap/datasharing/Janssen/463_194023_YODAProjectDataReleasePolicyJuly2014.pdf

[84] The Editors, "Secret Clinical Trial Data to Go Public", *Scientific American,* 1 June 2014. http://www.scientificamerican.com/article/secret-clinical-trial-data-to-go-public/?mobileFormat=true

[85] Ibid.

[86] Op. Cit., 84.

[87] Clinical trials of GlaxoSmithKline's diabetes drug Avandia (rosiglitazone) and Merck's anti-inflammatory Vioxx (rofecoxib) revealed an elevated cardiac risk from the drugs, but relevant findings were held back from regulators or never published.

[88] Mello, Michelle, "Should There be Public Access To Data From Clinical Trial" *Human Capital Blog,* 16 June 2014. http://www.rwjf.org/en/blogs/human-capital-blog/2014/06/should_there_be_publ.html

[89] Mello, Michelle M, J.D., Jeffrey. K. Francer, J.D., m.p.p., Marc Wilenzick, J.D., Patricia Teden, M.B.A., Barbara E. Bierer, M.D., and Mark Barnes, J.D., LL.M, "Preparing for Responsible Sharing of Clinical Trial Data" *The New England Journal of Medicine,* Vol. 369;17, October 2013, p.1651. http://www.nejm.org/doi/pdf/10.1056/NEJMhle1309073

[90] AbbVie, Inc. v. European Medicines Agency, Case T-44/13. Order of the President of the General Court (Apr. 25, 2013); InterMune UK Ltd. v. European Medicines Agency, Case T-73/13. Order of the President of the General Court (Apr. 25, 2013).

the private sector, GlaxoSmithKline have adopted policies allowing researchers access to participant level data from trials of approved products.[91]

3.2.3 The benefits of open access to health data

Open access to big health data produces a number of benefits for patients or consumers, as well as public and private health organisations. For example, the use of data analytics on data from thousands (or millions) of patients leads to the discovery of effective treatments, and using these insights to ensure that patients get the most timely and appropriate treatments.[92] Other positive impacts of open access to health data include enabling individuals to take an active role in disease prevention and treatment; matching patients with the most appropriate providers; ensuring the cost effectiveness of care; and identifying new therapies and approaches to delivering care. In terms of creating commercial opportunities and adding value to economies, open access policies and initiatives relating to big health data may raise the productivity of R&D by makers of drugs and medical devices sharing data (where possible), using open data to find patients for trials, speeding products to market, and spotting potential problems once products are in use.[93] However, capturing this value will require changes in how care is delivered and paid for, standards for data governance and usability, persuading providers to share data, and investing in the capabilities of all players to make the most of open data.[94]

Another benefit of open health data is greater transparency. Transparency assists health care professionals and patients in making more informed decisions. As mentioned above, this is especially important when open health data means open access to the results of clinical trials were not traditionally published, or misleadingly, the positive outcomes of clinical trials were published and the negative results were omitted. With greater transparency of data, there could be greater accountability for design, conduct, analysis, and reporting of clinical trials and more vigorous monitoring of products over their life cycle.[95] This benefit has been seen through the YODA Project discussed above.

There are also a number of commercial opportunities for private organisations that have the opportunity to capitalise on health data made available (predominantly in the public sector) in order to develop a variety of open health data products such as health and exercise monitoring devices and applications. However, in that context intellectual property rights of the original data creators must be respected.

3.2.4 Negative impacts associated with open access to big health data

Despite the obvious benefits of making health data open, open access policies relating large sets of health data also produce risks to personal rights, particularly as the majority of related

[91] GlaxoSmithKline. Clinical study register (http://www gsk-clinicalstudyregister.com).
[92] Manyika, James, Michael Chui, Peter Groves, Diana Farrell, Steve van Kuiken and Elizabeth Almasi Doshi, "Open Data: Unlocking Innovation and Performance with Liquid Information", *McKinsey & Company,* October 2013, p. 79.
[93] Ibid.
[94] Op. Cit., 2013, p. 22.
[95] Institute of Medicine, *The Future of Drug Safety: Promoting and Protecting the Health of the Public*. Washington, DC, National Academics Press, 2006; Institute of Medicine, *Ethical and Scientific Issues in Studying the Safety of Approved Drugs*, Washington, DC., National Academics Press, 2012 cited in Mello, Michelle M, J.D., Jeffrey. K. Francer, J.D., m.p.p., Marc Wilenzick, J.D., Patricia Teden, M.B.A., Barbara E. Bierer, M.D., and Mark Barnes, J.D., LL.M op.cit., 2013, p.1653.

data consists of private, and at times, sensitive information amounting to 'sensitive personal data' under the Data Protection Directive 95/46/EC.[96] Health data becomes highly sensitive data, especially when it is connected to an identifiable person. This may result in a breach of data protection and privacy rights. As this sensitive information is increasingly utilised for the benefit of patients and the wider public, especially through open access initiatives, organisations dealing with this information will need to continually update their internal data protection and privacy policies and procedures. There is also some concern over whether the privacy of research participants can be guaranteed, and even if it were so, whether effectively de-identifying data renders it useless.[97] This may occur if identifiers such as sex, age and geographic location are relevant and give context to the raw data. Another issue with de-identified that relates to genome data is that it can easily lead to re-identification when coupled with other publicly available information.[98] It may be necessary to accept limitations in potential big data discoveries given the risks to the privacy of data subjects. Other negative externalities relate to the risk that open data could be used to perpetuate abuses that exist in current systems, and/ or open data can provide information on which to build malpractice suits.[99] Further, open access to data could result in a reduction of private funding for fear that the potential gain is diminished if competitors have a right of access to scientific and commercial strategies. Similarly, open access can hinder potential patents when disclosed data poses an obstacle to obtaining patents on further claims. Legally, this early disclosure could be considered prior art under the relevant patent law and invalidate a patent application. Further, premature publication can trigger the initiation of the period of patent protection and data exclusivity that the inventor enjoys.[100] Thus, mandating open access to health data ought come with some safeguards for the prospective patents that would have developed from the data. Alternatively, patent protection must be encouraged to foster voluntary open access to health data.

The claim that health data, particularly raw data like participant trial clinical data is confidential commercial data akin to trade secrets has been the basis for the a refusal to allow access to that data. However, this issue is being dealt with in America where a policy[101] is proposing that pharmaceutical companies release de-identified participant-level data that are pooled within a product class and masked so that hey do not identify particular products. Thus, it is possible to provide open access whilst retaining its profitability and product identity that means the product designer and inventor retains its competitive advantage.

---

[96] Article 8 Data Protection Directive 95/46/EC.

[97] Mello, Michelle M, J.D., Jeffrey. K. Francer, J.D., m.p.p., Marc Wilenzick, J.D., Patricia Teden, M.B.A., Barbara E. Bierer, M.D., and Mark Barnes, J.D., LL.M,, op.cit., 2013, p.1653.

[98] Malin B, Karp D, Scheuermann RH., "Technical and Policy Approaches to Balancing Patient Privacy and Data Sharing in Clinical and Translational Research" *J Investig Med*, *Vol. 58(1),* January 2010, pp.11-8. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2836827/

[99] Manyika, James, Michael Chui, Peter Groves, Diana Farrell, Steve van Kuiken and Elizabeth Almasi Doshi, "Open Data: Unlocking Innovation and Performance with Liquid Information", *McKinsey & Company,* October 2013, p. 86.

[100] For a more detailed description of legal and regulatory issues associated with pen access to raw data such as clinical trial data see Mello, Michelle M, J.D., Jeffrey. K. Francer, J.D., m.p.p., Marc Wilenzick, J.D., Patricia Teden, M.B.A., Barbara E. Bierer, M.D., and Mark Barnes, J.D., LL.M, "preparing for Responsible Sharing of Clinical Trial Data" *The New England Journal of Medicine,* Vol. 369;17, October 2013, p.1651. http://www.nejm.org/doi/pdf/10.1056/NEJMhle1309073

[101] Food and Drug Administration. Availability of masked and de-identified non-summary safety and efficacy data; request for comments. June 4, 2013 https://www.federalregister.gov/articles/ 2013/06/04/2013-13083/availability-of-masked-and-de-identified -non-summary-safety-and-efficacy-data-request-for-comments? utm_campaign=subscription+mailing+list&utm_medium=email &utm_source=federalregister.gov.

3.2.5 Summary

Open health data produces a number of positive outcomes for public and private sector stakeholders. However, it appears that open access to large volumes of health data are provided by the public sector organisations, and being utilised for commercial gains in the private sector stakeholders. Nevertheless, the public sector is harnessing benefits of open access in terms of the provision and understanding of health care but collaborations between the sector organisations may yield greater benefits for both. As big health data presents obvious risks in term of personal information privacy of sensitive data, these risks can be mitigated with relevant technologies. Open access to big health data provides untapped business and public health opportunities for stakeholders.

3.3 CRISIS DATA

3.3.1 The relevance and importance of big crisis data to open access

Large data sets evolve as a result of the information published during crisis situations. This crisis data can create value for communities and citizens, as well as companies and their consumers, especially when that data are utilised in open access initiatives. There are a number of open access initiatives that assist in humanitarian efforts, such as crisis mapping, or that disseminate vital information about crisis situations, such as social media networks. Crisis mapping and the role social media networks play during times of crisis reflect the sentiment and necessity of open access initiatives involving big crisis data:

> Democratized access to incident-level information will empower global technologists to redefine the relationship between data and crisis. Crisis data is the most important and most urgent information produced by the digital age. People's lives are at stake, it's time to be open, work smarter, and think bigger.[102]

This report first examines crisis mapping as an open access initiative developed primarily in connection with humanitarian efforts during crises. Such initiatives produce a number of societal benefits during and immediately following a crisis situation. These benefits are realised because of collaborations between crisis mapping initiatives and private sector actors such as Google Inc. that provides access to its vast stores of geo data through Google Maps. Such initiatives are championed by international organisations. In particular, "The United Nations Global Pulse, a digital innovation initiative, supports the idea of 'data philanthropy' to convince corporations to make anonymized versions of their data available for use in crises and emergencies."[103] Second, this report examines the role played by social media, in particular Twitter Inc., in the dissemination of information and news during crises to inform and assist members of the public by providing real time updates, as well as providing a platform for services and companies to engage with citizens and consumers, and for citizen to citizen engagement. Thus, crisis mapping and social media networks reflect the potential usefulness of open and free access to crisis data for governments, citizens, journalists, health care workers and an array of other related public service and public interest organisations.

---

[102] Morgan, Jonathon, "The Future of Crisis Data", *Ushahidi,* 20 May 2014. http://www.ushahidi.com/2014/05/20/the-future-of-crisis-data/

[103] United nations Office for the Coordination of Humanitarian Affairs ("OCHA"), "Humanitarianism in the Network Age", *OCHA Policy and Studies Series, 2012, p. 29. https://docs.unocha.org/sites/dms/Documents/WEB%20Humanitarianism%20in%20the%20Network%20Age%20vF%20single.pdf*

Furthermore, crisis mapping and the use of social media during crises also highlight when information models can be both open and profitable, as well as achieving a public good. In terms of generating profit, open access initiatives involving crisis data provide new avenues for revenue to companies implementing them, and increase notoriety of organisations implementing them. Crisis situations tend to cause a spike in the number of people accessing these initiatives and subsequently, the volume of user generated content. This in turn provides commercial opportunities to social media companies and crisis mapping organisations to the extent that they can capture and make use of the increased amounts of personal and behavioural data that becomes available during this peak time. Thus, the opportunities created by open access initiatives involving crisis data are increased at the time of crisis because of the rate and volume of which the data is disseminated.

The very reason that social media platforms and crisis mapping are useful during times of crisis is because they are open access. Further, because these open access initiatives deal specifically with crisis data, some crisis mapping platforms need to be available when general infrastructure is disrupted by a crisis event. This can lead to further opportunities for technological innovation and infrastructure to support such models. Therefore, crisis mapping and the role of social media during crisis situations are indicative of finding ways to make big data useful to humanitarian decision makers and is considered to represent one of the great challenges, and opportunities, of the network age.[104]

Nevertheless, crisis mapping and the role of social media during crisis situations implicate important issues that reflect negative externalities and require understanding so that they can be minimized in order for the positive externalities to be fully captured. Open access initiatives involving big crisis data have the potential to compromise security and privacy, as well as potentially hamper emergency efforts when incorrect information is published. This is a particular risk with crisis data disseminated through social media when the reliability and validity of the data is not certain, or when graphic data including pictures and videos are published without the consent of those captured in that footage.

Despite these potential negative externalities, such open access initiatives provide good examples of the beneficial relationship between open access and big crisis data, and is an important area of growth. The latter is particualry relevant as there exist few examples of European led open access initiatives involving crisis data. This issue requires provides opporuntity to Euoropean big data ators:

> Whether it's social media, sensor data, crowd sourced or crowd-seeded reporting, or manually maintained databases exposed over web APIs, we need to mandate that crisis data be available immediately and continuously, and build our systems to accommodate that necessity.[105]

Ultimately, open access to big crisis data is a potenital area of innovation for European businesses and organisations. However, innovators need to remain mindful of minimising the negative implications that can flow from open access to crisis data, as illuminated below.

---

[104] United nations Office for the Coordination of Humanitarian Affairs ("OCHA"), "Humanitarianism in the Network Age", *OCHA Policy and Studies Series, 2012, p. 26. https://docs.unocha.org/sites/dms/Documents/WEB%20Humanitarianism%20in%20the%20Network%20Age%20vF%20single.pdf*

[105] Morgan, Jonathon, "The Future of Crisis Data", *Ushahidi,* 20 May 2014. http://www.ushahidi.com/2014/05/20/the-future-of-crisis-data/

3.3.2 Examples of open access initiatives utilising big crisis data

Crisis mapping

Crisis mapping is an emerging open access concept involving crisis data, intermixed with big geo data. Crisis mapping produces an array of public benefits, from enabling citizens to identify and stay away from crises and affected areas, as well as assisting public and private organisations that play a number of humanitarian roles during times of crisis. These benefits are realised because the data is made open access and is updated in real time:

> today, technology once limited to experts and institutions is available to anyone. This has allowed groups of self-organizing volunteers to place SMS messages and social media postings on dynamic maps, highlighting clusters of cries for help in an earthquake, or identifying where roads have been washed away after a flood.[106]

Further, the United Nations Office for the Coordination of Humanitarian Affairs defines crisis mapping as, "A process of sourcing, visualizing and analysing data from a humanitarian, political, ecological, financial or other crisis, often in real time, on a dynamic, interactive map."[107] Geographic Information Systems, combined with SMS and open-source data-sharing platforms, have facilitated the phenomenon of crisis mapping popularised by groups such as Ushahidi (discussed below). Thus, crisis mapping is an open access initiative that depends upon the accessibility of big geo spatial data via tools such as Google Maps that provide open access to interactive maps based on algorithms that can be updated in real time. This means that crisis mapping is also a strong example of effective cross-sector collaboration for the provision of open access.

There are a number of recognised and obvious benefits that flow from crisis mapping for citizens, individuals and private and public organisations when assisting with, and raising awareness of disasters and the associated issues that arise during such a time. At a fundamental level, it enables logistical planning of emergency efforts and the identification of areas and communities that require assistance and support. In fact, crisis mapping is so valuable that the United Nations Office for the Coordination of Humanitarian Affairs recommended in 2012 that humanitarian organisations develop standards and best practices for communication mapping in emergency response by 2015.[108] Further, crisis mapping utilises crowdsourcing techniques to facilitate technical or information management tasks, such as mapping or geo-tagging, are outsourced to a "crowd" of volunteers that can live anywhere.[109] This relationship between crisis mapping and crowdsourcing can lead to new opportunities, innovation and development. For example, Standby Task Force, which comprises over 1,000 volunteers in 80 countries, provides specialist support to humanitarian responders in a number of ways, least of all, by providing geo-locating by way of collaboration. Nevertheless, crisis mapping is not without some negative implications, especially when it implicates issues such as privacy and security that arise when factually incorrect data forms part of the crisis data made available through open access initiatives, or jeopardises the safety of citizens, and or their right to privacy, at a time when they are most vulnerable.

---

[106] United nations Office for the Coordination of Humanitarian Affairs ("OCHA"), "Humanitarianism in the Network Age", *OCHA Policy and Studies Series, 2012, p. 28. https://docs.unocha.org/sites/dms/Documents/WEB%20Humanitarianism%20in%20the%20Network%20Age%20vF%20single.pdf*

[107] Ibid., *p. 61.*

[108] OCHA, op. cit., 2012, p. 61.

[109] Ibid*., p. 26.*

The following examples of crisis mapping illuminate when big crisis data and open access produce both positive and negative externalities, as a stand alone initiative, and when in collaboration with private partners such as Google Maps.

Ushahidi, Inc.

Ushahidi provides open source software that deals primarily with crisis data. This example strongly reflects the humanitarian focus of crisis mapping, which is in itself an overarching positive externality. As with most examples of crisis mapping, Ushahidi employs the concept of crowdsourcing for social activism and public accountability, serving as an initial model for what has been coined "activist mapping"—the combination of social activism, citizen journalism and geospatial information.[110] Ushahidi evolved simultaneously with the concept of activist mapping when the platform was first developed to map reports of violence in Kenya after the post-election fallout at the beginning of 2008.[111] During that time, eyewitness reports of violence reported by email and text message were placed on a Google map.[112] This initial effort is outlined in the Ushahidi mission statement:

> The original website was used to map incidents of violence and peace efforts throughout the country based on reports submitted via the web and mobile phones. This website had 45,000 users in Kenya, and was the catalyst for us realizing there was a need for a platform based on it, which could be used by others around the world.[113]

Since its initial use, Ushahidi's open source software has enabled the mapping of incidents and global efforts by encouraging: "information collection, visualisation, and interactive mapping".[114] The platform maintained by Ushahidi utilises Google Maps to perform the geospatial aspect of its crisis data management. It allows users to mark events on a map that changes in near real time, and which is accessible online. This creates a picture of what is happening, when and where, so it can help responders make decisions. Another benefit of crisis mapping platforms, such as Ushahidi, is that it is made accessible by way of implementing "simplified the technology so that anyone can use it, and it is designed to take input from hundreds of people by cell phone or e-mail. It uses free software called FrontlineSMS that turns a laptop and a mobile phone into a text-broadcasting hub. As an SMS is sent from a hot zone, the message synchs with the Ushahidi software and shows up in a Web administrator's in-box. The Web admin can decide to send a text message back to the sender to verify the information, send out a blast alert to large numbers of people or post the information onto a Web page with location information from Google Maps (or do all three)."[115] Thus, crisis mapping can spur on technological innovation and create avenues for development and growth through the demand for specific technology and digital infrastructure. Further, the involvement of a text administrator is also beneficial in mitigating potential negative implications of the inclusion of false information that could hamper emergency responses.

[110] "Ushahidi", *Wikipedia,* no date. http://en.wikipedia.org/wiki/Ushahidi
[111] Ushahidi, "Mission", no date. http://www.ushahidi.com/mission/
[112] Bahree, Megha, "Citizen Voices", *Forbes,* 29 November 2008. http://www.forbes.com/global/2008/1208/114.html
[113] Ibid.
[114] "Ushahidi", *Wikipedia,* no date. http://en.wikipedia.org/wiki/Ushahidi
[115] Bahree, Megha, "Citizen Voices", *Forbes,* 29 November 2008. http://www.forbes.com/global/2008/1208/114.html

Ushahidi gained further notoriety as a useful and valuable open access initiative during the 2010 Haiti earthquake disaster. During that crisis, Ushahidi launched a crisis map that provided a visual representation of real-time reports and information generated by the affected community (via sms/text and the internet). The information consisted of reports relating to the damage, requests for assistance, and the establishment of relief centres. The concept of openness is central to this initiative and was further developed in line with the interoperability during the Haiti earthquake. Platforms such as Ushahidi have spurred the consideration of interoperability to support innovative and socially valuable platforms such as Ushahidi that depend on an extensive degree of openness. This represents another positive externality of open access initiatives that utilise big crisis data, as outlined on the ushahidi blog:

> the software systems we build need to be equally as flexible, resilient, and open. As technologists it's our job to empower actors local to the crises they're experiencing with unopinionated APIs and data processing pipelines that handle the heavy lifting of real-time data science without the heavy-handed perspective of direct analysis.[116]

The success of this open access initiative as a tool for humanitarian efforts has also lead to the creation and consideration of further technological developments, including interoperability across sectors to support the intermix of geo data and crisis data, and ultimately, support it as an access initiative.

Ushahidi is perhaps one of the better-known open access initiatives involving crisis data. This is largely because of its reach and the high levels of engagement that assisted humanitarian organisations and governments address crises, including unrest following the Kenyan election crisis in 2008, and the 2010 Haiti earthquake. This example also highlights the importance of cross – sector collaboration as it relies on the accessibility of big geo data.

Other examples of crisis mapping

Whilst crisis mapping is still very much in its infancy, there are a number of other examples of crisis maps that illuminate societal externalities of this open access initiatives. However, depth of detail about their effectiveness has not yet been recorded, which makes an analysis of their potential value difficult. Nevertheless, such open access initiatives have been implemented during times of crises as a humanitarian aid, despite the potential for crisis mapping to infringe privacy or threaten security of those already vulnerable as a result of the crisis that is the subject of the map. Looking to the future, crisis mapping also presents commercial opportunities for companies developing crisis mapping platforms alone, and in collaboration with other big data actors.

A crisis map was created following the March 2011 tsunami in Japan. That map resulted in almost 9,000 reports in the first few weeks, and it was still being updated in October 2012. This was a success story in terms of engagement, but there was no evaluation of the impact of the map in terms of more rescues or more efficient resource allocation.[117] However, this level of engagement is indicative of the potential growth of these platforms as a viable open access initiative, as well as a potential source of data collection for organisations implementing the initiatives. The latter is important because it highlights that such models can be open and profitable in light of the vast amounts of data that are created as a result of implementing this

[116] Morgan, Jonathon, "The Future of Crisis Data", *Ushahidi,* 20 May 2014. http://www.ushahidi.com/2014/05/20/the-future-of-crisis-data/

[117] United nations Office for the Coordination of Humanitarian Affairs ("OCHA"), "Humanitarianism in the Network Age", *OCHA Policy and Studies Series, 2012, p. 31. https://docs.unocha.org/sites/dms/Documents/WEB%20Humanitarianism%20in%20the%20Network%20Age%20vF%20single.pdf*

type of open access initiative during a time when it is in high demand. This may also prompt more private sector organisations to pursue crisis mapping, which is primarily thought of in relation to non-profit organisations. Thus, crisis mapping is an area ripe for innovation and growth, especially in terms of fostering collaborations between public, private and non profit organisations. The crisis map developed during the 2011 tsunami was a product of volunteers across the world combining data from sources such as satellite maps, World Health Organization maps of health facilities, and locations of police facilities from the Pacific Disaster Center, and data from mobile phone companies. This map became known as the OpenStreetMap project, and it became a critical source of reliable information to guide both governments and private aid workers and supplies to hospitals, triage centers, and refugee camps. This map helped responders better match the supply and demand for various resources in the aftermath of the disaster, significantly improving the services delivered.[118]

Another example of crisis mapping involves USAID, SBTF, GISCorps and several private-sector companies who collaborated to launch an exercise in June 2012 that compared the work of volunteers with the results of an automated process. Their task was to clean and map data highlighting the locations of loans made by private banks in developing countries. The data set had originally been mapped at the national level, but more detailed geographic information was available. The aim was to create a more precise map—a common need in humanitarian response, where a few miles can make all the difference.[119] The volunteers succeeded. According to the USAID case study of the process, the automated geocoding process "refined 66,917 records at 64 per cent accuracy", while the crowdsourcing process "refined an additional 7,085 records at 85 per cent accuracy."55 It was also fast: the entire project was completed in just 16 hours, which is 44 hours earlier than projected.[120] This provides an example of collaboration between private and public sector organisations and indicates that there is potential value for private companies engaging in crisis mapping open access initiatives. This is relevant because crisis mapping has its roots in humanitarian causes and is more closely aligned with non-profit organisations and public humanitarian efforts. However, such initiatives can benefit from the collaboration across sectors.

In addition, the Standby Task Force promotes a crisis mapping model reliant on real time updating of maps that provide information pertaining to crisis situations around the world in the pursuit of what the Standby Task Force refers to as "crisis mapping for humanitarian response."[121] The concept for the Task Force was launched at the 2010 International Conference on Crisis Mapping. The Standby Task force's most recent involvement in Ebola epidemic[122] also identifies the societal benefits flowing from crisis mapping as it enables citizens to be informed about affected areas, as well as identify access routes for emergency and health care workers. These are just some of the positive externalities of crisis mapping.

A final example of crisis mapping is Harvard's HealthMap that came into the spotlight after reporting the Ebola crisis earlier than the World Health organisation:

---

[118] Op. Cit., 2013, p. 21.

[119] United nations Office for the Coordination of Humanitarian Affairs ("OCHA"), "Humanitarianism in the Network Age", *OCHA Policy and Studies Series, 2012, p. 31. https://docs.unocha.org/sites/dms/Documents/WEB%20Humanitarianism%20in%20the%20Network%20Age%20vF%20single.pdf*

[120] Ibid.

[121] Standby Task Force, "About", no date. http://blog.standbytaskforce.com/about-2/

[122] For more information about the Standby Task Force involvement with crisis data during the 2014 Ebola epidemic, see Per, "Standby Task Force Activates to Support NetHope During Ebola Outbreak", *Standby Task Force Blog,* 27 August 2014. http://blog.standbytaskforce.com/category/all-posts/

> Much of the coverage of HealthMap's success has emphasized that its early warning came from using massive computing power to sift out early indicators from millions of social media posts and other informal media.[123]

Thus, the rapid pace at which crisis mapping can assist in early notification of incidents affecting citizens and the global society is valuable externality, that is made even more so due its large scale accessibility as an open access initiative. How the crisis map picked up the potential pandemic included a computer algorithm that picked up social media posts that referenced health are workers who had blogged about the outbreak of Ebola. HealthMap is thought to have picked up these references before others.[124] This model comprises "sophisticated computer algorithms sift through millions of data points and divine hidden patterns indicating a previously unrecognized outbreak that was then used to alert unsuspecting health authorities and government officials."[125] This again highlights the potential for innovation and technological development that can be prompted by such initiatives because of the type of algorithm required for its efficiency. However, the contention that data algorithms can identify trends before other traditional channels has been disputed[126]. Irrespective of whether such algorithms are faster than traditional channels or not, they can undoubtedly assist to confirm or deny what is first reported, and provide logistical teams with vital geolocational data in the support of better preparedness when responding to crises. Like Ushahidi, HealthMap relies on the open tool, Google Maps as the primary source of geospatial data and is another example of an important cross-sector collaboration. However, despite its effectiveness, HealthMap is also susceptible to producing some negative externalities such as privacy breaches and security risks, as well as unintended consequences that can arise when factually incorrect data is published on a crisis map. These issues are address below under the section regarding issues for users and crisis map organisations.

Overall, whilst Ushahidi has maintained its prominence as a leading crisis mapping platform, the subsequent examples addressed above are indicative of the growing interest and support for crisis mapping due to it being of great benefit to society. This is likely to increase into the future to provide humanitarian and commercial benefits to those involved. Relevantly, crisis mapping identifies examples of beneficial big data use that calls for collaborations between sectors, such as the between big geo data actors and those dealing with crisis data. This can promote harmonisations of digital services across sectors.

The relationship between crisis data and geo data

Geo data is crucial in determining location and affected areas of crises and enabling logistical and other decisions central to the carrying out of emergency efforts. Crisis mapping initiatives rely on large volumes of geo-data that comprises satellite imagery, photographic imagery, terrain data, business listings, traffic and street view perspectives, and other related informational data. A large source of this data is Google Maps. Google Maps is built on top of its large geo-spatial data sets. Google Maps is integral to crisis mapping not only because it provides open access to a lot of its big geo data but because it holds such vast amounts of geo

[123] Leetaru, Kalev, "Why Big Data Missed the Early Warning Signs of Ebola", *Foreign Policy,* 26 September 2014.
http://www.foreignpolicy.com/articles/2014/09/26/why_big_data_missed_the_early_warning_signs_of_ebola
[124] Ibid.
[125] Ibid.
[126] This is also the case with algortihms used for Google Flu Trends that seeks to identify trends by analysing health related search queries: Lazer, David, Ryan Kennedy, Gary King and Alessandro Vespignani, "The Parable of Google Flu: Traps in Big Data Analysis", *Science Mag,* vol. 343, 14 March 2013, p. 1204. http://gking.harvard.edu/files/gking/files/0314policyforumff.pdf

data from around the globe, and its data stores are constantly being updated. The relationship between crisis mapping and geo is observed:

> Crisis mapping is being promoted as a new development at the nexus of geographic practices and humanitarianism that employs a variety of geospatial and new information communication technologies (ICT) to power effective early warnings or rapid responses to complex humanitarian emergencies.[127]

However, Google Maps is not the only open access source of geo-spatial data and Yahoo Maps, and NASA provide other examples. There are also other open resources, like Wikipedia and its semantic version, DBPedia, and Geonames, for location information that are is important building blocks for future applications. However, the easy accessibility of Google Maps, and its sheer volume of geo data[128], makes it particularly compatible with crisis mapping initiatives, a number of which develop in the humanitarian sector that does not have unlimited funds to purchase the amount of geo data required to run the crisis mapping initiatives. Google Maps data is accessible and re-usable by the public, businesses and governments. The majority of this data is capable of being updated by users, subject to the website's terms and conditions. Some of the data incorporated into Google Maps, has been obtained from open access initiatives in the public sector. This collaboration between government data and privately owned data is another reason

Thus, big data actors such as Google Inc. can provide integral tools that support open access initiatives. The relationship between crisis data actors and geo data actors is indicative of how successful cross-sector collaboration can be beneficial to society, whilst maintaining their commercial viability.

Benefits for the users and crisis mapping providers

A number of positive externalities that flow from crisis mapping have been addressed above in relation to examples of crisis maps, however, a more general discussion of societal benefits is provided here to highlight the potential presented by crisis mapping platforms. These opportunities are important to recognise as this is still an emerging area. However, this means that their effectiveness, and conversely the issues they implicate, have not yet been exhaustively examined in the relevant literature. However, a number of positive externalities are obvious in terms of the assistance crisis maps provide to humanitarian organisations, governments, communities and citizens affected by crisis situations. This includes the provision of up to date logistical data that can assist in emergency efforts.

There are also a number of positive externalities that relate to commercial opportunities, including technological innovation. The significance of the benefits related to open access to crisis data are further observed by UNISDR's Information Management Coordinator Craig Duncan:

> Open data, models and knowledge can help us understand our disaster risks. As of May this year, we have a global risk model for earthquakes and wind speed. We are

[127] Gray, Barry, "Crisis Mapping and Its Use in Disaster Management", afac.com, no date, p.1. http://www.afac.com.au/downloaddoc.aspx?q=5b224dca-0f92-4ecb-acb0-19c1eb39cc5e

[128] To give an indication of the volume of geo-data held by Google Inc., and made available through the Google Maps, which has expanded its street view coverage and mapping imagery to 58 countries, and it has completed the "largest single update of Street View imagery ever" with photos of over 370,000 miles (600,000 km) of roadways: Albanesius, Chloe, "Google Street View Expands to 58 Countries", *PCMag*, 6 April 2014. Further, Further, the Google Maps teams are currently publishing more imagery data every two weeks than the total data it had at its disposal in 2006: Madrigal, Alexis C. "How Google Builds its Maps--and What it Means for the Future of Everything," *The Atlantic*, 6 September 2012.

> now working hard with the GAR15 team to come up with a comprehensive global coverage for earthquakes, floods, and tsunamis that will include economic loss information.[129]

Crisis mapping platforms can prompt a demand for related innovations, especially in collaborative efforts such as wit the example of Ushahidi that uses Google Map's big geo data. For example, functions can be developed alongside Google Maps, such as technological support for embed maps to be incorporated in crisis mapping websites that enable users to obtain all relevant geo-locational information relating to a disaster from just one site. There is also potential for organisations to collect personal and behavioural data from those accessing crisis maps via these open access initiatives. This data can be turned into other sources of revenue. For example, models incorporating open access aspects are presumably sufficiently lucrative to warrant Google making their geo data publicly accessible via Google Maps, or at least, a large chunk of it. Aside from the commercial benefits, the crisis mapping model has had a tremendous positive impact in terms of crisis management, and a general increase in interest of satellite imagery. This in turn benefits those accessing during crisis situations because the increasing detail of the maps means that people can better identify precise locations of disaster zones etc. Thus, this type of open access provides a platform for others to take advantage of the informational output of Google Maps, as well as create further possibilities by accessing this geo data for crisis mapping.

Another positive externality of open access initiatives such as Ushahidi and other crisis mapping examples mentioned above, is the development of crowdsourcing. Crowd sourcing is a digital tool that allows the expression of personal demands and online collaboration to reach a common purpose, such as crisis mapping. This is mainly facilitated by people's access to open digital works.[130] Thus, the accessibility of crisis data through crisis mapping initiatives leads either directly or indirectly to collaboration with willing participants that may produce positive benefits for other users of the application.

Therefore, whilst crisis mapping initiatives present obvious societal benefits in terms of assisting humanitarian organisations and citizens to prepare responses to crisis situations, they are also an opportunity for growth and development amongst European big data actors. These opportunities can be better captured if the potential issues that arise in relation to crisis mapping are identified and minimised.

Issues for the users and crisis mapping organisations

Open access to crisis data may also produce negative impacts. These impacts range from privacy and data security concerns to implications when the geo-locational and other information is incorrect. It is easy to overlook such negative implications during times of crisis when getting access to data overwhelms any other priority.

Crisis mapping highlights the benefits of cross-sector collaboration (as addressed above), but it can also result issues that arise in relation to a service or product in one sector being transferred to the subsequent service. This is seen with the collaboration between sources of big geo data and crisis maps. For example, users of Google Maps are concerned about their

[129] cited in Fung, Vincent, "Open Data Makes Disaster Risks Viable", *UNISDR,* 15 October 2013. http://www.unisdr.org/archive/35126

[130] Ortiz Ehmann, Cristóbal, Crowdsourcing - A Pandemonium for Disruptive Innovation (May 16, 2012). Available at SSRN: http://ssrn.com/abstract=2423471

privacy[131] in relation to Google's use of invasive technologies to obtain the data used in its Google Maps service, which is one of the major sources of geo data of crisis mapping. Thus, concerns over Google Maps flow through to crisis mapping examples. The same can be said for concerns regarding the collection of behavioural data whilst users access mapping services. Another potential negative externality can arise in relation to collaboration when the default reliance on one dominant geo data provider, such as Google Inc., has the effect of reducing competition. However, this also highlights the need for the development of similar and competitive models in the market, especially in light of the humanitarian focus of this open access initiative. The European Data Protection Supervisor draws attention to this market situation: The digital economy is marked by strong, dynamic growth, a high turnover of new services, market concentration involving a few overwhelmingly dominant players, and an ever greater imbalance between big companies on the one side, and SMEs and individual users on the other sides."[132] This suggests that it is important to identify the potential for competitive businesses in the market.

Further, aid agencies' use of satellite images and detailed maps may be considered to have security implications. Yet even though treaties, such as the International Charter on Space and Major Disasters, are supposed to provide a framework for the rapid release of data in an emergency, too often this process is delayed.[133] In addition, other security implications include: what information should be shown, is it feeding malicious actors with intelligence, does sharing data endanger people already at risk, what happens to vulnerable citizens if crisis mapping is wrong, what responsibility do humanitarian organisations have for incorrect information or the realisations of adverse repercussions for citizens if the open access data puts them in harms way or jeopardises rescue efforts. Similarly, open access crisis maps can result in the manipulation of data. This is referred to as "red team dynamics" when research subjects or web searchers attempt to manipulate the data generating process to meet their own goals, such as economic and political gain.[134]

Finally, as crisis mapping is in its infancy, there are no specific guidelines in place, and nor has it been around along for customary standards to evolve. This has been identified as being problematic:

> More than simply lacking an accepted doctrine or code of ethics, a body has not been designated to convene or coordinate the community of crisis mappers with the intent of developing agreed standards or consensus. If the crisis mapping community does not develop shared ethical standards or determined responsible practices, they will lose the trust of the populations that they seek to serve and the policymakers that they seek to influence.[135]

---

[131] Manyika, James, Michael Chui, Peter Groves, Diana Farrell, Steve van Kuiken and Elizabeth Almasi Doshi, "Open Data: Unlocking Innovation and Performance with Liquid Information", *McKinsey & Company,* October 2013, p. 53.

[132] EDPS, Preliminary Opinion of the European Data Protection Supervisor: on Privacy and Competitiveness in the Age of Big Data: The Interplay between Data Protection, Competition Law and Consumer Protection in the Digital Economy", March 2014, p.8. https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2014/14-03-26_competitition_law_big_data_EN.pdf

[133] OCHA, "Humanitarianism in the Network Age", *OCHA Policy and Studies Series, 2012, p. 37. https://docs.unocha.org/sites/dms/Documents/WEB%20Humanitarianism%20in%20the%20Network%20Age%20vF%20single.pdf*

[134] This has occurred in connection with Twitter and Facebook, see Lazer, David, Ryan Kennedy, Gary King and Alessandro Vespignani, "The Parable of Google Flu: Traps in Big Data Analysis", *Science Mag,* vol. 343, 14 March 2013, p. 1204. http://gking.harvard.edu/files/gking/files/0314policyforumff.pdf

[135] Gray, op. cit., no date,p.1.

Whilst open access initiatives such as crisis mapping present a number of issues for organisations and consumers, better understanding of the impact of these externalities can enable actors to minimise their negative effect. This is especially important in light of the array of societal benefits produced by open access initiatives that involve big crisis data.

Section summary: crisis mapping

Crisis mapping is an emerging open access initiative involving crisis data and reflects the relationship across sectors between crisis data and big geo data. The Haiti earthquake in 2010 drew more widespread attention to the use of crisis mapping as a tool, although contemporary crisis mapping attracted notoriety for its contribution to humanitarian efforts following the 2008 Kenyan election crisis. Open access initiatives for crisis data rely largely on geo data provided through applications such as Google Maps. The aforementioned open access initiatives rely heavily on big geo data, which is provided through applications such as Google maps. Google maps itself provides a good example of a big data which provides open access to geospatial information.

Social media: a disseminator of crisis data

Social media plays a prominent role during times of crisis, often publishing information and images long before emergency teams and or journalists and other aid agencies arrive at the scene. Users obtain information, education, news, and other data from electronic and print media sites such as Facebook and Twitter. In this section, we consider more generally the positive and negative externalities of social media as an open access initiative that involves crisis data, and we also consider externalities specifically in relation to the role Twitter Inc. plays during crisis situations. The potential role that social media plays during a time of crisis is great because of the high rates of usage. According to CNN, in 2010 75% of people got their news forwarded through e-mail or social media posts, while 37% of people shared a news item via Facebook or Twitter.[136] Additionally, the social media platform also offers anyone the means to publish information and enables anyone with access to view it, expand upon it, re-use it or collaborate with it. "the most important driver for social media is the wide availability of free and easy to use service applications".[137] Thus, Social media companies are an integral disseminator of information during crises. They also provide a means by which individuals can communicate directly to assist one another such as by providing up to date logistical information. For example, a traveller stranded near an affected area who us unfamiliar with that area, can tweet or post on Facebook requesting information in relation to possible transport routes or other information that may be of assistance to them during that time.

Social media is different to traditional forms of media in that the information is accessible to anyone with access to the requisite technology, and the information on social media is offered without entry barriers such as subscription fees. It is by and large open access. "The Internet as a whole and social media in particular has a strong tradition of being free for the end-users."[138] Thus, social media is an invaluable tool during times of crisis, as they become epicentres for large amounts of crisis data. However, this open access model is epitomised by the user exchanging their personal data for access to the social media platform. Therefore,

[136] Gross, Doug, "Survey: more Americans Get News From Internet than Newspapers or Radio", *CNN online,* 1 March 2010. http://edition.cnn.com/2010/TECH/03/01/social.network.news/index.html

[137] Ahlqvist, Toni, Asta Back, Minna Halonen and Sirkka Heinonen, *Social Media Roadmaps: Exploring the Future triggered by Social Media,* VTT, Technical Research Centre, Finland, 2008, p.5. http://www.vtt.fi/inf/pdf/tiedotteet/2008/T2454.pdf

[138] Ibid., p.17.

whilst social medial plays an important social role in alerting and informing citizens of crises, it provides that for free in the sense that the user does not pay a monetary subscription free. The user pays for the privilege in other ways, which enables the social media provider to make money from the user-generated data. This represents a benefit for social media companies, but it can also raise important legal and ethical issues such as privacy for users of these services from whom personal and behavioural data is collected.

Therefore, open access initiatives in the form of social media models support the presumption that business models based on the provision of access to big crisis (or the extent of access to data in which a company has a proprietary interest) can be profitable and competitive, as well as producing a number of societal benefits. However, they also present a number of issues when used in times of crisis.

Benefits for users and companies

The use of social media during crisis situations produces a number of benefits to society, as well as commercial opportunities to the companies implementing them. Social media platforms play a vital role in the dissemination of information during crisis situations. For example, they are an up to date source of information about crises at the time they are occurring, and they can acts as a platform for assistance by enabling communication between individuals and communities. The potential impact of open access initiatives such as social media is immeasurable in terms of delivering vital information to assist humanitarian and emergency responses to crisis situations, as well as inform citizens and communities around the world with up date information. Social media can also support individuals assisting one another during these times by, for example exchanging up to date logistical information.[139] In addition to the direct benefit to the public as an information source, and a tool for assistance and support, emergency response agencies have identified the power of the information being generated and established channels to support and build their communication and information collection via this medium.

Further, the social media business model is also a potential source of profit for social media companies, as well as spurring technological innovation in the form of measures that address the high demand for these networks during times of crisis. Commercial opportunities presented by social media usage during times of crisis include when "mobile social media makes use of the location- and time-sensitivity aspects of it in order to engage into marketing research, communication, sales promotions/discounts, and relationship development/loyalty programs".[140] The most common source of revenue for these models is advertising. Targeted advertising can be undertaken because users often disclose large amounts of data about themselves. The ability of social media companies to captialise on that data is improved during crisis situations when, particularly during the time of the crisis and the period of time shortly thereafter, the amount of data rapidly increases.

Positive externalities that flow from the use of social media during crisis situations are best examined in relation to the specific example of Twitter, discussed below.

[139] Watson Hayley and Rachel Finn, "Privacy and Ethical Implications of the use of social Media During a Volcanic Eruption: Some Initial Thoughts", Proceedings of the 10th International Iscram conference, May 2013.

[140] Kaplan, Andreas M, "If You Love Something, Let it Go Mobile: Mobile marketing and Mobile Social Media 4x4", *Business Horizons* **55** (2): 129–139, March-April 2012. http://www.sciencedirect.com/science/article/pii/S0007681311001558

Issues for users and companies

A major issue in relation to social media networks used during times of crisis is that they can threaten privacy and security of individuals and organisations, as well as hampering emergency efforts by publishing factually incorrect data.

Watson and Finn identify a number of social and ethical implications of the use of social media during the eruptions of the Eyjafjallajokull volcano in 2010. That incident led to the grounding of aircraft across Europe, and social media use at that time illuminated how privacy can be put at risk. The EDPS suggests "'Free' online services are 'paid for' using personal data which have been valued in total at over EUR 300 billion and have been forecast to treble by 2020."[141] The risks posed by this collection and analysis of personal data are rarely accounted for and not adequately addressed. The personal data provided by consumers may accessed, used and re-used by a number of stakeholders including a multiplicity of individuals, businesses, public institutions and non-profit organisations, including data brokers and cloud computing service providers.[142] Further, as mentioned above, the creation of consumer profiles that may lead to profiling and discriminatory practices such as different priced offerings geared towards different types of users also threatens consumer welfare and is considered a current issue for discussion in light of EU consumer protection law.[143] Another aspect of consumer protection is raised by the imbalance between provider and consumer with respect to online services. This refers to a situation in which the consumer not genuinely being presented with choice and they have limited room, if any, to negotiate the terms and conditions of the service contract which is characterized by the user giving access to their personal data in exchange for a popular and or dominant service to which there are few alternatives. During times of crisis in the digital age, there are few other dominant sources of up to date services that provide the benefits of access that social media networks do.

The role of social media during times of crisis has become vital in the dissemination of important and relevant information pertaining to that crisis. Its presence is bolstered by the fact that access to crisis data is free and up to date. Twitter provides useful example of the use of the social media during crisis situations, including the societal benefits it produces, as well as highlighting when it can raise issues of concern for users of the service.

Twitter Inc.

Twitter provides a clear illustration of the interplay between big crisis data and open access. The Twitter business model represents an "open ecosystem"[144] and is based on the open sharing of information. Twitter's mission is: "to give everyone the power to create and share ideas and information instantly, without barriers." This mission encompasses the ethos of the open access movement. It may also allay fears that open access models are not profitable.

Former CEO and co-founder of Twitter, Inc., Ev Williams, describes the evolution of Twitter:

[141] EDPS, Preliminary Opinion of the European Data Protection Supervisor: on Privacy and Competitiveness in the Age of Big Data: The Interplay between Data Protection, Competition Law and Consumer Protection in the Digital Economy", March 2014, p.8. https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2014/14-03-26_competitition_law_big_data_EN.pdf

[142] Ibid., p.10.

[143] see EDPS, op. cit., 2014, p.23.

[144] Twitter, Inc., "Your Rights", *Terms of Service*, 25 June 2012. https://twitter.com/tos

"Twitter was really more of an information network than it is a social network".[145] It has also been suggested that Twitter is "a combination of broadcasting service and social network, classes as a "social broadcasting technology".[146] All Tweets are instantly indexed and Twitter co-founder, Biz Stone, describes Twitter as "[…] a discovery engine for finding out what is happening right now.[147] This premise of this model makes it specifically useful in times of crisis as it acts as a large-scale source of information relating to the crisis.

Tweets are public by default (although users can restrict their privacy settings so that approved followers only view tweets), which makes it a powerful tool in real time dissemination of news and related information about crisis situations. However, these followers are still able to re-tweet the content to other users. Whilst tweets are limited to 140 characters in length, there is an expectation that Tweeters will post links to information relevant to their tweets.[148] The fact that tweet data also combines geo data referencing the location of the Tweet can be useful during times of crisis. Up to date information is presumably likely accurate when it is tweeted from within a crisis zone, although regard must be had for the state of the person tweeting. For example, if it is a news organisation that is located within the affected area, this may have more currency than a distressed civilian or another user seeking to spin the information for political gain for example

The Twitter Terms of Service provide: "By submitting, posting or displaying Content on or through the Services, you grant us a worldwide, non-exclusive, royalty-free license (with the right to sublicense) to use, copy, reproduce, process, adapt, modify, publish, transmit, display and distribute such Content in any and all media or distribution methods (now known or later developed)".[149] Further, the Terms of Service summarise rights with respect to content: This license is you authorizing us to make your tweets available to the rest of the world and to let others do the same."[150] The data posted to, and shared on, Twitter undoubtedly makes an impact. According to its internal statistics, there are 255 million active monthly Twitter users, and 500 million Tweets are sent per day, in 35 plus different languages.[151]

Benefits for users and consumers

Although Twitter operates in the private sphere, and does so on the basis that its open access to content generates revenue, largely generated through paid advertising on its platform, it forms part of the global move towards open access as a source of information and material that may be developed by users, re-used and shared with great benefit. The sharing of this information can greatly assist emergency services, humanitarian organisations, journalists, civilians, governments and communities affected by crises or implore other communities to assist in emergency efforts. An overriding positive externality of Twitter in the provision of access to crisis data is that it serves a as public objective by informing the pubic of emergency

[145] Lapowsky, Issie, "Ev Williams on Twitter's Early years", *Inc.* 4 October 2013. http://www.inc.com/issie-lapowsky/ev-williams-twitter-early-years.html?cid=em01011week40day04b

[146] Shi, Zha, Rui Huaxia and Andrew B Whinston, "Content Sharing in a Social Broadcasting Environment: Evidence from Twitter". *MIS Quarterly*, 4 august 2013. http://misq.org/content-sharing-in-a-social-broadcasting-environment-evidence-from-twitter.html?SID=86knf65su27v98u9npsa31heh3

[147] Stone, Biz, "Twitter Search for Everyone", *Twitter Blog*, April 30, 2009.

[148] Holton, Avery E., Kang Baek, Mark Coddington, and Yaschur, Carolyn, "Seeking and Sharing: Motivations for Linking on Twitter". *Communication Research Reports* **31** (1), 2014, pp. 33–40. http://www.tandfonline.com/doi/abs/10.1080/08824096.2013.843165?queryID=%24%7BresultBean.queryID%7D#.U5fJU5SSz_4

[149] Twitter, Inc., "Your Rights", *Terms of Service*, 25 June 2012. https://twitter.com/tos

[150] Ibid.

[151] Twitter, Inc., "About", 2014. https://about.twitter.com/company

situations and breaking news, and disasters. Although it has also been used in relation to civil unrest as was the case during the Arab Spring. Twitter is also being used as an informal educational tool. However, twitter accounts lack rigorous fact checking. It doesn't have to be real or true to post about it.

Social media also provides users and organisations with other benefits during emergency situations. In fact, it is accepted that "a common thread running through all definitions of social media is a blending of technology and social interaction for the co-creation of value".[152] For example, social media acts as a point of information and assistance by requesting real time information about the crisis or consequential issues such as the need to plan alternative travel or make plans for someone caught in the area of crisis.[153] In the example of the Eyjafjallajokull volcano, Watson and Finn underline the usefulness of social media: "numerous instances of those involved in the aviation industry, including authorities and airlines, using social media to communicate with passengers".[154]

The context in which crisis data emerges, and the rapid increase in the volume of that data during a concentrated period of time means it is best supported by open software that facilitates the collection, storage and availability of large amounts of data in one place. This further opens up opportunities for analysis, which in turn leads to the discovery of trends or assists in the development of solutions and assistance during crisis events. Open access initiatives that incorporate crisis data reflect contemporary big data models that can create a number of benefits for citizens and society, as well as opportunities for organisations and consumers. The very reason crisis data initiatives are useful and produce benefits is because they are open access.

Issues and problems with the openness and accessibility of social media channeled crisis data

Nevertheless, social media sites such as Twitter can hamper emergency services efforts as well as assist them. The latter can occur when users update publish false and inaccurate information. This can also risk personal privacy of already vulnerable people in affected areas, as well as threaten their security. For example, looting has become wide spread during times of crisis and access to information can assist criminals in identifying vulnerable areas. Further, in relation to crisis data, tweets also have the potential to misinform, or confuse or scare the public when the 140-character limit means that short sharp bursts of information can mean that they can fuel perceptions that of events that are more alarming than the event itself. Further, as tweets are not modified or checked by a moderator, false information can be tweeted, which has the potential to worsen situations.

[152] Morgan, Nigel, Graham Jones and Ant Hodges, "Social Media", *The Complete guide to Social Media From the Social Media Guys,* PDF version, November 2010 *p.1. http://rucreativebloggingfa13.files.wordpress.com/2013/09/completeguidetosocialmedia.pdf*

[153] Watson, H and Rachel Finn, "Privacy and Ethical Implications of the Use of Social Media During a Volcanic Eruption: Some Initial Thoughts", *Proceedings of the 10th International ISCRAM Conference*, Baden-Baden, Germany, May 2013, 417.

[154] Watson, Hayley and Rachel Finn, "Privacy and Ethical Implications of the Use of Social Media During a Volcanic Eruption: Some Initial Thoughts", *Proceedings of the 10th International ISCRAM Conference*, Baden-Baden, Germany, May 2013, 417.

Finally, while consumers stand to gain from open crisis data, privacy concerns represent a negative impact these models have on consumers.[155] This is particularly so in disaster situations when graphic and distressing images become common place on social media without the consent of those captured in the images. The open accessibility of social media like twitter and the vulnerability of tweets (even where private) being re-tweeted can easily be seen by friends or families or associate of those captured in related footage. There is no moral compass or moderator for these sensational tweets. Companies will also need to have strategies about how, when, and under what circumstances they open their data, taking into account the potential impact the release of their data could have.[156] Privacy is also implicated directly in relation to the commercial practices of social media companies. The collection of personal information raises ethical and legal issues with respect to the tracking of digital behaviour, as well as being potentially privacy invasive, and/ or manipulative practices.[157] The European Data Protection Supervisor ("EDPS") observes, "Online services are driving the huge growth in the digital economy. Many of those services are marketed as 'free' but in effect require payment in the form of personal information from customers."[158] The personal information collected from and about users may also result in discriminatory practices where differing price points and services are offered to different sectors of society following the creation of consumer online profiles.

Therefore, social media networks such as Twitter can produce a number of negative externalities when used in times of crisis. Although Twitter has the ability to disseminate vital information relating to disasters at a rapid rate, this open dissemination implicates a number of issues, not the least of all is privacy.

3.3.3 Summary

Crisis data is an obvious candidate for open access initiatives that produce an array of societal benefits. However, open access initiatives using crisis data can also produce some negative externalities that are often overlooked during times of crisis when obtaining as much information about a situation as possible becomes the primary focus. Despite the obvious humanitarian benefits of open access initiatives involving crisis data, and the commercial opportunities that flow from providing open access to crisis data, social media involving crisis data presents challenges for individuals and organisations as well as governments.

Whilst there is potential value in these open access initiatives, little is yet known about how that is captured in relation to crisis mapping initiatives. On the other hand, the social media model is better understood as providing open access to information, in this instance crisis data, in return for the vast amounts of personal and behavioural data collected from users when they access social media sites, which is later marketed to advertisers for example.

Google maps collaborate with Ushahidi to provide up to date and crowd sourced information about crisis situations, as they did during the Kenyan post election crisis of 2008. The collaboration between platforms such as Ushahidi and other crisis maps and private actors

[155] Manyika, James, Michael Chui, Peter Groves, Diana Farrell, Steve van Kuiken and Elizabeth Almasi Doshi, "Open Data: Unlocking Innovation and Performance with Liquid Information", *McKinsey & Company,* October 2013, p. 21.
[156] Ibid., p. 23.
[157] see Finn, Rachel L., and Kush Wadhwa, "The Ethics of Smart Advertising", *Emerald Insight,* Vol. 16, No. 3, 2014.
[158] EDPS, op. cit., 2014, p.3.

such as Google Maps highlights the room for innovative collaboration in the area of big crisis data.

Further, providing open access to crisis data, such as information and alerts about dangerous political, social or natural disasters via social media sites such as Twitter provides a useful public service. However, these models that facilitate knowledge and awareness of crisis situations are not without risks associated with privacy. However, it is difficult to identify relevant European open access examples that deal largely with crisis data, which represents a discrepancy in the market and also an opening for the fostering of new platforms, and the opportunities that flow for society and organisations implementing them.

Overall, open access initiatives involving crisis data offers the companies providing these services the ability to generate profit from the relationship between big commercial data and open access, as well as providing a public good.

3.4 ENERGY DATA

3.4.1 The relevance of open data to the energy sector

Energy data comprise a wide range of government and non-government data sources such as renewable energy installations, oil & gas reservoirs, or energy customer data to name a few. Initiatives for opening energy data aim to stimulate entrepreneurship and create new products and services that may serve to reduce costs, protect the environment, and ensure a safe and reliable energy supply. For example, the Energy Data Initiative from the US government aims to "liberate government data and voluntarily contributed non-government data as fuel to spur entrepreneurship, create value, and create jobs in the transition to a clean energy economy."[159] There are similar initiatives in other countries, such as the EPIM E&P Information Management Association for the oil & gas industry in Norway, whose main objective "is to facilitate the best possible flow of information between our users. To meet this need we deliver innovative IT solutions, based on open standards and semantic technology."[160] They have a series of services for gathering and sharing information, which many are based on the general model of ISO 15926 and the Norwegian Petroleum Directorate FactPages dataset. Here we present this case as a well-known and heavily used open dataset that can serve to illustrate how to better integrate and manage information in energy companies and between cooperating companies.

3.4.2 The case of the Norwegian Petroleum Directorate FactPages

The Norwegian Petroleum Directorate (NPD) is a governmental specialist directorate and administrative body which reports to the Ministry of Petroleum and Energy in Norway. NPD's main objective is to "contribute to creating the greatest possible values for society from the oil and gas activities by means of prudent resource management based on safety, emergency preparedness and safeguarding of the external environment."[161] One of the datasets that the NPD manages is the NPD FactPages,[162] or FactPages for short. The FactPages contain data about petroleum activities on the Norwegian continental shelf (NCS)

[159] https://www.data.gov/energy/energy-data-initiative
[160] http://www.epim.no/epim/main/about
[161] http://www.npd.no/en/About-us/
[162] http://factpages.npd.no/factpages/

ranging from operating companies and production licences, fields and discoveries, facilities and pipelines, to wellbores and stratigraphic data; some data dating back to the start of oil production on the NCS in the early 1970s. The data in the FactPages is collected from companies that operate on the NCS, and this information forms the basis for the authorities' planning of future activity and their judgement of existing activity. Additionally, an important purpose of the FactPages is to secure efficient sharing of information between the companies, and to provide sufficient information to the public. To this end, the FactPages are published online under the following open licence:

> The content on the Norwegian Petroleum Directorate's website can be copied and used free of charge as long as all materials are marked with their date, reference and link to their source. This information must appear prominently. The content on the Norwegian Petroleum Directorate's webpages may be used in accordance with Norwegian Licence for Open Government Data (NLOD) http://data.norge.no/nlod/en/1.0.[163]

The dataset may be examined and searched in with an ordinary web browser, and large parts of it are available for download in bulk, in CSV, Excel and XML format. Following the popular Tim Berners-Lee's Linked Open Data star rating system[164,165], we may award the FactPages three stars: the data are made available on the web under an open licence, and the data are available in structured and non-proprietary formats. The benefits of having a three star dataset are already substantial; for the consumer of the dataset the first star is very important as it warrants any use of the dataset at all. For the second and third star, having the dataset on machine-readable, structured and non-proprietary format means the user is not confined to a specific tool-set when processing and manipulating the dataset. For the data provider, serving three star data is simple, and amounts setting up a simple regular export from the current database system to a structured non-proprietary format, and releasing it under an open licence. The NPD use Microsoft's SQL Server Reporting Services that allows one to create reports from a relational database, which may be generated in a range of output formats including HTML, Excel, PDF, CSV and XML.

However, further benefits can be achieved by publishing the FactPages as Linked Open Data and, thus, obtaining a five star rating. The process of converting the FactPages into semantic web data is described here[166]. Briefly, every resource, e.g. a wellbore, is assigned a URI that serves as a global identifier and can be looked up to obtain a description of the resource. Resources are represented in RDF[167], a simple, uniform and universal data model that simplifies data exchange and consumption. Using the RDF language is thus possible to describe the information of a domain such as the petroleum activities on the Norwegian continental shelf. For example, we can say in RDF that Ekofisk is an oil field operated by the company ConocoPhillips. Due to the RDF format, adding new data to a dataset or merging it with other RDF datasets is as simple as taking the union of the sources, so there is no need to harmonise the structure of database tables. It is also worth noting that this structure makes RDF "schemaless", and to some extent self-describing; the dataset may carry its own schema by containing facts that define the schema vocabulary used by the dataset. The resulting

[163] http://www.npd.no/en/About-us/Information-services/Use-of-content/
[164] Tim Berners-Lee. Linked Data. 2010. URL: http://www.w3.org/DesignIssues/LinkedData.html
[165] Tom Heath and Christian Bizer. Linked Data: Evolving the Web into a Global Data Space.
[166] Martin G. Skjæveland, Espen H. Lian, and Ian Horrocks. "Publishing the Norwegian Petroleum Directorate's FactPages as Semantic Web Data". Proc. of the 12th International Semantic Web Conference (ISWC 2013).
[167] Graham Klyne and Jeremy J. Carroll. Resource Description Framework (RDF): Concepts and Abstract Syntax. W3C Recommendation. W3C, 2004. URL: http://www.w3.org/TR/rdf-concepts/

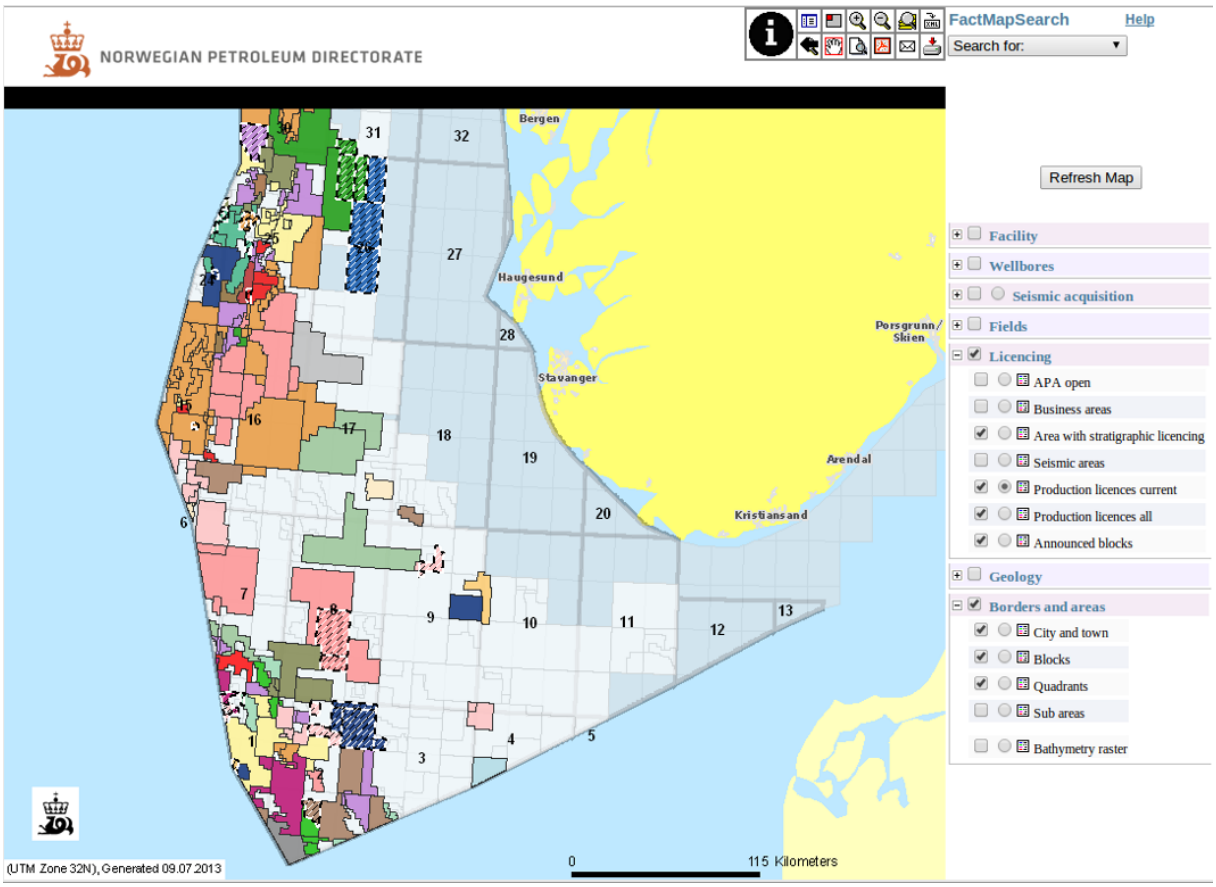
dataset is exposed for querying through a SPARQL[168] endpoint and its data is made available for linking and display with a linked open data frontend. The endpoint is an interface that adheres to the SPARQL protocol and allows users to issue SPARQL queries over the RDF data in the triple store. So instead of setting up a set of web services, where each service corresponds to a query over a relational database, a triple store containing RDF data may safely expose its contents to the Internet through a SPARQL endpoint allowing arbitrary SPARQL queries over its dataset. In the Linked Open NPD FactPages, the URI for querying the SPARQL endpoint is http://sws.ifi.uio.no/sparql/npd-v2?query= followed by a URL encoded SPARQL query. Hence, setting up a report for export in structured format, e.g., similar to the FactPages' bulk data, may be as simple as specifying a SPARQL query. All the information about the Linked Open NPD FactPages can be gathered at the website http://sws.ifi.uio.no/project/npd-v2/.

For the consumer of data the benefits of a five star dataset compared to only three stars are profound. Each data item has a globally unique identifier which can be referenced, bookmarked, and information about the data item can be retrieved both in human and computer intelligible formats by simply using the identifier as a look-up address. Data is no longer kept in blocks, such as spreadsheet files, so now only selected parts of datasets may be easily used. Since data are linked and linked to others' data, more data can be collected by plainly crawling the dataset. Additionally, given that the dataset also includes its schema definition, the same methods for exploring the dataset can be used for learning about the semantics of the data. For the data provider, serving five star data requires more than that of a three star dataset. URI schema identifiers need to be designed, both with the intention that the URIs should stay as stable as possible and that the URI must be de-referenceable. Moreover, a thorough restructuring of the dataset may be necessary in order to adapt to the graph format of RDF, and finding other relevant datasets of high quality can be a difficult task. On the technical side, if the provider should decide to keep the old information system, which might very well be reasonable choice, then an apparatus for converting the data to RDF needs to be installed, maintained and updated. However, there are tools that are built to solve the problem of exposing relational databases as RDF graphs, e.g. the D2RQ platform.[169]

A further benefit for users is that arbitrary SPARQL queries can be posed to the Linked Open NPD FactPages. Indeed, a dataset query interface open to users is an extremely valuable and powerful asset as it allows them to extract, aggregate and join data in ways that are practically impossible with the official FactPages; take the screenshots in Figure 1 as examples. The top image shows the NPD FactMap, which is published alongside the FactPages and allows users to display geographical information from the FactPages on a map. All selections are predefined — the current selection shows all production licences and blocks — so customising selections like "show (only) licence areas where Statoil is partner or operator" is impossible. The bottom image shows a general-purpose web application developed for the Linked Open NPD FactPages. It can display the result of SPARQL queries containing geographical information on a map, using free and openly available software. The given display shows four layers, each layer corresponds to a SPARQL query result: all licence areas, all active licence areas, all licence areas where Statoil is partner, and all licence areas where Statoil is operator. It is easy to specify the SPARQL queries necessary to have the same query functionality as in the NPD FactMap application. This shows that by making query endpoints available it is possible to present both the existing predefined queries and

[168] The W3C SPARQL Working Group. SPARQL 1.1 Overview. W3C Recommendation. W3C, 2013. URL: http://www.w3.org/TR/sparql11-overview/
[169] http://d2rq.org/

reports, and additionally have available the more advanced feature of executing custom-built queries. Note also that the SPARQL map visualisation application does not need to run on the same site as the dataset is hosted, illustrating that data publishers of high quality open data and their consumers can exploit available software to present and process data. Data publishers should hence concentrate on providing high quality data, and leverage the growing set of available tools developed for linked open data instead of developing potentially costly special-purpose applications themselves.



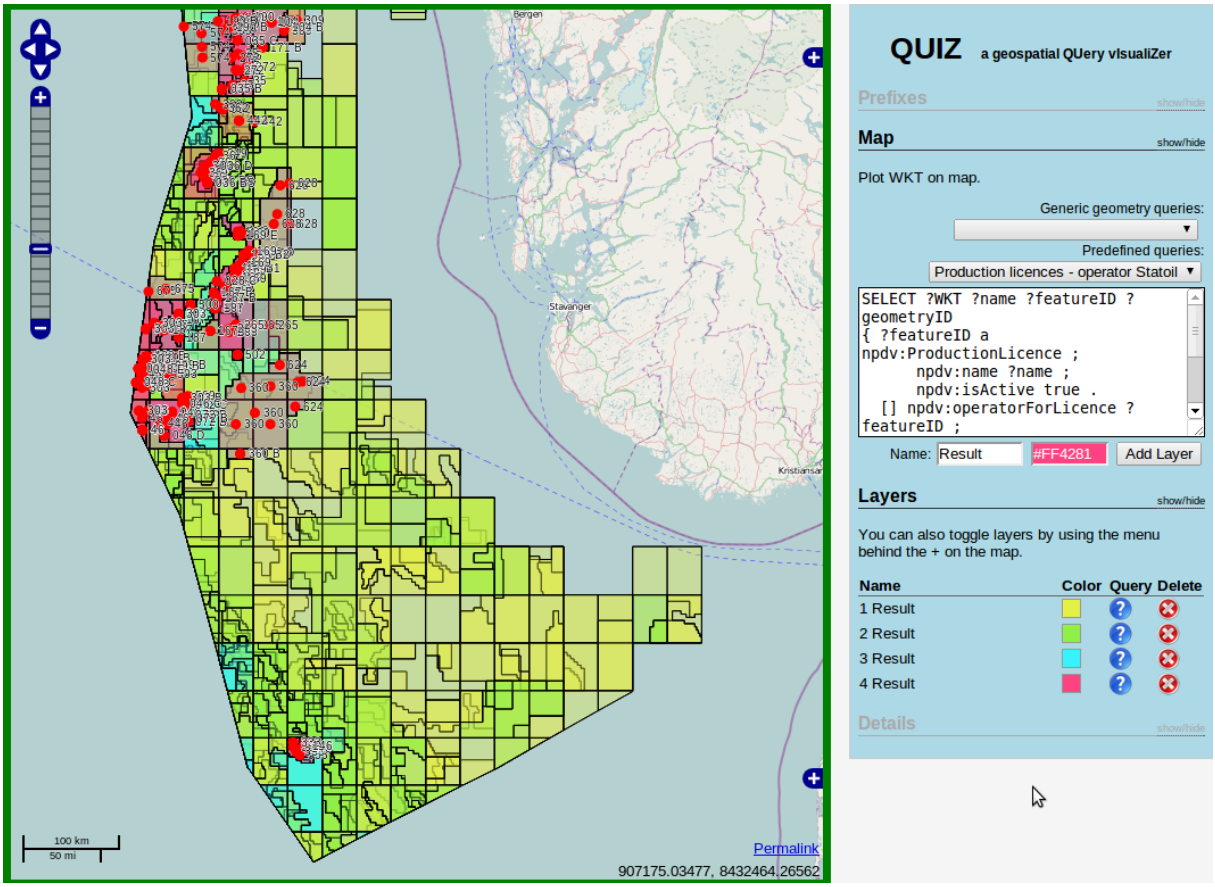


Figure 1: NPD FactMap vs. SPARQL map visualizer.

Data integration can be greatly simplified by exploiting the RDF model as illustrated with the following example: the FactPages contain data about the ages of the stratigraphy that a wellbore passes through, e.g., Jurassic era. When drilling a wellbore the operator often takes samples of the matter that is drilled in; this sample is called a core. Parts of these cores are physically stored by the NPD so that other companies and researchers can make reservoir studies by examining the cores. Suppose we want to find all wellbores that have the same true depth (TD) age as the ages where the discovery wellbores of Ekofisk have found petroleum, but we are only interested in those wellbores which are drilled earlier than 1990 and for which physical cores and biostratigraphic data are available. This question is impossible to answer using the FactPages alone. The first problem is that the FactPages does not contain any information about biostratigraphy. As a remedy, it is possible to convert the Norlex database[170], which contains this information, into RDF and integrated it with the Linked Open NPD FactPages. Also missing from the FactPages is a proper representation of the global time scale giving the relations between the different eras, e.g. the Jurassic period is split into different epochs, Upper, Middle and Lower Jurassic, so a string match on the names of the eras is not sufficient, since the eras in the FactPages are given at different levels of granularity. To solve this problem we can simply import the Geologic timescale vocabulary[171], which is maintained by members of the geologic community, and then create the mappings to the names of geologic eras in the FactPages. Once accomplished these two

[170] Felix M. Gradstein et al. “Norwegian Offshore Stratigraphic Lexicon (NORLEX)”. In: Newsletters on Stratigraphy 44.1 (Oct. 2010), pp. 73–86.
[171] Simon Cox. Geologic timescale (2012). RDF vocabulary. URL: http://resource.geosciml.org/classifierscheme/ics/2012/ischart

integration steps, it is possible to create a SPARQL query that answers our previous question. More information about the FactPages case can be found here[172].

3.4.3 The benefits of open access to energy data

The companies operating in the Norwegian continental shelf continuously access the FactPages dataset. This initiative thus serves to illustrate how the release of open data can help to improve the exploration and operation activities in the oil domain. While NPD publishes the official FactPages as three stars data, the case we have presented demonstrates that further benefits can be obtained by providing semantically annotated and query enabled linked open data. On the one hand, access to data is much more flexible and can be exploited in unanticipated ways, e.g. allowing petrophysicists to pose *ad hoc* queries over oil data. On the other hand, linked open data facilitates the integration of datasets that were not originally purposed to be merged – see the previous example of the integration of FactPages with biostratigraphic and geologic eras data.

Additionally, open data can be even more useful if provided in a "raw" condition. In the FactPages case, NPD should (also) publish the original data, as it exists in their internal databases, allowing users to a greater degree create their own views on the data. The raw data should be accompanied by available metadata, such as database schema definitions, and models, e.g. ER or UML diagrams. This will make the users able to better understand the data and reduces the possibility of incorrect usage of the dataset. Serving the raw data removes the problem of keeping the export definitions updated, like NPD's attribute explanations. By exposing the raw data, publishers do not need to invest resources on developing ad hoc tools for manipulating and visualizing data.

3.4.4 Negative impacts associated with open access to energy data

In the energy domain there are many useful datasets that do not include personal data, such as the FactPages, so privacy concerns are not normally an issue. Nevertheless, publishing open data entail non-negligible time and monetary resources, especially when exposing five stars data. These costs can be substantially reduced if adhering to a three stars open data scheme, as in the case of NPD for the official FactPages. Publishing open data also requires licensing to clarify what users can do with the data. In this regard, data providers should strive to use an existing licence developed by standardisation organisations or the government. The Norwegian Agency for Public Management and eGovernment (Difi) has developed a licence especially for open government data,[173] and opendatacommons.org provide three different licences for data/databases.

3.4.5 Summary

Energy is a broad field that encompasses many disciplines. Releasing open energy data produces a number of opportunities for innovation, as illustrated with the NPD FactPages case. Note that releasing both the raw data and schema of a database to the public can be a terrifying thought for database administers, and requires properly implemented change management processes since internal changes are also exposed to the public. However, these

[172] Martin G. Skjæveland and Espen H. Lian. "Benefits of Publishing the Norwegian Petroleum Directorate's FactPages as Linked Open Data". In: Norsk Informatikkonferanse 2013.
[173] http://data.norge.no/nlod/en/1.0

are procedures that should be in place for any important dataset anyway, and we believe the benefits of having truly open data outweighs the costs of producing and maintaining them.

3.5 ENVIRONMENTAL DATA

3.5.1 Relevance and importance of open access to big data in the environmental sector

The Digital Agenda for Europe (DAE)[174] recognizes that the big data revolution brings about novel ways of understanding and addressing environmental challenges. A better use of globally available national and local datasets helps scientists in their research and enables policy-makers to make informed and evidence-based decisions to fight against climate change and reduce costs. At a local/regional level, positive and negative implications of open access to big environmental data overlap with the context of utilities/smart cities, as far as the management of the environment is concerned. We elaborate on the utilities/smart cities sector in chapter 2.2.4; here we focus on Earth Sciences data, particularly from Earth Observation.

In fact, especially in the sector of Earth Observation, we are assisting to an explosion of data, due to many factors including: new satellite constellations, increased sensor technology, social media, crowdsourcing, and the need for multidisciplinary and collaborative research. In this area, there are many expectations and concerns about big data, not only by scientists, but also by vendors, who are attempting to use big environmental data for their commercial purposes. It is necessary to understand whether big data is a radical shift or an incremental change for the existing digital infrastructures for environmental and geospatial data.

3.5.2 Examples of policies and initiatives relating to environmental data

GEO and the GEOSS Data Core

The Group on Earth Observation[175] (GEO), a global effort of voluntary nature grouping around 80 nations and other international organisations coordinating and sharing information on the Earth, is implementing data sharing across many different scientific disciplines, by means of its Global Earth Observation System of Systems (GEOSS). GEOSS is a System-of-Systems based on a brokering/mediation infrastructure, which has proven able to provide harmonized discovery and access to heterogeneous multi-disciplinary data, according to a scalable approach. Besides, GEOSS focuses particularly on the problem of data discovery and access, analysing search tools and techniques involving use of metadata, relevance indicators, keyword searches, to enable researchers and the general public to find their data of interest through the mass of available scientific data and information, and to access disparate content (e.g. heterogeneous encoding formats) through the same platform.

In that framework, GEO has recommended a set of specific principles and technologies for open data discovery, access, and use, which are highly relevant to BYTE[176]. The GEOSS 10-

[174] European Commission, "Digital Agenda for Europe", no date. http://ec.europa.eu/digital-agenda/digital-agenda-europe

[175] Group on Earth Observations, "GEO - Group on Earth Observations", no date. http://earthobservations.org/index.shtml

[176] As per the project Description of Work, BYTE will seek to establish a relationship with GEOSS, in particular with the Infrastructure Implementation Board, which coordinates the Data Sharing activities.

Year Implementation Plan explicitly acknowledges the positive implications of data sharing in achieving the GEOSS vision and anticipated societal benefits: "The societal benefits of Earth observations cannot be achieved without data sharing"[177]. To that extent, the Implementation Plan sets out a set of Data Sharing Principles for full and open exchange of data:

- There will be full and open exchange of data, metadata and products shared within GEOSS, recognizing relevant international instruments and national policies and legislation;
- All shared data, metadata and products will be made available with minimum time delay and at minimum cost;
- All shared data, metadata and products being free of charge or no more than cost of reproduction will be encouraged for research and education.

A task group established within the GEO Work Plan drafted a white paper[178] and developed a set of guidelines[179] for the GEOSS Data Sharing Principles. In GEOSS terms, "full" and "open" are interpreted "taking into account international instruments and national policies and legislation", whereas "minimum cost" is interpreted as "free or cost of reproduction". The GEOSS Data Sharing Principles apply to data, metadata, and products. The intended data users and topics of application are essentially research and education.[180] Upon these guidelines, noting that the Data Sharing Principles may remain an abstract goal until all parties (members, contributors, users) can appreciate how they take form concretely, an Action Plan[181] was developed, which identifies some of the negative implications of open access to big environmental data in the GEOSS context.

A prominent negative implication is of financial nature. Various data providers have the perception that the implementation of the full and open exchange of data, metadata and products in GEOSS could pose challenges to their development, resulting in limited revenue, in particular as payments for reuse are not consistent with the accepted Implementation Guidelines for the GEOSS Data Sharing Principles. Further, many providers cannot see a clear articulation of a business model linked to the adoption of the principle of full and open exchange. Yet in many cases, requiring users to pay for access to data impedes its use, especially if acquiring the necessary funding to purchase data is a long and arduous process. Hence the data provider can at best only realise very limited societal benefits if the product is not attractive to the user.

To rectify the above and mitigate the reluctance of providers to share their data and products openly, the Action Plan suggests that the GEO Community demonstrate that the full and open exchange of data can lead to new applications, additional jobs and more open competition as opposed to the old model of data protection. One action to address this issue is the GEO

[177] Group on Earth Observations, "*10-Year Implementation Plan Reference Document*", ESA Publications Division, Noordwijk (The Netherlands), February 2005, p. 139, 205.
[178] Group on Earth Observations, *White Paper on the GEOSS Data Sharing Principles [Review Draft]*, CODATA, Paris, 27 September 2008; subsequently published concurrently as: Group on Earth Observations, "Toward Implementation of the GEOSS Data Sharing Principles", *Journal of Space Law*, Vol. 35, No. 1, 2009; and *Data Science Journal*, Vol. 8, 2009.
[179] GEOSS Data Sharing Working Group, *GEOSS Data Quality Guidelines*, 19 June 2013.
[180] Doldirina, op. cit., October 2013.
[181] Group on Earth Observations, *GEOSS Data Sharing Action Plan*, GEO-VII Plenary document, Beijing, China, 3-4 November 2010.
https://www.earthobservations.org/documents/geo_vii/07_GEOSS%20Data%20Sharing%20Action%20Plan%20Rev2.pdf

Appathon[182], a global app development competition that aims to develop new, exciting, and (most importantly) useful apps using Earth Observation data. The first ever GEO Appathon runs from 7th May 2014 until 26th September 2014 and is open to non-commercial individuals and teams from any background; although, the event itself is aimed at Earth observation students, scientists and developers. All Apps will be judged and the top 3 winners will receive cash prizes with a top prize of $5,000, and GEO endorsement of the App.

Another negative implication noted is that different disciplines, sectors and countries have developed different socio-cultural approaches to open data in the environmental sector, resulting in language barriers and different rate of development of countries across the globe. GEO recognizes that a commonly endorsed vision is needed to bridge these gaps and overcome such barriers.

Incompatibilities in the legal frameworks in different countries are also seen as inhibitors that need to be adapted, in order to remove legal barriers that could slow the implementation of the GEOSS Data Sharing Principles. In some cases the principle of full and open exchange of data is inconsistent with the current national policies. GEO tries to address this issue encouraging national and international bodies to adopt the principle of full and open exchange of data.

GEO recognizes the need to fully understand negative implications, to be able to remove them. For example, paying for data may hinder their use not just for the price, but because the mechanisms for paying is too cumbersome. In fact, the barriers to data access and use are not simply a matter of pricing policies, but also one of varying policies across data providers and countries, so that negotiating access with each provider is extremely complex and long, thus creating a de facto barrier. Problems regarding the availability, quality, organisation, accessibility and sharing of data and information are common to a large number of policy and information themes and are experienced across the various levels of public authority. Solving these problems requires measures that address exchange, sharing, access and use of interoperable data and services across the various levels of public authority and the different sectors.

To implement open access to the increasing amount of environmental data offered by GEO participants, GEOSS provides a central service framework, termed the GEOSS Common Infrastructure (GCI), that is the primary tool where the interaction between data providers and users are materialized. The GCI plays a critical role in efficiently and effectively support the implementation of the Data Sharing Principles. In fact, it is based on a number of mediating services, named brokers, that transparently address some of the mismatches described above, such as technical and language issues. This in turn has another implication: the requirement for the long-term sustained operation of the GCI itself. Until now, the GCI has been maintained on a voluntary basis, in accordance with the GEOSS implementation methodology. The Action Plan calls for the GEO Members and Participating Organisations to provide resources for the sustained operation of the GCI and the other initiatives set out. However, the governance of GEOSS beyond the time frame of the Action Plan is not yet defined.

GEOSS introduces the notion of Data Collection of Open Resources for Everyone (Data-CORE), a distributed pool of documented datasets with full, open and unrestricted access at

[182] GEO, “GEO Appathon”, 2014. http://geoappathon.org/

no more than the cost of reproduction and distribution. Data CORE has been a key mechanism to advocate openness in data provisioning, and address the non-technical negative implications identified. There has been a big push in the last year 2013 to increase the stock of the CORE, leveraging the voluntary nature of GEOSS. GEO Members are strongly invited to encourage data providers to abide by the Data-CORE terms in publishing their datasets, and specific features of the GCI, such as result ranking, are meant to highlight the value added by Data-CORE.

Copernicus

Copernicus[183], previously known as GMES (Global Monitoring for Environment and Security), is the European Programme for the establishment of a European capacity for Earth Observation. The rationale behind Copernicus is the awareness that the well being and security of future generations are more than ever dependent on everyone's actions and on the decisions being made today on environmental policies. To take the right actions, decision makers, businesses and citizens must be provided with reliable and up-to-date information on how our planet and its climate are changing. Hence, environmental information is of crucial importance. It helps to understand how our planet and its climate are changing, the role played by human activities in these changes and how these will influence our daily lives. The main users of Copernicus services are policymakers and public authorities that need the information to develop environmental legislation and policies, or to take critical decisions in the event of an emergency, such as a natural disaster or a humanitarian crisis.

Copernicus addresses six thematic areas: land, marine, atmosphere, climate change, emergency management and security. Those support a wide range of applications, including environment protection, management of urban areas, regional and local planning, agriculture, forestry, fisheries, health, transport, climate change, sustainable development, civil protection and tourism. The architecture of Copernicus comprises three components: in-situ installations, ensuring observations through airborne, seaborne and ground-based sensors; a space component, ensuring sustainable space borne observations, that consists of both missions contributed by Copernicus members, such as commercial/national satellites, and dedicated satellite missions, namely the Sentinel constellation; and services to ensure access to the massive amount of data and information expected from Copernicus (as an example, the data volume of Sentinel-1, -2, -3 A-series production is roughly equivalent to 25 Envisat missions[184]).

Based on the Copernicus services and on the data collected through the Sentinels and the contributing missions, many value-added services can be tailored to specific public or commercial needs, resulting in new business opportunities. In fact, several economic studies have already demonstrated a huge potential for job creation, innovation and growth. This is a major positive externality expected from the Programme, in terms of: strengthening Earth observation markets in Europe, in particular the downstream sector, with a view to enabling growth and job creation; and support the European research, technology and innovation communities, in making the best use of these data to create innovative applications and services[185].

[183] European Commission, "Copernicus, the European EO Programme", no date. http://www.copernicus.eu/
[184] European Space Agency, *Sentinel Data Policy and Access to Data*, Workshop on GMES Data and Information Policy, Brussels, 12-13 January 2012.
[185] Koch, Astrid-Christina, *Copernicus Data Policy*, presentation at Copernicus Today and Tomorrow, Geneva, 16 January 2014.

As a strategic pan-European Programme requiring significant resource investment, Copernicus is coordinated and managed by the European Commission. The development of the observation infrastructure is performed under the aegis of the European Space Agency for the space component and of the European Environment Agency and the Member States for the in situ component. The Member States and the European Parliament have mandated the EC to define the overall Copernicus data and information policy, which takes full and open access to information produced by GMES services and data collected through GMES infrastructure as the basic principle. As the Copernicus Regulation on the access to GMES dedicated data and GMES service information implies a commitment to follow the GEOSS Data Sharing Principles (see chapter 3.1.1), the Copernicus data policy is compliant with the definition of GEOSS Data-CORE.

In fact, the policy promotes the access, use and sharing of Copernicus information and data on a completely full, free and open basis. To understand the extent of this freedom, it is interesting to highlight the key general principles of the Copernicus data and information policy:

- No restriction on use nor on reproduction and redistribution, with or without adaptation, for commercial and non-commercial purposes;
- Free of charge version of any dataset always available in pre-defined format on the Copernicus dissemination platform (COFUR [Cost of Fulfilling User Requests] is envisaged as well);
- Worldwide (European and non-European users) without limitation in time.

Security restrictions and licensing conditions, including registration, may limit these general principles. For example, access limitations are foreseen for conflict of rights, where the Copernicus open dissemination affects IPR from third parties (potential cascading effect of conditions imposed on input data used in the production of Copernicus service information), and rights and principles recognised by the Charter of fundamental rights of the European Union. Other limitations may apply for security reasons, where the Copernicus open dissemination may affect the security of the Member States of the European Union, or for urgency. Anyway, the decision must be balanced between the protection of security interest and the social benefits of the open dissemination.

While no warranty is made on the data and information provided, the only obligation imposed by the policy is an attribution clause, citing the source of data ("Copernicus") and notifying of any modification made. As for user identification, the policy allows quasi-anonymous use, specifying no registration for discovery and view services and a light registration for download service (registration is possible for users from countries contributing to the Copernicus programme, and stricter conditions may apply, e.g. for protection of security interests).

It is worth noting that data generated by missions contributed by Copernicus members, such as commercial/national satellites, as well as in situ data and information, are considered external to Copernicus, hence they are not covered by the policy. However, Copernicus follows/negotiates the rules set by the data providers for such external data. Instead, the policy applies to service information and, more importantly, to Sentinel mission data.

As the space component of Copernicus, the Sentinel data policy has been jointly decided by the ESA Member States and the EC[186], who agreed on the following principles[187]:

- The licenses for the Sentinel data itself are free of charge;
- Anybody can access acquired Sentinel data; in particular, no difference is made between public, commercial and scientific use and in between European or non-European users (restrictions may apply based on applicable security rules and regulations, as above; in the event security restrictions apply to specific Sentinel data affecting data availability or timeliness, specific operational procedures will be activated);
- The Sentinel data will be made available to the users via a "generic" online access mode, free of charge and subject to a simple user registration process and to the acceptation of generic terms and conditions (additional access modes and the delivery of additional products may be tailored to specific user needs, and therefore subject to tailored conditions).

The above principles have been implemented by the Copernicus regulation[188] adopted in April 2014, repealing the previous regulation[189] on the initial operations (2011 to 2013) of GMES. The fact that the Copernicus policy is supported by a regulation has both positive and negative implications: on the positive side, as a formal normative document, it could be aligned with other relevant directives, such as the EU INSPIRE Directive 2007/2/EC and the EU Public Sector Information – PSI Directive 2003/98/EC. This facilitates consistent implementation of open access in the environmental sector throughout the whole EU.

On the other hand, as a formal EU regulation, its provisions are legally binding for European entities, whereas they cannot have the same efficacy on foreign entities, external to the EU. The principle of worldwide (European and non-European users) access, without limitation in time, coupled with the absence of restrictions on the purpose of use (including commercial exploitation) have raised major concerns, particularly by the European industrial sector, about indirect negative implications on competitiveness. Among other concerns, industry has expressed the view that granting non-European entities free access to Copernicus data and information, including entities from countries such as China and India, with less expensive cost structures, may result in a competitive advantage over European industry.

It was therefore proposed to set clear criteria defining targeted users, their legal status and origin in order to ensure that the implementation the Copernicus Data Policy will not reduce the market share of European earth observation industry. Implicitly, industry thus requested the Commission to review the current version of the Copernicus Data Policy and to consider introducing limitations on data access for non-European entities, namely for commercial entities and their commercial use of Copernicus data and services.

---

[186] European Space Agency, *Sentinel-2 Preparatory Symposium*, April 2012, slide 9; cited in: Desnos, Yves-Louis, *The GMES/Copernicus Sentinels Missions and their Exploitation for Science and Applications*, no date. https://earth.esa.int/documents/10174/642954/ESASentinels062013.pdf

[187] European Space Agency - Earth Observation Programme Board, *The Joint Principles for a Sentinel Data Policy*, ESA/PB-EO(2009)98, rev. 1, Paris, 23 October 2009.

[188] European Parliament and the Council, Regulation (EU) No 377/2014 of the European Parliament and of the Council of 3 April 2014 establishing the Copernicus Programme and repealing Regulation (EU) No 911/2010, OJ L 122, 24.04.2014.

[189] European Parliament and the Council, Regulation (EU) No 911/2010 of the European Parliament and of the Council of 22 September 2010 on the European Earth Monitoring Programme (GMES) and its initial operations (2011 to 2013), OJ L 276, 20.10.2010.

The European Commission is analysing the request from industry and potential legal, policy and other impacts arising from measures restricting the principle of full, free and open access to Copernicus data and information for non-EU commercial entities. On 27 September 2013, the Committee on Industry, Research and Energy submitted a draft proposal amendment on the Copernicus Regulation. In particular, the report submitted that Article 14 of the Copernicus Regulation should be amended as follows:

> Copernicus data and information shall be made available on a full, open and free-of-charge basis for all participating Member States, for emergency situations and for development aid purposes. In all other cases a policy of pay-for-data shall be adopted or a reciprocity principle shall be applied.

However, the issue is very complicated and it is likely that both the industry and the ITRE proposal of potential restrictions in accessing and using Copernicus data and information for non-EU entities could lead to a violation of EU obligations and commitments under the WTO General Agreement on Trade in Services (GATS).[190]

3.5.3 Summary

The environmental sector illustrates several implications of accessing big amount of data in conjunction with open policies, in particular, as regards data stemming from Remote Sensing and Earth Observation, where a "big data revolution" is predicated. The increasing availability of multidisciplinary data available from new observing platforms is expected to empower scientists and society with unprecedented resources to understand our planet and better control or mitigate the environmental dynamics. The examples briefly discussed above highlight that the sector is witnessing a general push to abandon the traditional model of data protection, in favour of full and open exchange of data, believed to lead to new applications, additional jobs and more open competition. The major negative implications of open environmental data sharing can be identified in: interoperability issues, due to the large heterogeneity of technologies, applications, languages, and legal frameworks characterizing the context; as well as of financial nature, given the investments required for Earth Observation, and the strong interest of the industrial sector to protect their investments and competitiveness. Addressing these problems requires mutual policies on the exchange, sharing, access and use of interoperable data and services across the various levels of public authority and the different sectors of society, at a global level.

3.6 TRANSPORT/ LOGISTICS DATA

3.6.1 The relevance and importance of open access to big data in the transport and logistics sector

The Digital Agenda for Europe (DAE)[191] recognises that the transport sector can clearly benefit from big data collected through sensors, GPS data and social media in particular. A smart use of big data supports governments in optimising multimodal transport and managing

[190] Arena, Amedeo, and Ingo Baumann, *Applicability of WTO law to the Copernicus Data and Information Policy*, Specific Contract No GMES/G.4/2013/Lot3- SI2.646761 implementing Framework Service Contract 89/PP/ENT/2011- LOT3(GMES/H4/201KÌ3), BHO Legal, Köln, Germany, November 2013, p. 11.

[191] European Commission, "Digital Agenda for Europe", no date. http://ec.europa.eu/digital-agenda/digital-agenda-europe

traffic flows, making our cities smarter. Citizens and companies can save time through the use of route planning support systems.

Big data is seen as a propeller also in the logistics sector. For example, ship-tracking data, in general, is emerging as a platform for many new services, from mobile ship finder apps, intended for the general public, to more professional services aimed towards the maritime industry. In their position paper on the impact of big data on the industry, DNV GL researchers believe:

> that there are clear business opportunities for actors that can combine in depth knowledge of continuous advanced analysis of ship traffic with other data such as weather, ship parameters, ship equipment, cargo, safety condition etc. […] We also see the emergence of other data aggregators that compile data about the ocean in general. There are millions of sensors throughout the coastal areas and even across the oceans deployed by academic and research institutions, national bodies and by commercial organisations in order to monitor things such as water temperatures, currents, waves, chemical compositions, sounds, fish movements etc. Some of these may also become influential in the maritime and oil and gas domains. One example is Marinexplore who offers the combination of a public and commercial big data platform for everybody with data from the oceans to contribute, combine and analyse. Another example is Google who is expanding it´s Google Earth platform to include oceans data (Google Ocean).[192]

The concept of intelligent transport systems and the consideration to the contribution of logistics to the transport economy in Europe had already been introduced in the EU Common Transport Policy in 1998[193]. In particular, the policy recommended an examination of the need for and feasibility of a European Transport Data System, to improve further the provision of information, data and statistics to decision makers. In FP5, the Commission proposed a specific key action to carry out a number of targeted research actions aimed at improving the efficiency and sustainability of the transport system and at enhancing safety and intermodality. FP7 has promoted the investigation of open exploitation of massive data in the transport/logistics sector, e.g. in the VIAJEO project[194], fighting road congestion; or in the SimpleFleet project[195], address the geomarketing domain, which uses travel information in various geo-statistical analysis methods as well as visualizations of the data. These examples confirm the relevance and importance attributed by the EC to the issue of open access to big data in the transport/logistics sector.

3.6.2 Examples of policies and initiatives relating to transport data

As typically observed in the context of smart cities, transport data is amongst the most popular for re-use by mobile application developers, such as real-time train and bus information services and applications around road congestion, traffic information, and issues such as finding the best fares and personal navigation services[196]. However, possibly due to the conflicts of interest in the mixture of public and private companies operating, it seems that

[192] DNV GL, *Big Data - the new data reality and industry impact*, Strategic Research & Innovation Position Paper 4, 2014, p. 18.
[193] European Commission, Sustainable mobility: perspectives for the future, Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions, COM(1998) 716 final, Brussels, 01.12.1998.
[194] Viajeo Consortium, "Viajeo.eu :: Welcome", Viajeo Project, 2009. http://www.viajeo.eu/
[195] SimpleFleet Consortium, "SimpleFleet |", SimpleFleet Project, no date. http://www.simplefleet.eu/
[196] UK Department for Business, Innovation and Skills, *Smart Cities: Background paper*, October 2013, p. 23.

there are also major concerns with open data sharing in the transport sector. Considering for example car traffic, although many applications are based on peer-to-peer schemes and shared data generation and exploitation, the actual data are actually proprietary, inaccessible, and used for producing (and selling) services. Examples include: the popular Waze[197] social mobile application providing free turn-by-turn navigation based on the live conditions of the road; the TomTom Floating Car Data technology[198], an application of big data based on positions collected by anonymous devices, with over six trillion measurements since 2008, and over six billion new records per day, enabling commercial, global, applications developed by big data analysis, such as the TomTom Congestion Index.

Concerns on the costs and benefits of opening up traffic data, and ultimately the issue of business models, are echoed in the declaratory statements of several initiatives, such as:

- Polis[199], a network of European cities and regions working together to develop and deploy innovative technologies and policies for local transport and a more sustainable mobility, among others, has produced a position paper[200] on access to transport data. Therein, they remark the recent uptaking of the open data movement in the transport domain, noting that local authorities are increasingly moving towards open data as part of a wider public sector transparency agenda. They also note that open transport data is rarely a legal requirement but is rather policy driven. They conclude that *"*caution should therefore be exercised when adopting legislation: regulation could serve to stifle innovation in the fast evolving area of open data, and could even cause some existing open data initiatives to shut down."[201]
- The ePSI Platform[202], a EC DG CONNECT initiative for promoting a dynamic Public Sector Information (PSI) and Open Data re-use market across the EU, organized an Open Transport Data workshop on 17 September 2012, bringing together an informal group of 30 policy makers, experts, opinion leaders and other stakeholders within the transport data community, from over 15 countries. The workshop tried to capture and highlight the potential, the obstacles and the solutions at hand and, accordingly, propose an allocation of responsibilities to meet these challenges to the appropriate stakeholders. As noted: "many operators and incumbent service providers, in particular those relying on income from sales of data, still regard selective and exclusive access to transport data as a competitive advantage, restricting access and re-use through the exercise of intellectual property rights."[203]

These examples confirm the general reluctance by the private sector to embrace open access policies regarding the vast amount of transport data they manage, and may indicate that regulatory actions to force them to share their data are not an effective way to address their concerns and promote a change of mentality.

[197] Waze, "Free Community-based Mapping, Traffic & Navigation App", 2014. https://www.waze.com/
[198] Krootjes, Peter, *Applications of large scale Floating Car Data collection from consumer navigation systems*, Big data from Space Conference, Frascati (Rome), 5-7 June 2013.
[199] POLIS, "Polis Network - Home", 2011. http://polisnetwork.eu/
[200] Polis Traffic Efficiency & Mobility Working Group, *The Move Towards Open Data In The Local Transport Domain*, Polis Position Paper, June 2013.
[201] Polis Traffic Efficiency & Mobility Working Group, op. cit., June 2013, p. 2.
[202] European Commission, "Europe's One Stop Shop on Public Sector Information (PSI) Re-use", epsiplatform.eu, 2014. http://www.epsiplatform.eu/
[203] European Public Sector Information Platform, *2012 ePSI Open Transport Data Manifesto*, Helsinki, 17 September 2012. http://www.epsiplatform.eu/transport

UK National Public Transport Data Repository

We exemplify a specific case of open access to big data in the transport sector: the UK National Public Transport Data Repository (NPTDR)[204], the UK's largest transport dataset. It contains the position of every bus stop, station and airport, and a snapshot of every public transport journey in Great Britain for a selected week in October each year (data is currently available for October 2004 to October 2011). The dataset is compiled with information from many sources, including local public transport information from every region in the country (excluding Northern Ireland), coach services from the national coach services database and rail information from the Association of Train Operating Companies.

NPDTR used to require pre-registration and restrictions on commercial use, with access through and only for public sector agencies and/or projects. Since August 2010, NPTDR has been framed in the UK Open Data programme (data.gov.uk), as part of the Department for Transport contribution to the National Information Infrastructure, which contains the government data likely to have the broadest and most significant economic and social impact, if made available and accessible outside of government, where possible[205].

As such, NPDTR is free and available to use without restriction, under the Open Government License for Public Sector Information, which allows to copy, publish, distribute, transmit, adapt, and exploit the information, both commercially and non-commercially (e.g., by combining it with other information, or by including it in another product or application). The only condition is to acknowledge the source of the information, by including any attribution statement specified by the provider and, where possible, by providing a link to the licence.

NPTDR is mainly used by local authorities and other organisations in the production of accessibility indicators and transport planning (for example via software tools such as Accession[206]). However, it benefits even individual network and accessibility planning specialists, who now have at their disposal all the data needed to do their work. In fact, although a full analysis of all the folders of data that detail routes and services may require considerable computational power, as typical in any big data application, the NPTDR may be creatively elaborated to retrieve useful information even on an ordinary computer running a simple spreadsheet software package. In an instructive blog article[207], NPTDR is merged with population statistics and administrative boundaries, obtained from other open access database, to show which locations in Britain have the highest and lowest density of bus stops.

If coupled with the availability of free and open software, the advantage with respect to the previous required investment (and the return on the investment) is significant. This suggests that the use of open data by transport planners may already be routine: a quick poll conducted by the author of a magazine article[208] to gauge use by transport planners showed that most

[204] Transport Direct, "National Public Transport Data Repository (NPTDR) - Datasets", data.gov.uk, no date. http://data.gov.uk/dataset/nptdr
[205] Cabinet Office, "National Information Infrastructure", gov.uk, 31 October 2013. https://www.gov.uk/government/publications/national-information-infrastructure/national-information-infrastructure-narrative
[206] Citilabs, "Accession | Citilabs", no date. http://www.citilabs.com/products/accession
[207] Rogers, Simon, "UK transport mapped: Every bus stop, train station, ferry port and taxi rank in Britain. Welcome to the ultimate transport data", *The Guardian*, 27 September 2010. http://www.theguardian.com/news/datablog/2010/sep/27/uk-transport-national-public-data-repository
[208] Drew, Keith, "Opening up opportunities to benefit the individual", *TransportXtra*, Issue 598, 11 June 2012. http://www.transportxtra.com/magazines/local_transport_today/news/?ID=30979

were aware of open data, particularly Ordnance Survey data, with a vast majority actively using open data in some ways.

As the same author notes, the availability of open big transport data open new possibilities to explore and innovate, which are expected to turn to value in the future. Perhaps this is not so significant for contractors in the public sectors, where contract revenues are not expected to change, as the data has always been freely available through contractor licenses. However, innovation fostered by the use of open data may still further benefit the public sector. Furthermore, the accessibility to large amount of transport information may open new opportunities in the private sector, where data charges are typically prohibitive.

3.6.3 Summary

Applications in transport and logistics can clearly benefit from the open availability of increasing quantities of data. Sharing, analyses and cross-combination of transport data related to people and goods support both the public and the private sector in optimising multimodal transport, managing traffic flows, and reducing delivery expenses (e.g. time, fuel and human resources) through the use of route planning support systems. Dynamic traffic optimization in urban areas is an obvious application, whose positive externalities directly benefit the citizens' well being, as well as the overall productivity of the economic ecosystem, making our cities smarter. On a global/regional scale, logistics and transport data can be exploited, for example, to improve the safety of marine shipping, together with open big data in the environmental sector. In the example presented, a comprehensive, although limited, sample of public transport data proves a significant support for local authorities as well as for private professionals, and possibly represents a success story of the whole "open data" approach, especially when coupled with open software packages capable of analysing that data. However, negative implications are also perceived, mostly by the private sector, regarding the costs and benefits of a generalized uptaking of the open data approach in the transport domain, and the resulting business model. The reluctance of commercial stakeholders does not seem to be easily addressable in a policy-driven way, that is via regulatory actions. Deeper analysis of positive experiences like the one presented may illuminate more clearly the economical advantages of open exploitation of massive transport and logistics data, with respect to the traditional closed market model.

3.7 CULTURAL DATA

3.7.1 Relevance and importance of open access to big data in the cultural sector

The re-use of digital content is an essential tenet of the Digital Agenda for Europe. Several activities are already stimulating the re-use of cultural heritage in order to demonstrate the social and economic value of cultural content. One example is Europeana Creative[209], which aims at facilitating the re-use of digital objects by the creative industries.

[209] Europeana Professional, "Europeana Creative Homepage", no date. http://pro.europeana.eu/web/europeana-creative/home

According to (Lilley 2013)[210], the current approach to the use of data in the cultural sector is out-of-date and inadequate: the sector as a whole and its policy and regulatory bodies are failing to fully exploit the considerable financial and operational benefits that could arise from better use of data. In addition, a significant opportunity to better understand and possibly increase the cultural and social impact of public expenditure is missed. They advocate a step-change in the approach of arts and cultural bodies to data and for them to take up and build on the management of so-called "big data" in other sectors.

One of the fundamental barriers to the use of big data approaches in arts and cultural institutions is related to the funding environment, where the perspective towards data is often seen as too limited. Often, the gathering and reporting of data is seen as a burden and a requirement of funding or governance rather than as an asset to be used to the benefit of the artistic or cultural institution and its work. This point of view, arising partly from a reported "philosophy of dependence, subsidy and market failure that underpins much of the cultural sector, including the arts and public service broadcasting"[211], is at risk of holding the sector back.

The analysis of big data in other sectors is starting to uncover the possibility of new ways of measuring the impact of arts and cultural investment on our wider society in terms of social capital and cultural value creation. Increasingly-sophisticated approaches to the measurement of the structure of and activity on social networks, including sentiment analysis and behavioural analytics, are giving glimpses of a future in which it will be increasingly possible to track, measure and influence the spread of ideas and the coming together of groups of people and associated changes in their behaviour both on- and off-line.

3.7.2 Examples of policies and initiatives relating to big cultural data

The interest in programmes such as Wikimedia Foundations's "Wikipedian in residence"[212] by a growing number of cultural heritage institutions (e.g. the Royal Library in the Netherlands, the National Library of Scotland) now hiring dedicated people to coordinate their Wikimedia activity, may be an indication of the perceived positive externalities of providing open access to the existing amount of cultural heritage data.

Europeana[213], Europe's flagship resource for making Europe's cultural heritage accessible to the world, is arguably the most obvious initiative to exemplify open access to cultural heritage data and illuminate some of the related issues, especially as regards policies.

Europeana

Europeana is a Web portal that acts as an interface to different types of content (books, paintings, films, museum objects and archival records) that have been digitised throughout Europe by different types of heritage institutions.

[210] Lilley, Anthony, and Paul Moore, "Counting What Counts: What Big Data can do for the Cultural Sector", magiclantern.co.uk, February 2013.
http://www.nesta.org.uk/sites/default/files/counting_what_counts.pdf
[211] Lilley, op. cit., February 2013, p. 3.
[212] Wikimedia Foundation, "Wikipedian in Residence", Outreach Wiki, 12 August 2014.
http://outreach.wikimedia.org/wiki/Wikipedian_in_Residence
[213] Europeana Foundation, "Homepage", no date. http://www.europeana.eu

Europeana provides a framework for many other current initiatives promoting open access to big data in the cultural sector. One example is the European Library, the biggest content provider to Europeana (although it does not focus purely on open access), with over 6 million items added to the Europeana database[214]. The European Library is the aggregator of digital content from national libraries for Europeana and delivers digital content from national libraries on a monthly basis to Europeana. Another example is the OpenGLAM initiative[215], coordinated by the Open Knowledge Foundation[216], that promotes free and open access to digital cultural heritage held by Galleries, Libraries, Archives and Museums.

Europeana takes a federated and distributed approach: the digital objects that users can find in Europeana are not stored on a central computer, but remain with the cultural institution and are hosted on their networks. Europeana collects contextual information – or metadata – about the items, including a small picture, useful for discovery. Users search this contextual information and once they find what they are looking for, they can click through to the original site to access the full content of the item of interest.

As typical of federated systems, the participants retain a large degree of autonomy. For example, Europeana does not make any decision about digitisation. The decision about which objects are digitised lies with the organisation that holds the material. The terms for user contributions[217] governs the terms on which users contribute data to Europeana. These ensure that Europeana can combine objects and data provided by individual users with the rest of the Europeana metadata.

In fact, different types of cultural heritage organisations – libraries, museums, archives and audio-visual collections – catalogue their content in different ways and to different standards. Approaches also vary in different countries. To make the information searchable, it has to be mapped to a single common standard (the "federal model", in federated systems terminology), known as the Europeana Semantic Elements. This metadata standard at present takes a lowest common denominator approach to the integration of different types of digital content. However, the introduction of a richer metadata standard, the Europeana Data Model, will help to give users more and better information.

The Europeana Data Exchange Agreement (DEA) [218] is the central element of the Europeana Licensing Framework and governs the relationship between Europeana and its data providers. The Agreement is the result of a yearlong consultation process with the network of data providers and aggregators and, as of 1 July 2012, it replaced all the existing agreements in place. The DEA sets out a number of simple licensing principles that basically establish free and open access to the descriptive metadata provided to Europeana by data providers.

In summary, all descriptive metadata provided to Europeana are published under the terms of the Creative Commons CC0 1.0 Universal Public Domain Dedication. This means that all metadata provided to Europeana can be re-used by third parties without any restrictions. Europeana does not intend to make direct commercial use of providers' metadata, however the

[214] Europeana Foundation, "Providers, Overview of collections included in Europeana", 25 August 2014. http://www.europeana.eu/portal/europeana-providers.html
[215] OpenGLAM, "OpenGLAM", Open Knowledge Foundation, no date. http://openglam.org/
[216] Open Knowledge Foundation, "Open Knowledge: Home", no date. https://okfn.org/
[217] Europeana Foundation, "Europeana Terms for User Contributions", no date. http://www.europeana.eu/portal/rights/terms-for-user-contributions.html
[218] Europeana Professional, "Data Exchange Agreement", no date. http://pro.europeana.eu/data-exchange-agreement

contribution of data to Europeana does not prevent from selling metadata to a third party. Besides, data providers are not required to provide complete metadata for digital objects, nor providing metadata about some works in a collection implies the obligation to provide metadata about the complete collection.

Data providers also grant Europeana the right to publish the thumbnails and previews possibly provided to Europeana. However, thumbnails and previews related to an object may not be re-used by third parties, unless the rights label that describes its copyright status (that each digital object available via Europeana needs to carry), allows such re-use. In any case, Europeana is committed to consultation with the network of data providers.

These provisions have the positive consequence of facilitating and supporting the implementation of discovery services, such as catalogues and search engines, which further contribute to advertise the existence and characteristics of the original cultural heritage. Thumbnails and previews are not directly used for discovery (support Content-based image retrieval [CBIR], also known as query by image content [QBIC], is experimental), so they are not required, or the provider may restrict their distribution. However, they are made available to support evaluation, when present.

From the user viewpoint, Europeana defines Terms of Use[219] that govern what a perspective user can or cannot do with the metadata and previews that are published on Europeana.eu (separate terms are defined for the Europeana Application Programming Interface). Moreover, Europeana publishes non-binding guidelines that set out the responsibilities that people who want to re-use the data should be aware of, when working with Europeana metadata as well as for using digital reproductions of Public Domain works found via Europeana. In addition, Europeana provides a list of the data sources used, to enrich the metadata provided by data providers.

Europeana follows several other policies, for example on user privacy, language and accessibility. A policy of particular interest in our context is the Europeana Public Domain Charter[220], which clearly states that Europeana and the Europeana Foundation believe in and wish to strengthen the concept of the Public Domain in the digitised world. The rationale for this is that having a healthy and thriving Public Domain is essential to the social and economic well being of our societies, as the Public Domain (with such examples as Diderot's Encyclopédie, the paintings of Leonardo, Newton's Laws of Motion) is the raw material from which we make new knowledge and create new cultural works.

In fact, a large part of the content currently accessible via Europeana consists of pre-20th century works, mostly no longer protected by copyright and in the public domain. Part of Europeana's mission is to strengthen the digital Public Domain[221], ensuring that works that are in the public domain are made available without restrictions and will continue to remain so. With this approach, Europeana strives to help both users and providers to understand the public domain and to ensure that it continues to function in the technological environment of the networked information society. This is not intended to prevent organisations from commercially exploiting public domain works that they have in their collections. Instead, it

[219] Europeana Foundation, “Terms of use”, no date. http://www.europeana.eu/portal/rights/terms-of-use.html
[220] Europeana Foundation, “Public Domain Charter”, no date. http://www.europeana.eu/portal/rights/public-domain-charter.html
[221] Europeana Professional, “Europeana and the Public Domain”, no date. http://pro.europeana.eu/public-domain-content

provides a set of minimum standards, ensuring that the public domain functions in the digital environment.

One of the practical actions that Europeana is taking with these regards is to encourage content providers to make their content fully open access, rather than using restrictive licences, such as CC-BY-NC. Of course, many cultural heritage institutions desire recognition for their considerable efforts in preserving and digitising public domain works and making them available. However, Europeana advocates that claiming copyright in digital reproductions of public domain works is the wrong strategy to realise this and promote responsible re-use of public domain works, including attribution for institutions that have invested in processing them and other usage guidelines for public domain content[222].

Another practice that Europeana is recommending to face a common challenge is proper rights labelling of the public domain content, either in the digitised version or not. Europeana and Creative Commons have jointly developed the Public Domain Mark (PDM)[223], a simple mark that indicates that a work is in the public domain. The DEA requires that data providers label works that are in the public domain by applying the PDM. Works that are labelled as being in the public domain can be re-used by anyone without any restrictions. Rights information icons are displayed alongside search results.

The distinction between the rights statement concerning public domain content (e.g. the PDM) and the one concerning metadata (e.g. CC0) is important. CC0 is specifically designed for use with (meta) data sets and in the context of Europeana is unlikely to be used as a rights statement describing content, but is primarily indicating that metadata can be used without any restrictions (in fact, the CC0 waiver is automatically applied to all metadata that is being provided to Europeana). More specifically, the Creative Commons CC0 1.0 Universal Public Domain Dedication can be applied to objects or data that is subject to copyright, but where the rights holder wants to waive the rights and dedicate the object to the public domain. Instead, the PDM is used to mark the cultural heritage digital objects submitted to Europeana that are in the public domain. In other words, the PDM applies to objects that are not subject to copyright either because copyright has expired (e.g. the author died many years ago) or because the object was never subject to such rights and is therefore in the public domain.

Apart from contributing to consolidate a clear and sound copyright framework in the context of cultural heritage data, this distinction is important because it facilitates the collection of public domain artefacts. In fact, only the rights holder or someone authorised by the rights holder can apply the CC0 waiver, whereas anyone can apply the PDM to an object if they know the object is in the public domain. Apart from realising Europeana's mission to strengthen the digital Public Domain, this mechanism to enable the general user to tag PDM material has the positive implication to promote citizen engagement, responsibility, and awareness of European cultural heritage.

3.7.3 Summary

Considerable financial and operational benefits could arise from better use of data in the cultural sector, its policy and regulatory bodies (and the actors in this sector as a whole) would be failing in exploiting them. The process of value creation in the cultural sector,

[222] Europeana Foundation, “Europeana Usage Guidelines for public domain works”, no date. http://www.europeana.eu/portal/rights/pd-usage-guide.html
[223] Creative Commons, “Public Domain Mark 1.0”, no date. http://creativecommons.org/publicdomain/mark/1.0/

sometimes characterized by a certain lack of interest in market success (if not by relying on subsidies), may also impact negatively on the perception of the social and economic value of cultural content. The management of data is often considered just a requirement of funding or governance, rather than an asset that benefit the artistic or cultural institution and its work. However, the increasingly proclaimed benefits of big data in other sectors may imply a change of perspective in the approach of arts and cultural bodies to data. In fact, the analysis of big data in other sectors is starting to uncover the possibility of new ways of measuring the impact of arts and cultural investment on our wider society in terms of social capital and cultural value creation. New technological approaches to cultural data, such as sentiment analysis and behavioural analytics of social network interactions, are suggesting that it will be increasingly possible to also quantify and improve the value of our cultural expressions. The growing interest in initiatives that promote open cultural heritage, like the ones exemplified above, may be an indication of the positive externalities that are starting to be perceived from providing open access to our existing amount of cultural heritage data, and the idea that the cultural works, including the vast amount in the Public Domain (with such examples as Diderot's Encyclopédie, the paintings of Leonardo, Newton's Laws of Motion), are essential to the social and economic well being of our societies, as the raw material from which we make new knowledge and, in turn, create new cultural works.

3.8 UTILITIES/ SMART CITIES DATA

3.8.1 Overview

Smart cities are essentially grounded on the idea that open access to multidisciplinary information can improve decision-making, resource management, and ultimately the living conditions of the citizens. For example, smart cities may host data centres adapting the power consumption of public buildings, depending on the availability of renewable energy and on other indicators computed from open access data sources, as investigated by the FP7 DC4Cities project[224]. In another example, data mining and integration may support a utility company optimizing a wind farm location, by considering several criteria such as wind power, local environment characteristics, potential interference on communication systems, visual impact, or the existence of archaeological sites, as in the Sopcawind FP7 project[225].

Given the complex mixture of public and private interests, especially in the utilities sector, smart cities can still be seen as a grey area in terms of open access to data and their policy of access and use. Publicly available data may include profiles of consumption of renewables, weather forecasting data, or traffic information, as well as the correlation of this data with the utilities' proprietary information, such as high-resolution metering data and customer billing information. In general, utility companies and other commercial organisations tend to be conservative with respect to making data available, whereas there is an opposite tendency by the public administrations. This implies that smart cities data are usually subject to various policies, such as the PSI Directive for data held by governments, open access arrangements for research initiatives, and proprietary policies for other purposes.

In this complex scenario, the European Commission is trying to coordinate the development and implementation of a smart cities strategy through the European Innovation Partnership (EIP) on Smart Cities & Communities, in cooperation with Directorates General ENERGY,

[224] DC4Cities, "DC4Cities", 2013. http://www.dc4cities.eu/
[225] Sopcawind Consortium, "SOPCAWIND", Sopcawind Project, 2012. http://www.sopcawind.eu/

MOVE (mobility and transport) and CONNECT (communications Networks, content and technology), and building on well structured research and development initiatives, notably Energy Efficient Buildings, Intelligent Transport and Smart Grids.

The European Commission has clearly expressed[226] its vision of "smart cities" that provide public services to their citizens in a more convenient way, that are more responsive and citizens-centred, that provide the right information in real-time to allow for better everyday and business decision-making, and that achieve all this in an economically viable way so as to improve environmental sustainability. Despite the flurry of activity and media attention around smart cities, the EC admits that are few, if any, cities that are recognised as coming close to the above vision of smart cities.

In fact, the fragmentation of initiatives, scarcely synergetic (e.g., through joint uses of physical and digital infrastructures), with sometimes insufficiently clear goals other than infrastructure deployment, too little attention to commercial viability and, thus, eventual scale-up, is one of the main barriers that hinder the achievement of this vision. There is a lack of cross-sectorial business models, as well as of metrics/performance indicators for smart city projects to assess environmental and other benefits, and return on investment, which makes the administrations reluctant of investing, for fear of vendor/technology lock-in. This in turn implies a lack of well-tested template solutions and interoperability problems, i.e. missing adaptability to new user requirements and technological change, insufficient attention to citizens needs and inclusiveness. A specific negative implication of big data in the utilities sector is the growing energy requirements of ICT, due to proliferation of data and data storage requirements. The digital infrastructure and economy is estimated to use a tenth of the world's electricity supplies, likely to grow following the explosion in data production. It is estimated that by 2020 we will produce 35 zettabytes (1 trillion Gb) annually.[227]

On the other hand, as the EC recognises, data will be the key ingredient for any "smart city solutions". Policy-driven approaches, especially to promote interoperability and standardization, are emerging as an effective measure to mitigate the negative externalities of the big data revolution in the sector of utilities / smart cities, and to address the current challenges and needs, for example identifying schemas to promote openness and accessibility of data as well as common structures and formatting (building on public sector information and open data policy); or standardising the way of representing data and information. Standardisation enables to pursue more clarity around the solutions for data privacy and security as well as integrity of data, transparency around energy costs of data collection and storage, and ease-of-use and accessibility to all the smart city solutions.

3.8.2 Examples of policies and initiatives relating to smart cities/ utilities data

The following cases are examples of "smart" approaches to open data by two urban administrations: one in Asia, Jakarta, focusing on improving mobility and safety, one in Europe, Florence, focusing on promoting transparency and citizen engagement.

Florence

[226] European Commission - DG CONNECT, "Smart Cities", no date. http://ec.europa.eu/dgs/connect/en/content/smart-cities-0

[227] The Best Computer Science Schools, "Big Data, Small Footprint?", no date. http://www.bestcomputerscienceschools.net/big-data/

Together with Rome and Aquila, the city of Florence has been recently awarded the title of 'Smart City' from the Milan-based Information and Communication Technology fair (SMAU), for its work to increase the city's social infrastructure and make data and information more accessible to the public[228]. The Tuscan city of Prato was also shortlisted for having launched a management and control system of street lighting.

The Smart City designation is an acknowledgement of the cities' efforts to develop innovative ideas for their citizens. Florence has been awarded it for its Open Data website[229], which to date includes 400 datasets and has become Italy's third biggest producer of data. Florence is the only city government in Italy to have a 'five-star rating' for its level of quality and accessibility of information. In a statement, former mayor Matteo Renzi noted that Florence has long worked to "open" its data up to the public: *"*We aim to be completely transparent and have all data verified online so it is clearly presented to the public. It is a "revolution" that aims to change the rules of public administration in Italy and we need to invest more and more in this direction."[230]

Apart from an effort to increase transparency (in the words of the city's director of innovation, "the data belongs to the people"[231]), Florence administration also sees open access as an opportunity for economical growth, for service providers to provide information that may be useful to everyday life, or for those wishing to open a business in the city, to find information that will help them better analyse the market. Besides, open access to data in urban contexts helps the Florence administration promoting citizen engagement and participation, for example by means of interactive applications to report graffiti, illegal dumping and other kinds of urban blight, through messages and photographs.

Users can access a range of datasets from different categories, such as the environment, the work of the local administration, with live streaming of local council meetings, culture and tourism, education, traffic and safety, public works, health and social, sports and urban planning. Among the data is such information as the most popular baby names in Florence, the number and locations of all the commemoratory plaques in the city, the number of infrastructure projects currently underway, the number of non-Italian residents and number of speeding fines issued each year. There is a specific section called Open Budget, which gives full details on the municipal budget.

The data on the site is free to use, reuse and redistribute through a Creative Commons license. It is data and information produced or commissioned by Palazzo Vecchio, and the site's purpose is to encourage transparency and democratic control, civil participation, research on new products and services, innovation, and governmental efficiency.

---

[228] SMAU Servizi, "Premio smart city di Firenze: dalla Toscana i migliori progetti in tema di città intelligenti per trasformare il territorio in una fucina di innovazione", *SMAU news*, 8 July 2014. http://www.smau.it/news/premio-smart-city-di-firenze-dalla-toscana-i-migliori-progetti-in-tema-di-citta-intelligenti-per-trasformare-il-territorio-in-una-fucina-di-innovazione

[229] Comune di Firenze, "I dati aperti del Comune di Firenze", no date. www.opendata.comune.fi.it

[230] La Nazione Firenze, "Al Comune il premio Smart City dello Smau per il portale 'Open Data'", 20 March 2013. http://www.lanazione.it/firenze/curiosita/2013/03/20/861890-firenze-tra-i-comuni-vincitori-del-premio-smart-city.shtml

[231] Dionisi, Brenda, "Open data – Palazzo Vecchio puts government data online", *The Florentine*, issue no. 159/2012, 15 March 2012. http://www.theflorentine.net/articles/article-view.asp?issuetocId=7536

As Florence is globally recognized for its cultural treasures, the open data portal is also used for the promotion of cultural assets, as part of a partnership between the city and the Wikitalia association, what has gained to Florence the definition of Italy's first "wikicity". This illuminates an interesting implication of open data access in smart cities, as related to the overlapping with open access and big data in the cultural sector, discussed in chapter 2.2.3.

Jakarta

Indonesian capital city, Jakarta, prioritises big data and open government for public safety and transport.[232] The nearly 10-millions inhabitants capital of Indonesia held its first open data challenge, HackJakarta, on 26 April 2014, opening up over 50 datasets in machine-readable format and a crucial initiative, as Indonesia is chairing the Open Government Partnership[233]. Jakarta is using open government and big data to help its city managers make more informed decisions, particularly in public safety and transport.

Jakarta's Intelligent Transport System, adapted from Australia, connects to the traffic management centre and local police for CCTV and traffic signals. There are currently 1,000 CCTV cameras in Jakarta, expected to grow to 5,000, improving also the security of the city. Residents are able to get public transport information and the city is able to prioritise and manage its public transportation network. Another important area for big data application is flood management, as 13 rivers flow into the city. Jakarta's new early warning system monitors risk indicators so that agencies are able to take quick action before the situation escalates.

A crucial step in Jakarta's smart city roadmap is its Intelligent Operations Centre (IOC), which will serve as the command and control centre bringing together different data sources for a holistic view of movements in the city. The IOC takes inspiration from the system in use in Rio de Janeiro, and will leverage sensors to collect, combine and analyse data to anticipate problems and automated coordinated responses.

The city sees around 15,000 projects being implemented every year and numerous parameters need to be taken into account for managing these projects. Jakarta's project monitoring system combines all the parameters into one of three possible signals (similar to red, yellow or green on a traffic signal) so that city managers are able to quickly interpret the data and take action. A big challenge for Jakarta's city government is extracting value from and taking action on the huge volumes of data in the city systems.

Jakarta's commitment to implement open access to public information is part of a wider country's effort to help meet increasing public demand for easily accessible official government data, synergetic with the global open data movement[234]. Indonesia has recently launched the country's very own open data project, the Indonesia Data Portal, which is proving a useful tool for business, encouraging innovation, as testified by the numerous applications for business start-ups in using the data. An important lesson learnt is that

[232] Basu, Medha, "Indonesian Capital City Prioritises Big Data And Open Government For Public Safety And Transport", *FutureGov*, 17 April 2014. http://www.futuregov.asia/articles/2014/apr/17/indonesian-capital-city-prioritises-big-data-and-o/

[233] Open Government Partnership, "Open Government Partnership |", 2014. http://www.opengovpartnership.org/

[234] Salim, Tama, "Government joins open data project", *The Jakarta Post*, 9 September 2014. http://www.thejakartapost.com/news/2014/09/09/government-joins-open-data-project.html

inspirational leadership is important for the success of the smart city vision, along with good communication, fair negotiation and consistent law enforcement.

The example of Jakarta, and of Indonesia in general, is very interesting to analyse to what extent open access to big data is genuinely useful to tackle practical problems in developing and emerging countries. Noticeably, Jakarta hosts the World Wide Web Foundation's Open Data Lab[235], a project in cooperation with policy-makers, local and national government officials, researchers, and practitioners across the Southeast Asia region, that aims to identify new ways to unlock the benefits of open data for Indonesia, Asia and the Global South in general.

Web Foundation research shows that the use and impact of open data in developing and emerging countries remain very low, notwithstanding a growing number of developing countries are pursuing open data initiatives to proactively provide easy access to more, better, and machine-readable government data, seen as a key ingredient in tackling development-related challenges, including inefficiencies in the public sector.

One factor that negatively impacts open data policies in improving governance of developing countries is the very weak or absent vertical accountability of officials to citizens, as well as a traditional lack of "evidence-based" policy-making. It is unclear whether app contests and hackathons, popular in North America and Europe to stimulate the growth of a community of data re-users (see for example the GEOSS example, in chapter 2.2.1), which in turn is often seen as sufficient to produce impact, are the most effective route in emerging country contexts.

3.8.3 Summary

As we have observed in chapter, positive and negative externalities of big data in the context of utilities and smart cities are often similar, or related, to the issues implied by open access to big data in the environmental sector, as far as the management of the local environment is concerned. Besides, the utilities/smart cities sector often subsumes the positive and negative implications typically observed in the context of open access to big transport data, that we discussed in chapter 2.2.2. As well exemplified by one of the cases discussed, the smart analysis of traffic data flows (e.g. to dynamically adjust traffic signals, the frequency of public transport, road maintenance planning, safety) is among the most common (and possibly most appreciated by citizens) features of a smart city.

The examples discussed illuminate that open access to big data is seen as having positive implications in the specific context of smart cities, for instance on decision-making, resource management, and the living conditions of the citizens, through resource optimization and planning, support to economical growth, innovation and value-added services (e.g. for tourism), transparency and citizen engagement, safety and control, creation and promotion of cultural value.

Open standards and data interoperability are priorities in various policies. Flexible data reuse in the complex urban ecosystem requires standard and shared formats and protocols for gathering and processing data from different sources in a coherent and interoperable manner across sectors and vertical markets (energy, transport, environment, smart cities, retailing,

[235] World Wide Web Foundation, "Open Data Lab Jakarta", 2014. https://webfoundation.org/projects/open-data-lab/

security, etc.) This is reflected in ongoing EU initiatives to set EU-wide standards within important economic sectors, such as transport, geographic information, and public sector information in general, like the ISA programme[236] aiming to facilitate the use of common core data standards for national administrations. To help create a climate of open data exchange, the EC specifically supports the mapping of existing relevant standards for a number of big data areas (e.g. smart grid, health, transport, environment, retail, manufacturing, financial services).[237]

A UK government report[238] reinforces the importance of standardisation: as the vision of smart cities involves data from a wide range of sources and systems, it requires a focus on protocols for data sharing and communication. This has the positive implication to accelerate the development of standards for smart city solutions, through providing the necessary guidance, frameworks, specifications, protocols and vocabulary to create a common understanding of the issues, manage the risks effectively and optimise communication across the city. It is probably the "variety" of data that makes for the most significant aspect of open access to big data in the utilities/smart cities sector.

[236] European Commission, "Interoperability Solutions for European Public Administrations - ISA", 22 September 2014. http://ec.europa.eu/isa/

[237] European Commission, Towards a Thriving Data-Driven Economy, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, COM(2014) 442 final, Brussels, 2.7.2014.

[238] UK Department for Business, Innovation and Skills, op. cit., 2013, p. 16.

4 GOOD PRACTICE LESSONS FOR OPEN ACCESS TO BIG DATA

This report identifies a number of good practice examples that are relevant to open access initiatives involving big data, and that can be translated across sectors, as well as promoting the collaboration between private and public sectors in the development of open access projects. This is important as it might result in diversity of the information available through open access models, and the availability of more data can have a significant, cross-sector impact. Ultimately, good practice policies support maximising the value to be derived by open access to big data, and potentially minimize negative externalities that are produced by open access policies and initiatives for big data.

Good practices need to be adapted to the diversity of the big data enterprise in the public and private sectors and accommodate the interests of the innovators and institutions that fund and publish big data, as well as private organisations who are increasingly developing business models with open access elements. This is especially important to better exploit positive externalities of open access to big data as well as diminishing the negative externalities such as barriers presented by legal or other issues. Good practice policies may also encourage a shift away from the protection of big data as a private reserve and the encouragement of open access initiatives in the private sector, or as a result of collaborations between the public and private sectors. Thus, it is important to balance the rights and responsibilities of those who generate big data and those who access big data. Attention must also be had for the sustainability of open access particularly in terms of ensuring adequate financial support for the initiative as well as supporting technological infrastructure. This must extend to metadata and ensure the integrity of the data so that it is valuable useful in open access policies and initiatives.

In public and private sectors, stakeholder policies will need to embrace the data-driven approaches so that the best use of open data can be made. Best practice policies can address a number of the following issues that relate to the following aspects of open access policies and initiatives:

Technology

- Focus on the development of e-infrastructure and interoperability
- Promote technical security and safety aspects such as privacy by design
- Investment in technology is needed to collect data, create suitable platforms for sharing data, and perform analyses to uncover valuable insights

Collaboration

- Support information and education on the benefits of open access policies and initiatives in both public and private sector
- Encourage cooperation, collaboration and partnership between the private and public sector to produce integrated services

Legal

- Promote regulations governing open access
- Develop standard licensing agreements to protect intellectual property rights
- Raise awareness of ways in which regulations can be complied
- Raise awareness of legal and ethical issues as they arise in relation to open access such as risks to privacy and security

- Develop frameworks that take into account the risks and benefits associated with data, particularly sensitive data such as health data, to ensure compliance with legal and ethical obligations
- Develop policies that appropriately identify risks associated with open access, particularly when the data relates to identifiable people require development. E.g. if a patient is participating in a health care study or clinical trial, they ought be informed about the potential for a re-identification of data
- Develop privacy protecting policies that enshrine consumer trust to ensure secure competitive advantage, especially in areas of concern such as banking and finance and health. This includes moving towards shorter, more concise and user friendly privacy policies

Commercial

- Support the development of business models that incorporate open access aspects by promoting ways in which these elements can add value to business, as well as increase their relevance and market position in the digital market
- Encourage private organisations to open their data once they have utilised it to meet their business objectives. This enables potential new uses of the data
- Raise awareness of how open access can be integrated into business models to generate growth and profit
- Educate organisations on the importance of de-identification of products that enable organisations to retain their competitive edge whist granting access to raw data
- Innovate in the privacy enhancing services and technologies especially where commercial data is exchanged for personal data

Ultimately, good practice needs to be shared between public agencies and private organisations, as well as encouraging and facilitating collaboration between the sectors. Governments, companies, and individuals will need to understand how to take advantage of open data. Governments can develop policies to facilitate the continued release of data, and also regulate to diminish the negative externalities associated with open access to big data. Companies in the private sector can choose to share proprietary data to create benchmarks that can improve overall industry performance. Thoughtful policies and practice guidelines will ultimately give consumers and institutions confidence to move forward with open access initiatives that involve data. Thus, deriving valuable insights from open data will require new rules and procedures and new attitudes as well as investments in technology and capabilities.

5 CONCLUSION

The relationship between open access and big data involves a number of public and private sector based stakeholders. These stakeholders deal with large volumes of health data, crisis data, energy data, environmental data, transport and logistical data, cultural data and utilities and smart cities data. Whilst the open data movement is most strongly connected to the public sector and such policies are motivated by achieving a public good, the benefits of open access policies related to big data are being embraced in the private sector as they are incorporated in business models in the pursuit of profit. The burgeoning relationship between big data open access policies and initiatives are recognised as being of great benefit to society, and open data is heralded as providing an abundance of opportunities for Europe, despite the threat from negative externalities. In the digital economy, big data represents a significant tangible asset. Encouraging asset holders to provide free and open access to that asset requires both voluntary and proscribed policies and initiatives so that the socio-economic benefits of big data can be fully realised. Whilst open access to big data in the public sector is grounded in the belief that publicly funded research ought be made available to the public to spur further developments and foster education and innovation, access to big data in the commercial sector is spurring innovation in terms of business opportunities. In the private sphere, the data is collected by organisations and their partners, as well as from open access initiatives implemented in the public sector. In particular, in the private sector, data providing information about users' behaviours stands in the place of subscription fees to services because the service providers are able to make profit from the on-sale or internal use of that data to generate profit. However, there is an imbalance between the number of open access policies and initiatives produced by the public sector, compared with fewer in the private sector. What this means is that private sector organisations are at an advantage as have they access to big data made openly accessible by the public sector, as well as their own data stores. On the other hand, public sector organisations can access only data made openly accessible by public sector initiatives and mandates.

Open access initiatives involving health data are more commonly found in the public sector than the private sectors at this time. Some of those initiatives examined in this report include the GOSgene health initiative, Gen Bank and Protein Data Banks (in Europe and internationally), Teralab initiative and the Yale University Open Data Access initiative. An examination of open access initiatives in health care reveal that open health data produces a number of positive outcomes for public and private sector stakeholders. However, it appears that open access to large volumes of health data are provided by the public sector organisations and being utilised for commercial gains in the private sector, whilst private health care companies drive fewer open access initiatives. We reveal that collaborations between the public and private sectors may yield greater benefits for both.

Moreover, crisis data is increasingly being made open via crisis mapping and through social media networks. Thee examples serve not only as a public service, but additionally as initiatives capable of generating profit by acquiring personal data, including behavioural data, from users who access these services during peak times. Examples of crisis data creating value for communities and organisations through open access initiatives include the Ushahidi crisis mapping platform and Twitter. However, the move towards developing open access initiatives such as criss mapping, and social media to disseminate crisis data present challenges as well as opportunities. However, negative implications of the use of crisis data ar often overlooked because such initiatives are produced largely as a public benefit and as part

of humanitarian efforts. Relevantly, it is difficult to identify examples of European open access initiatives involving crisis data, which represents a discrepancy in the market and also an opening for the fostering of such initiatives.

Open access initiatives involving energy data, such as NPD operated FactPages initiative, serves to illustrate how the release of open data can help to improve the exploration and operation activities in the oil domain. This is despite issues arising in relation to publishing open data entail non-negligible time and monetary resources, especially when exposing five stars data. These costs can be substantially reduced if adhering to a three stars open data scheme, as in the case of NPD for the official FactPages. Therefore, energy is a broad field that encompasses many disciplines and thus, releasing open energy data produces a number of opportunities for innovation.

This report also looks at two examples of open access initiatives for environmental data. Environmental data brings about novel ways of understanding and addressing environmental challenges, and two examples that reflect the sentiment of this statement are the following cases of open access to big data in the environmental sector: the GEO initiative and the Copernicus Programme. These examples illustrate several implications of accessing a big amount of data in conjunction with open policies, in particular, as regards data stemming from Remote Sensing and Earth Observation, where a "big data revolution" is predicated. The increasing availability of multidisciplinary data available from new observation platforms is expected to empower scientists and society with unprecedented resources to understand our planet and better control or mitigate the environmental dynamics. The examples briefly discussed above highlight that the sector is witnessing a general push to abandon the traditional model of data protection, in favour of full and open exchange of data, believed to lead to new applications, additional jobs and more open competition. However, a number of negative implications of open environmental data sharing include: interoperability issues due to the large heterogeneity of technologies, applications, languages, and legal frameworks characterizing the context; and financial issues given the investments required for Earth Observation, and the strong interest of the industrial sector to protect their investments and competitiveness. Addressing these problems requires mutual policies on the exchange, sharing, access and use of interoperable data and services across the various levels of public authority and the different sectors of society, and at a global level.

Transport and logistics data is another focus of open access initiatives. The Digital Agenda for Europe (DAE) recognises that the transport sector can clearly benefit from big data collected through sensors, GPS data and social media in particular. Big data is seen as a propeller also in the logistics sector. For example, ship-tracking data, in general, is emerging as a platform for many new services, from mobile ship finder apps, intended for the general public, to more professional services aimed towards the maritime industry. The National Public Transport Data Repository (NPTDR) is the UK's largest transport dataset and provides a strong example of an open access initiative. This example reveals that applications in transport and logistics can clearly benefit from the open availability of increasing quantities of data. Sharing, analyses and cross-combination of transport data related to people and goods support both the public and the private sector in optimising multimodal transport, managing traffic flows, and reducing delivery expenses (e.g. time, fuel and human resources) through the use of route planning support systems. However, negative implications are also perceived, mostly by the private sector, regarding the costs and benefits of a generalized uptaking of the open data approach in the transport domain, and the resulting business model. The reluctance

of commercial stakeholders does not seem to be easily addressable in a policy-driven way, that is via regulatory actions.

Big cultural data is also a focus of open access policies and initiatives. The re-use of digital content is an essential part of the Digital Agenda for Europe. Several activities are already stimulating the re-use of cultural heritage in order to demonstrate the social and economic value of cultural content. One example is Europeana Creative, which aims at facilitating the re-use of digital objects by the creative industries. This section reveals that considerable financial and operational benefits could arise from better use of data in the cultural sector. However, the process of value creation in the cultural sector, sometimes characterized by a certain lack of interest in market success (if not by relying on subsidies), may also impact negatively on the perception of the social and economic value of cultural content.

Examples of open access initiatives involving utilities data and smart cities illuminate the complex mixture of public and private interests. Publicly available data may include profiles of consumption of renewables, weather forecasting data, or traffic information, as well as the correlation of this data with the utilities' proprietary information, such as high-resolution metering data and customer billing information. Our examination of two smart cities, Jakarta and Florence, suggests that smart cities can still be seen as a grey area in terms of open access to data and their policy of access. However, these examples illuminate that open access to big data is seen as having positive implications in the specific context of smart cities, for instance on decision-making, resource management, and the living conditions of the citizens, through resource optimization and planning, support to economical growth, innovation and value-added services (e.g. for tourism), transparency and citizen engagement, safety and control, creation and promotion of cultural value. Despite these positive externalities, the fragmentation of initiatives, scarcely synergetic (e.g., through joint uses of physical and digital infrastructures) with sometimes insufficiently clear goals other than infrastructure deployment, and too little attention to commercial viability can act as barriers to capturing the positive externalities that flow from smart cities that utilise big utilities data.

A number of private sector examples from outside of the EU have been addressed in this report because there are fewer examples to be observed within the EU. What this means is that there is room for competitive growth by incorporating open access elements into business models, as well as promoting cross-sector collaboration in the delivery of open access initiatives. Alternatively, private organisations can grant access to big data once they have used that data to meet their business needs. This opens up a number of other possibilities by industry to use and re-use data in a manner not envisaged by the original collector.

Finally, this report identifies a number of good practice examples that are relevant to open access initiatives involving big data, and that can be translated across sectors, as well as promoting the collaboration between private and public sectors in the development of open access projects. This is important as it might result in diversity of the information available through open access models, and the availability of more data can have a significant, cross-sector impact. Thoughtful policies and practice guidelines will ultimately give consumers and institutions confidence to move forward with open access initiatives that involve data.

This report has identified the relationship between big data and open access. Scientific researchers and industries in accessing and harnessing the economic potential (including improved access to data, better targeting of resources, deeper analyses and further opportunities to increase personnel skills and employment opportunities) associated with big

data in a socially responsible manner. Open access policies offer powerful advantages to unlock previously inaccessible insights from new and exiting data sets in a variety of fields and industries such as environmental data, transport data, energy data, cultural data, crisis data and utilities/ smart cities data. Open access policies can fuel developments and discoveries in these fields and across sectors and empower economies, governments and consumers. Initiatives contribute to the creation of a culture of data sharing. Big data is creating value for companies and consumers. Ultimately, open access can provide more sectors of society with access to information than would be provided in the absence of open access initiatives. However, strong governance is required to protect private and commercial interests when considering open access initiatives. Whilst many open data initiatives, particularly in the public sector, have been motivated by societal goals such as improving the transparency and accountability, initiatives in the private sector cannot be overlooked as spurring growth, innovation and adding value to national and international economies. The development of open access initiatives in the private sector, and collaboration between public sector and private sector organisations require support to produce sustainable and profitable open access.