Big Data and Environmental Sustainability: A Conversation Starter

Smith School Working Paper Series

December 2014

Working paper 14-04

Alan Keeso

Author: Alan Keeso

Acknowledgements

Dr. Paul Jepson supervised this work in its original dissertation form as required for the fulfilment of the MSc in Biodiversity, Conservation and Management Programme at the University of Oxford. Anna Hushlak, a DPhil student in the Department of Geography at Oxford, also assisted in editing the dissertation and working paper. This paper would not have been possible without the time and insight offered by those who were interviewed for this study. Their enthusiasm for this topic lends great hope for future developments and innovation in the field of environmental sustainability. I'd also like to thank Gordon Clark, Caitlin McElroy, and Patrizia Ferrari of the Smith School of Enterprise and the Environment for their support throughout this process.

Abstract

Big data, referring broadly to, "the ability of society to harness information in novel ways to produce useful insights or goods and services of significant value" (Mayer-Schönberger and Cukier, 2013, 9), has been heralded as, "the next frontier for innovation, competition, and productivity" (McKinsey Global Institute, 2011, 1). Despite these claims, a review of literature that highlights big data's revolutionising effects across sectors and industries revealed that environmental sustainability is largely not yet part of the popular lexicon of big data in action. This study addresses this gap. By interviewing 14 organisations across sectors, I examine how big data is perceived, employed, hindered, and enabled. I conclude that while big data adoption has broadly been slow to coalesce with sustainability efforts, emerging factors such as collaborative partnerships and business model innovation are positioning big data to become an integral element of environmental sustainability and vice versa.

Keywords: Big data, environmental sustainability, open data, circular economy, regulation, innovation, technology, strategy, conservation

Table of Contents

1 Introduction

1.1 Overview and Definition of Big Data

Big data, referring broadly to "the ability of society to harness information in novel ways to produce useful insights or goods and services of significant value" (Mayer-Schönberger and Cukier, 2013, 9), has been heralded as, "the next frontier for innovation, competition, and productivity" (McKinsey Global Institute, 2011, 1). While the term's initial coining and rise in popularity was paired with sciences such as astronomy and genomics throughout the 2000s (Mayer-Schönberger and Cukier, 2013), big data has more recently been the subject of popular literature that details both the breadth of the term and its ability to revolutionise individual sectors.

Despite big data's continuing increase in popularity, it has yet to receive a concrete definition (Dumbill, 2013). One of the more popular definitions includes the three 'Vs' of data management: volume, velocity, and variety (Laney, 2001). Volume refers to the increased amount of data to be managed, velocity refers to the increased pace of data generation as well as its use and interaction, and variety refers to the many and often incompatible data formats, structures, and semantics (ibid). Over time however, additional 'Vs' have been added, including veracity, visualisation, and value (Sowe, 2014, 23).

McKinsey Global Institute (MGI) has defined big data as, "datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyse" (2011, 1). MGI explained that big data classification changes not only over time as a result of technology enhancements, but also across sectors due to the variation in dataset size, software, and tools that are common to each sector (ibid). As a result, MGI does not identify a minimum volume requirement for big data classification. Simon (2013) noted that the agendas of different companies shape the definition of big data. Given that different organisations have vested interests in big data thought leadership, Simon (ibid) suggested that a perfect definition of big data does not exist.

Despite the different definitions of big data, agreement has broadly been reached on its potential to enhance decision-making (Dumbill, 2013). Increases in the capabilities and number of cell phones, computers, and information technology (IT) systems have been well documented, but the amount of information now available, combined with the rapid growth of that information (Mayer-Schönberger and Cukier, 2013) has shifted the focus to big data.

1.2 Range and Impact of Big Data Applications

Mayer-Schönberger and Cukier's book (2013), Big Data: A revolution that will transform how we live, work and think, is now one of a growing number that provides accounts of compelling applications of big data analytics. The authors told the story of how Google used search queries to predict the spread of flu in the United States, highlighting just one of many ways in which big data is transforming healthcare. The authors also explored the famous big data story from the retailer Target, who can predict customer pregnancies based on purchasing habits (ibid). As for the future of retail in the United States, MGI stated that furthering big data applications has the potential to, "increase sector-wide productivity by at least 0.5 per cent a year through 2020," growing operating margins amongst individual firms by more than 60 per cent should they remain at the forefront of the big data movement (2011, 64).

MGI has also forecasted profound potential benefits from public sector applications of big data, stating specifically for the European Union that cost reductions for administrative activities could be achieved by up to 15 to 20 per cent, and annual productivity growth could be accelerated by "0.5 percentage points over the next ten years" (54). Researching big data and organisational productivity in general, Tambe (2013) found that a single standard deviation higher utilization of big data technologies can equate to one to three per cent higher productivity than the average firm and vice versa for one standard deviation lower utilisation of big data (Provost and Fawcett, 2013).

1.3 Why Link Big Data and Environmental Sustainability

In summary, a review of the literature revealed a similar pattern: environmental sustainability is largely not yet part of the popular lexicon of big data in action. This study addresses this gap. Big data is playing a transformative role in sectors such as retail, manufacturing, and healthcare. The environmental sustainability effort however, which broadly, "creates and maintains the conditions under which humans and nature can exist in productive harmony, that permit fulfilling the social, economic and other requirements of present and future generations" (Environmental Protection Agency) requires a transformation of its own.

Climate change continues to move to the forefront of world-scale problems. The Intergovernmental Panel on Climate Change's (IPCC) latest report signalled the unequivocal warming of the climate system and highlighted the central role that human activity has played in driving this change (IPCC, 2013). Furthermore, the health of ecosystems is in decline as indicated by the Millennium Ecosystem Assessment, stating that humans have drastically altered the world's ecosystems, resulting in "substantial and largely irreversible loss in the diversity of life on Earth" (2005, 1).

Statements have been made that call for a reinvigorated effort to address matters of environmental sustainability in light of the new forms of analytics and insight that big data could generate. John Elkington, founder of the consulting firms SustainAbility and Volans, stated that big data and how we use numbers and data will play significant roles in sustainable outcomes. In turn, this will drive a shift from, "values as a driver of change to valuations" in what Elkington (2014) calls a 'Breakthrough Decade'.

How then is big data being applied toward the environmental sustainability effort? What transformative big data applications are as relevant to the environment as they have been to healthcare? What efficiencies are being gained through big data applications that are as beneficial to sustainability as they are to retailer profits? This paper highlights how big data is being applied in the realm of environmental sustainability across sectors and industries. By gathering insights from sustainability advocates (NGOs), service providers (consultancies), policy specialists (think tanks and intergovernmental organisations), compliers (corporates), and finally governments, I aim to provide a multi-perspective view of:

- How big data is being applied to the environmental sustainability effort;
- To what extent big data is impacting environmental sustainability while other sectors are experiencing revolutionising effects;
- Who the leaders and late or non-adopters are; and
- What some of the reasons are for late or non-adoption of big data tools and techniques.

1.4 Importance of this Study

As big data discussions, developments, and applications combined with calls for a heightened sense of urgency toward environmental sustainability efforts continue to emerge, the timing of this paper is important – if for no other reason – because it is late to arrive. Breadth on the topic is lacking while other fields are receiving extensive coverage.

Big data and sustainability have been combined and featured with depth, however. In the field of geography, researchers Sui, Elwood, and Goodchild (2013, 4) claimed that, "we can safely say geospatial data is becoming an important part of the big-data torrent." Similarly, Graham (2013, 255) stated that because big data produced can often be mapped, measured, and analysed, "big data possesses the potential to produce fundamentally new ways of knowing, enacting, and being in the world." Graham (2013) indicated that geographers have been wrestling with some of big data's central issues for some time, providing examples that date as far back as the 1950s.

Like Graham, Haklay (2008) focused on 'neogeography', described by Turner (2006, 3) as, "people using and creating their own maps, on their own terms and by combining elements of an existing toolset." The rise of websites from Google, Yahoo, and Microsoft, such as Google's Maps and Earth platforms, feature these novel methods of collecting and packaging information (Haklay, 2008).

Ultimately, by building on existing depth, I look to generate further discussions on big data applications to the environmental sustainability effort, adding breadth to the topic and elevating it closer to parity with other fields.

2 Methods

2.1 Classifications of Organisations Engaged for this Study

To explore big data's application to sustainability, I generated a sample of 14 organisations with environmental mandates and classified them into five categories:

- Environmental NGOs – These organisations play an environmental advocate role. They engage in fieldwork, promote and generate corporate environmental accountability, and make specific aspects of sustainability their core competency;
- Consultancies – These organisations advise on a range of functions and contain environmental sustainability as an area of focus, providing advice and support across sectors. Data management and analysis is typically a competency;
- Corporates – Corporates are required to comply with government environmental regulations. A corporate social responsibility (CSR) function often exists to factor in environmental impacts. Data analytics are critical to corporate performance;
- International Institutions – These organisations often specialise in policy research, working across sectors to promote their aims. Sustainability is a concentration embedded within their mandate, and data analysis varies as a competency; and
- Governments – Many governments have an open data initiative whereby information is made available for anyone to use for free and for any purpose (opendefinition.org). Open data is mixing with environmental topics.

2.2 Organisations Engaged for this Study and Interview Structure

Within these broad categories, I selected up to four organisations (listed in Table 1) to interview depending on a mixture of pragmatism and willingness to discuss the topic.

Table 1. *Organisations interviewed by category

Sustainability Advocates (NGOs)	Service Providers (Consultancies)	Compliers (Corporates)	Policy Specialists (Int'l institutions and think tanks)	Open Data Providers (Governments)
British Trust for Ornithology (BTO)	Accenture	BT Group	International Union for Conservation of Nature (IUCN)	**Government of the Netherlands
Conservation International (CI)	Anthesis Group	LinkedIn	World Bank Group	**The United States Government
World Wildlife Fund – UK (WWF – UK)	Volans		World Resources Institute (WRI)	
Zoological Society of London (ZSL)				

*Views provided under organisations are shorthand for the interviewees' personal views, rather than the organisations'.
**The Netherlands and United States governments are shorthand for the interviewees' departments or centres.

Organisations were selected for interview on the basis of the following:

- Environmental NGOs
    - The British Trust for Ornithology (BTO) is a charitable research institute providing information mainly on birds to inform the public, policy makers, and NGOs (bto.org). BTO's focus on data is appropriate for this study;
    - Conservation International (CI) is focused on conserving nature through improved coexistence between nature and people (conservation.org). CI was contacted based on my familiarity their big data initiatives;
    - The World Wildlife Fund – United Kingdom's (WWF-UK) goal is to achieve harmony between people and nature (wwf.org.uk). I participated in a workshop at their UK headquarters to brainstorm big data uses; and
    - The Zoological Society of London (ZSL) seeks to promote global conservation of animals and their habitats (zsl.org). ZSL was approached based on my familiarity with their innovative approach to conservation.

- Service Providers
    - Accenture is a global consultancy providing consulting, technology, and outsourcing services (Accenture.com). I contacted them due to their leadership in digital and information technology and sustainability;
    - Anthesis Group is a consulting firm focused on linking sustainability with business success (anthesisgroup.com). I contacted them based on their Software & Systems practice that explores big data solutions; and
    - Volans is a consultancy and think tank co-founded by John Elkington, who is a leading authority on sustainability. The organisation has coined 'The Break-Through Challenge', which factors in big data (Elkington, 2014).

- Corporates
    - BT Group (BT) is the largest provider of telecommunications networks and services in the UK (btplc.com). BT was approached based on my familiarity with their Better Future programme; and
    - LinkedIn operates the world's largest professional networking website with a user base of over 313 million members (linkedin.com). I contacted LinkedIn based on the their influence as a large internet company.

- International Institutions
  - The International Union for Conservation of Nature (IUCN) is the largest global environmental organisation (iucn.org). I contacted them based on their facilitation of scientific environmental efforts across sectors;
  - The World Resources Institute (WRI) is a global research organisation, identifying sustainability as "the foundation of economic opportunity and human well-being" (wri.org). Their Data Lab employs big data; and
  - The World Bank Group is focused on ending extreme poverty and boosting shared prosperity (worldbank.org). They, along with governments, are a leader in the movement towards open data.

- Governments
  - The Government of the Netherlands (Dutch government) and the United States (US) Government are active in promoting open data platforms. I contacted them based on open data's links to big data in the literature.

I determined whom to interview by asking the organisation for the relevant person and in some cases interviewed more than one person based on his or her expressed interest. Interviews took place over Skype or in-person. They were semi-structured with a template of questions used when necessary. Interviews were voice recorded, transcribed, and coded for themes.

3 Results

3.1 How the Organisations are Engaging with Big Data

To assess how the organisations are engaging with big data, I asked questions to identify the level of specificity the organisation has on the definition of big data, where big data skillsets are housed, and how its impact is viewed by the organisation.

3.1.1 Defining Big Data

A broad range of views on big data's definition was encountered throughout this study. One NGO's email reply to my interview request contained the question, "what is big data?" (anon, *personal communication*). At the other end of the spectrum, five of the 14 organisations provided a website definition of big data to varying degrees of specificity. Only the NGO category lacked a website definition.

A similar spectrum was found throughout the interviews. Just over half of the organisations indicated that clarity around the definition of big data is important. None of the participants had a refined definition of big data. Rather, a working interpretation was provided or the participant simply stated that he or she does not have a definition for – or interest in defining – big data.

WRI has directly engaged with big data, but they are not concerned with the general lack of agreement around big data's definition. Dan Hammer, who was Chief Data Scientist at WRI before moving on to work with the National Aeronautics and Space Association (NASA) and the White House on open data policy, suggested that there is no single definition of big data within WRI (*personal communication).* He described big data as "a placeholder term for datasets that are too big to hold in memory…" and explained that it is defined with as much fluidity within WRI as it is externally (ibid).

Tariq Khokhar, Data Scientist at the World Bank, was more specific on the definition of open data. Khokhar uses a widely accepted definition that clearly states the legal and technical parameters for data to be considered open (opendefinition.org). He stated that it is not uncommon for open data to be included in big data discussions. Certain big data sources are considered open if they conform to the open definition (*verbally*), but at present, similar, broadly agreed guidelines for big data are lacking.

From a government perspective, Mary Boatman, Environmental Studies Chief at the Bureau of Ocean Energy Management and manager of the Ocean Community at Ocean.Data.gov, added that focus is simply placed on data rather than the concept of big data (*verbally*). This perspective was echoed by most NGOs interviewed. Andy Musgrove, Head of Monitoring and Senior Research Manager at BTO, highlighted that BTO has been collecting data since the 1930s and has accumulated a lot of it, but the organisation identifies with data as opposed to big data (*verbally*). Robin Freeman, Head of Indicators and Assessments at ZSL, added that remotely sensed data for forests and climate science generated some of the earliest big datasets, but the term big data has not been utilised to describe them (*verbally*). Despite this, remotely sensed data is included as a factor in the enabling of big data initiatives (Sui, 2013).

The most detail applied to big data's definition came from the consultancy category. Andrew Armstrong, Managing Partner at Anthesis Group, drew from the International Institute for Analytics' (IIA) 'Analytics 3.0 Framework' to generate a definition and an understanding of how big data becomes impactful (*verbally*). The IIA placed big data as a second-generation development, following traditional analytics and preceding Analytics 3.0 (2013, 16). Analytics 3.0 combines traditional analytics and big data to yield "insights and offerings with speed and impact" (ibid). Armstrong uses this model to place big data in an evolutionary framework, separate it from traditional analytics, and link big data to its potential to generate impact (*verbally*).

In summary, alignment is lacking between these organisations with sustainability mandates on big data's definition, but as noted, this is consistent with the literature. Organisations like Anthesis however, seek to better understand big data's unique characteristics. I'll assess why this is important in the discussion section of this study.

3.1.2 Big Data Skillsets

To further examine how these organisations are engaging with big data, I enquired as to the availability of skillsets for big data's mobilisation. In 2010, Harvard Business Review claimed that the data scientist is the "sexiest job of the 21st century" (Simon, 2013, 16). The data scientist, as Simon noted, has yet to find a clear description, aligning well with the fluid definition of big data. Simon motioned to IBM's description, which states that the data scientist combines business acumen with the skills to communicate with business and IT professionals not only to address business problems, but also to pick the right problems to address (ibid).

Only five of the organisations employ data scientists in name, highlighting the potential absence of the skills needed to manipulate big datasets. None of the NGOs interviewed employ a data scientist in name. Several participants indicated however, that their organisation contains a variety of roles that touch upon elements of the data scientist's description, such as statisticians and programmers (R. Freeman, T. Cambridge, A. Musgrove, E. Fegraus, *verbally*).

BT Group and LinkedIn employ data scientists, as do WRI and the World Bank. The IUCN is the third organisation in this latter category and does not employ data scientists in name. Of the consultancies, Accenture is the lone organisation that employs data scientists, but Anthesis works with data specialists external to the organisation in addition to employing data analysts that function as data scientists (A. Armstrong, *verbally*). The housing of these skillsets signals that the organisation is capable of directly engaging with big data.

Table 2 highlights several indicators of big data engagement discussed in this paper. While the indicators are not exhaustive, they provide a picture of how each organisation and category engages with big data in addition to the prevalence of the indicators themselves. For example, the employment of data scientists in name is visibly rare. Figure 1 summarises how each category scores as per the indicators presented in Table 2.

Table 2. Organisations' level of engagement with big data

Category	Organisation	Big Data Appears on Website?	Big Data is Defined on Website?	*Definition of Big Data has Significance?	Data Scientist is a Position Employed?
NGO	BTO	NO	NO	NO	NO
NGO	CI	YES	NO	YES	NO
NGO	WWF-UK	NO	NO	NO	NO
NGO	ZSL	NO	NO	NO	NO
Consultancy	Accenture	YES	YES	YES	YES
Consultancy	Anthesis	YES	NO	YES	NO
Consultancy	Volans	YES	NO	YES	NO
Corporate	BT Group	YES	YES	YES	YES
Corporate	LinkedIn	YES	YES	NO	YES
Int'l Institution	IUCN	NO	NO	YES	NO
Int'l Institution	World Bank	YES	YES	NO	YES
Int'l Institution	WRI	YES	NO	NO	YES
Government	Dutch Government	YES	NO	YES	*NO
Government	US Government	YES	YES	YES	*NO

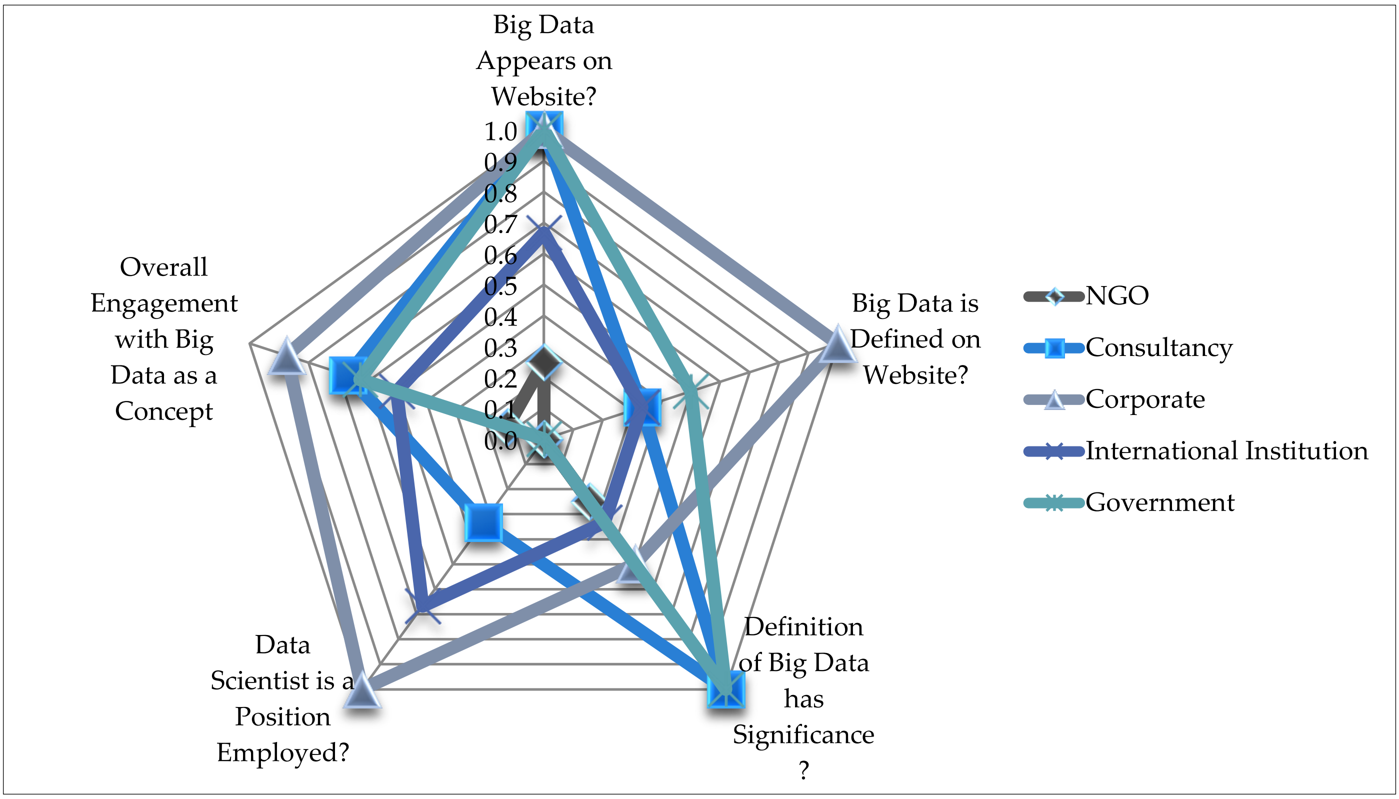
*'Definition of Big Data has Significance' has a 'yes' recorded if participants identified at a minimum certain components of the definition as important.
**The employment of the data scientist in name was limited to the centres represented within governments.

3.1.3 Big Data's Impact on Environmental Sustainability Efforts

Greater consistency was found on perspectives of big data's impact on environmental sustainability efforts. The exceptions largely align with the ambiguous definition and meaning of big data. The results of the big data engagement scoring pre-empt how different categories and the organisations themselves are adopting big data as an organisational tool or approach. The category on the left of Figure 1, 'Overall Engagement with Big Data as a Concept', indicates the extent to which the overall category is engaging with big data based on its scoring across the factors. On a scale of zero (no engagement) to one (full engagement), the corporate category has the best score at 0.88, followed by consultancies at 0.67, governments at 0.63, international institutions at 0.50, and lastly NGOs with a score of 0.13.

LinkedIn and BT, who form the highest scoring category, are making big data a core element of their environmental sustainability efforts. BT has established research programmes focused on big data. Eric Anderson, Senior Consultant in Group Strategy within BT, manages the Better Future Forum, which is an annual stakeholder forum to generate discussions and gather insights on topics that BT feels require more attention (E. Anderson, *verbally*). This year's forum focused on how to use big data for social good. LinkedIn also believes that big data will have an impact on the organisation's environmental sustainability efforts (K. Shea, *verbally*). Kelly Shea, Global Critical Sustainability Engineer at LinkedIn, was hired as a result of LinkedIn's recognition that big data could be applied to the energy efficiency of its data centres (ibid).

Figure 1. Each category's scoring on big data indicators listed in Table 2 and overall score



Scoring was calculated by counting the ‘yes’ answers for an indicator for all organisations within a category and dividing that number by the total number of organisations. An overall engagement score was calculated by adding the ‘yes’ answers across a category and dividing the sum by the total number of opportunities for a “yes” response. Each criterion is weighted evenly.

The consultancies, which as a category scored second highest on basic big data engagement indicators, have largely integrated big data into their practices as a competency, led by Accenture and Anthesis. Armstrong has found that the supply of big data initiatives is met with demand from customers, highlighting its relevance to their sustainability efforts (*verbally*). Volans is active in asking questions about the location of data and what it is saying (A. Feldman, *verbally*). Amanda Feldman, Director of Impact and Innovation, indicated that Volans is engaging with the topic in order to facilitate their coined, ‘Breakthrough Decade’.

Both the Dutch and US governments stated that they recognise the utility of big data applications toward environmental sustainability efforts (J. Holm, M. Boatman, T. Nijs, *verbally*). The US government’s open data effort includes an oceans focus, a climate change focus, and a new ecosystems focus (J. Holm, *verbally*). Ton de Nijs, Coordinator of Living Environment within the Centre for Sustainability of the Dutch government, acknowledged that his division is only beginning to explore the full potential of big data’s utility within the realm of environmental sustainability, but its open data effort is already proving to be effective, especially in the environmental planning space (*verbally*).

The World Bank and WRI largely uphold the international institution category, placing fourth. Martin Sneary, Programme Director for the Integrated Biodiversity Assessment Tool (IBAT) at IUCN, stated that a lot of interesting activity is happening around tool development to better collect and analyse data, but also stressed that traditional biodiversity data, which undergoes strict assessment and documentation, does not enter the realm of big data (*verbally*). WRI employs data scientists and software engineers that form a team called The

Data Lab. It was specifically built to work with big data (D. Hammer, *verbally*). As a result, WRI is poised to benefit from big data applications.

NGOs expressed less optimism on the impact of big data, placing the category far back of the others. CI was the only organisation to register any 'yes' responses. Eric Fegraus, Director of Information Systems, identified CI as a unique environmental NGO from the standpoint that it is a primary data collector, meaning that the organisation's teams go into the field and collect their own data, whereas many NGOs will rely on data that other organisations collect (*verbally*). The CI-HP partnership will be covered in greater detail in the next section, but it is worth noting here how Fegraus described the impact of CI's engagement with big data:

> We're basically building production level software for conservation data.... To produce fully robust systems that can grow as your data grow and build upon as your requirements change, to me that is the biggest win of it all. And basically it makes it sustainable because the traditional use of data in conservation science is…nice for papers but it doesn't work for a programme that wants to continually run (ibid).

ZSL and BTO also highlighted technology improvements that are helping them to collect data in new ways (R. Freeman, A Musgrove, *verbally*). They expressed doubt however, as to whether big data as a concept was having a significant impact because projects that might be thought of as big data initiatives are not referred to as such. Fegraus, on the other hand, stated that while the volume of data CI collects may not be considered big data in some sectors, it could be considered big to the field of biodiversity conservation and especially big to some of the remote regions in which CI is working (*verbally*).

WWF-UK acknowledged the impact that big data is having toward the organisation's contributions to the international WWF network's datasets as well as external to the organisation's efforts, but to a lesser extent on its own operations (N. Gunn, T. Cambridge, R. Perkins, *verbally*). Big data was described as "off the map" (R. Perkins, *verbally*). However, the NGO was considering a hackathon aimed at determining what could be done with the organisation's data. Neil Gunn, Websites Project Manager at WWF-UK, hoped that this event would be a catalyst for the organisation to use big data more often (*verbally*).

Having presented how these organisations are defining big data, the big data skillsets they house, and finally how they view the potential or realised impact of big data, I will now turn to the big data applications in detail that some of the organisations are engaged in.

3.2 Applications of Big Data towards Environmental Sustainability Efforts

First, I will detail big data applications, as labelled by the organisations, from one organisation in each category. I will then present an additional example that an organisation described as innovative or novel, but not as a big data initiative. This serves to emphasise differing perceptions as to what constitutes a big data application.

3.2.1 Private Sector – NGO Partnership: HP Earth Insights and CI's TEAM Network

CI's Tropical Ecology Assessment Monitoring (TEAM) Network serves as an "early warning system for nature" (teamnetwork.org), focusing on biodiversity monitoring in tropical forests (E. Fegraus, *verbally*). TEAM operates in 16 tropical forests across Asia, Africa, and Latin America, deploying people to collect data in a standard format that is publicly shared and updated in real time (teamnetwork.org).

In the earlier stages of the TEAM Network, Fegraus had built the supporting IT-infrastructure by himself, then with partners, and eventually this evolved into a partnership with HP (*verbally*). HP Earth Insights was formed to apply a big data solution to biodiversity loss by combining CI's data collection efforts with an HP software tool that allows scientists to "quickly analyse large and fast-growing volumes of data" through a platform that manages "the full spectrum of structured, semi-structured, and unstructured data" (conservation.org).

HP reports that scientists can analyse data 87 per cent faster, reducing months or weeks-long analyses by scientists to hours by a single person (hp.com). Moreover, HP built a project dashboard and analytics tool called the Wildlife Picture Index (WPI) Analytics system that allows for the visualisation of "near real-time data-driven insights" (ibid). The system leverages the shift from manually driven data collection to data generated by camera traps and climate sensors (ibid). The initiative is generating "findings about the environment that were previously unknown", such as species population declines in comparison to baseline levels (conservation.org).

3.2.2 Private Sector – International Institution Partnership: Global Forest Watch

WRI's Global Forest Watch (GFW) merges the latest technology with on the ground partnerships to enhance forest information (globalforestwatch.org). Now a Senior Fellow at WRI, David Wheeler piloted a programme with the World Bank's Development Research Group on public disclosure of pollution by collaborating with local Environmental Protection Agencies (EPA) in several countries (D. Wheeler, *verbally*). Wheeler indicated that over time data sources have become richer, allowing for greater specificity and locality of pollution identification (*verbally*).

Wheeler and Hammer began to think about how deforestation could be identified in a similar fashion to pollution (D. Wheeler, verbally). Lacking information at the time as to where deforestation was happening, Hammer, who was just out of college with a mathematics degree, began to build a dataset under Wheeler's leadership before they formed The Data Lab at WRI (ibid). Hammer and his recently graduated teammates provide an example of the skillsets that can generate large-scale environmental outcomes through big data mobilisation. The team eventually partnered with Google to reduce costs, implement algorithms, and add cloud technology (wri.org).

GFW also brings together satellite technology and crowdsourcing to generate a mapping application that is used by non-profit organisations, governments and other organisations interested in these monitoring efforts (D. Hammer, *verbally*). For example, the Jane Goodall Institute uses GFW to monitor chimp habitats, park rangers use it to monitor the boundaries of protected areas, and indigenous groups use GFW to monitor borders (ibid). Stakeholders are also able to set up email alerts for specific countries when forest lost occurs.

Wheeler explained that the process is now consistent: an environmental problem effecting surrounding communities is identified, an effort is made to use publicly available data at the highest resolution, and the data is transitioned into digestible formats such as maps and rating systems (*verbally*). These formats are shared publicly to empower people to zoom in on an area, see where the problem is and who is associated, and then take action (ibid). The result is an application that rivals that of Facebook's ability to identify people's faces in a photo. The same technology can be used to identify resource use and extraction that is geographically dispersed (ibid).

3.2.3 The United States Government's Open Data Initiative

The idea that open data could provide transparency into US government operations was generated under a new Chief Information Officer (CIO) in 2009, and in that same year Data.gov opened (J. Holm, *verbally*). Data.gov is the central site for US government data that provides "Federal, state and local data, tools, and resources to conduct research, build apps, design data visualizations, and more" (data.gov).

Jeanne Holm, Chief Knowledge Architect at the National Aeronautics and Space Administration (NASA) and Data Evangelist at Data.gov, explained that data.gov collates and harvests metadata, defined simply as data that describes data (Sweet, 2013), from 175 different Federal agencies. 100 non-Federal or non-governmental agencies, including states and cities contribute (J. Holm, *verbally*), as do 44 international countries and 163 international regions (data.gov). By combining these sources, the project team has catalogued inventories of data and combined them with technical guidelines and manuals (ibid). People can use this data however they would like, and the government only asks that this be conducted transparently (ibid).

Boatman was put on detail, bound to the Executive Office of the President, following the signing of the National Ocean Policy in 2010 (M. Boatman, *verbally*). She highlighted several ways in which these massive open datasets are applied, including:

- To better understand how offshore wind development is impacting birds;
- To factor in endangered species' locations for ocean-based work performed by the energy industry and military; and
- To map the seafloor for areas of biological sensitivity (ibid).

New methods of ocean-based information collection have led to the creation of apps. For example, to protect endangered North Atlantic right whales off the coast of Boston from getting hit by boats en route into harbour, an app has been developed that connects to passive acoustic monitoring signals that warn boats of the whales' location (ibid).

The volume of data, the environmentally focused communities, and the unique applications being generated through the US government's open data initiative are growing and expanding. Holm stated that there are at present roughly 500,000 data resources, which are grouped and classified so they can be easily found and accessed by users (*verbally*). In addition to the oceans focus, the climate and ecosystems communities opened in 2014.

3.2.4 A Corporate Sustainability Effort: BT's Better Future Programme

BT's Better Future programme endeavours to place sustainability as an important focus within the organisation's strategy. One of the programme's goals is to help customers reduce carbon emissions by at least three times the carbon impact of BT's business (btplc.com). BT states that it must also demonstrate internal efforts to reduce its carbon footprint if the organisation expects others to partner with it, especially considering that BT is one of the top energy users in the UK (ibid).

Anderson indicated that BT applies big data to further enhance its environmental sustainability efforts through smart energy management (*verbally*). BT implemented some of the common practices to saving energy such as turning off the lights in rooms that are not being used, refurbishing some of the fans in its data centres, and using fresh air cooling in its data centres (ibid). Anderson stressed however that BT needed to better understand the energy profiles of its buildings (ibid).

BT has achieved this by adding more than 50,000 smart meters and implementing a smart energy management system, delivering annualised savings of nearly five million pounds (ibid). Each of the smart meters generates a vast number of data points, and the system automates energy use. Typically, implementing smart metering systems presents a shift from one energy reading per month to one every 15 minutes (ibm.com). For BT, that translates to more than 4.8 million reads per day. This scaling up of data points to manage BT's energy usage is one example of how Anderson sees BT utilising a big data solution that reduces its environmental impact (*verbally*).

3.2.5 Building Big Data into Sustainability Advisory Services

Anthesis deploys big data solutions in its work with clients. Brad Blundell, Director at Anthesis, stated that there is an emerging area of work for the firm that is reliant on big data, and it is in demand by venture capital and private equity groups (*verbally*). This source of demand provides unique insight into some of the latest thinking around approaches to environmental sustainability. Investors are increasingly asked questions about the future issues that some of their investments are likely to face, and they are concerned about environmental and legislative impacts, including water and resource scarcity and trade blocks (*verbally*).

Anthesis was asked to create a tool that looks at critical issues grouped around recognised standards to identify circumstantial impacts and opportunities that a business might be exposed to (ibid). The tool will evaluate either the potential cost of mitigating an issue or benefit of realising an opportunity. A beverage manufacturer operating in an area where water scarcity is a concern, or a company developing a technology that allows for the substitution of water for concepts such as waterless washing machines exemplifies the risks and opportunities the tool will quantify (ibid).

The factors to be weighed are largely outside of the organisations' and investors' control, such as how climate change impacts food security, social unrest, or resource nationalisation. These factors also often overlap, which needs to be accounted for in the model (ibid). The overlap lends relevance to the definition of big data brought forward earlier by Armstrong, who views big data as a combination of internal and external data sources that begin to add value when correlations are drawn (*verbally*). For example, external information, such as the Organisation for Economic Co-Operation and Development's (OECD) output table data, could be overlaid with internal data, which could then be added to forecasts to see how costs will change over intervals of time (B. Blundell, *verbally*).

Anthesis hopes to access some of this data through open sources, such as flood risk resources in the UK (ibid). While the firm will be collecting massive amounts of data for this work, the important 'V' is variety, which Blundell substantiated by highlighting the focus needed on the most important aspects of the situation in order to find the right mix of information for the model (ibid).

3.2.6 One Additional Example: ZSL's Innovative Data Collection and Analyses

I will present some initiatives discussed by ZSL in order to lend insight into how the fluidity of the definition of big data and its relative scale across sectors and geographies are interpreted. While the organisation does not identify with big data on these specific initiatives, others might consider it to be within the big data realm.

ZSL manages large datasets. One of these datasets is the Living Planet Index, which includes population time series for roughly 3000 species and 11,000 populations and contains approximately 30 to 40 years of data for each (R. Freeman, *verbally*). Moreover, this data encompasses metadata about that species, such as threat status and whether or not it's a migrant species (ibid). ZSL analyses this data and publishes the Living Planet Report in partnership with WWF-UK as a science-based report on the health of the planet and impact of human activity (wwf.panda.org).

The greatest benefit derived from this data however, is yielded when it is paired with other datasets, namely habitat change data and the IUCN Red List (R. Freeman, *verbally*). The IUCN reviews the data provided by ZSL to determine which category of risk that a species falls under (ibid).

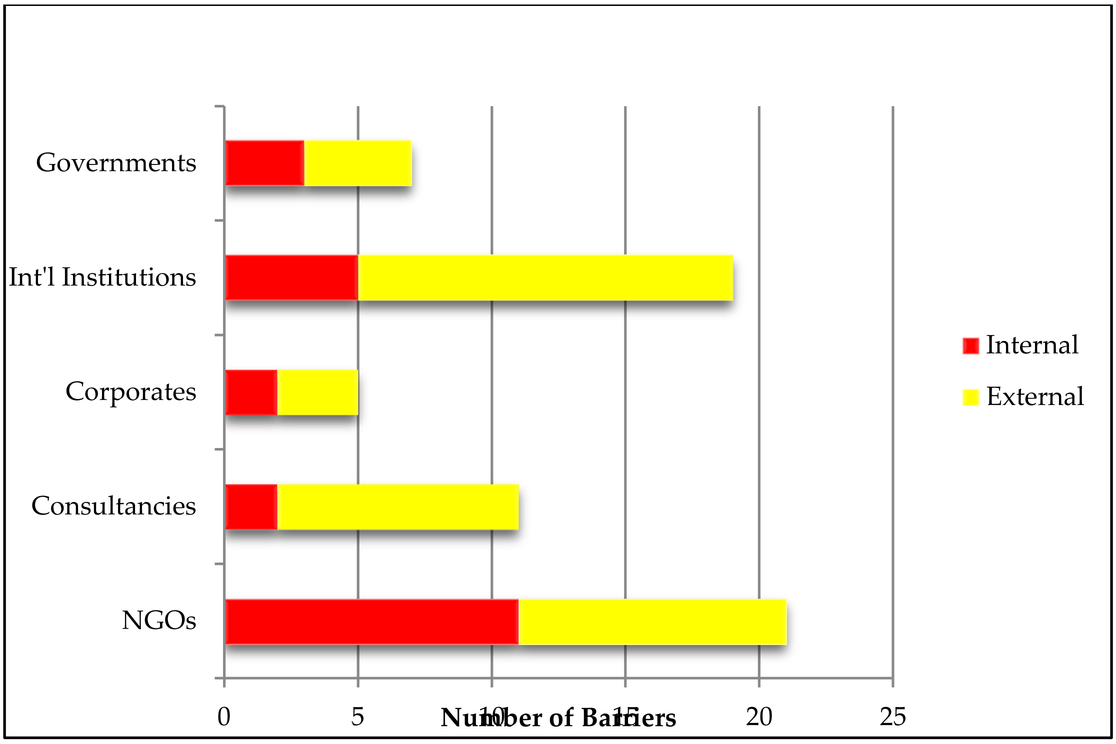
Another example is derived from Freeman's interest in movement ecology, which is the gathering of spatial information on individual organisms (ibid). He applies novel means of data collection, such as the placing of a GPS tracker on birds to record their locations and migration patterns over time. This data is also tied to other datasets, specifically predator activity and temperature (ibid). The shift in technology to more affordable and lighter devices is enabling and changing the understanding of the behaviour and ecology of animals (ibid) through the provision of new ways to collect information, regardless of whether or not the information is labelled as big data.

3.3 Barriers to and Opportunities for Big Data and Sustainability

3.3.1 Barriers to Big Data and Environmental Sustainability Efforts

The organisations were asked to identify barriers to big data adoption. Figure 2 provides a summary of how many internal (experienced within the organisation) and external (observed elsewhere) barriers were identified by each category. Figure 3 provides a heat map of which organisations discussed different barriers.

Figure 2. Number of internal versus external barriers identified by category

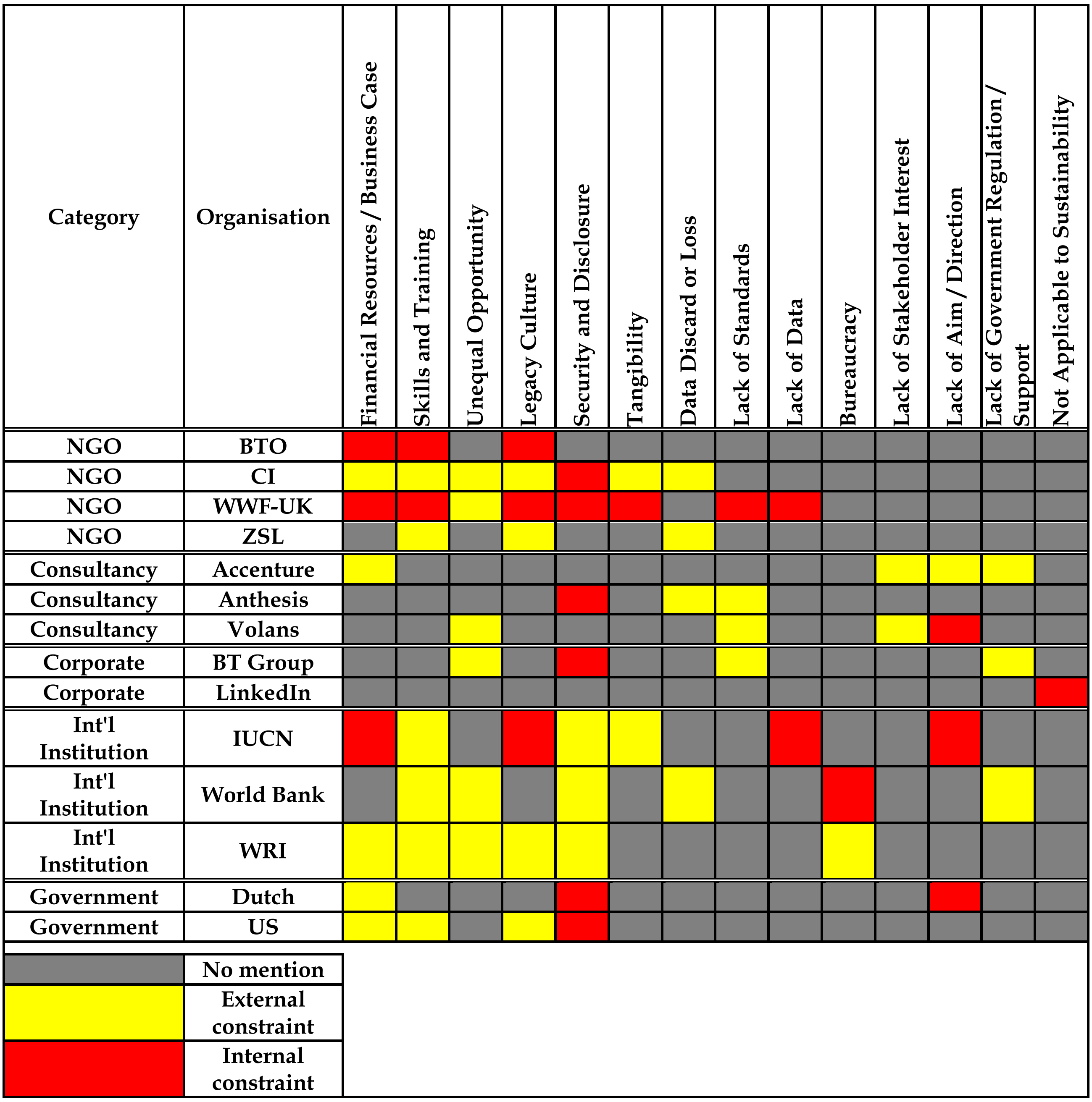


Only the NGO category identified more internal than external barriers, indicating that the category experiences or perceives significant internal challenges to big data adoption. This might also align with the high total number of barriers identified, but an additional consideration is that more interviews were conducted with the NGO category than with the others.

International institutions followed with a total of 19 barriers mentioned. The comparatively low percentage of internal barriers identified might suggest that international institutions are well positioned to observe challenges and risks to big data uptake given their cross-sector involvement, but experience fewer themselves. The exception within this category is the IUCN, who is the most conservation focused of all three organisations and works closest with NGOs. Four internal barriers were mentioned by the organisation, which is second only to WWF-UK.

Consultancies identified the lowest percentage of internal barriers, reflecting their outward looking perspective through the provision of expertise to clients. The corporate and government categories generated internal barriers that accounted for roughly 40 per cent of the total within their categories.

Figure 3. A Heat Map of barriers to big data adoption by organisation



Barriers are allocated to different organisations based on verbal feedback during interviews. Grey squares do not mean that the organisation disagrees with the barrier, but rather that it was not mentioned during the interview.

I will detail the barriers to lend context and insight into how they are relevant to each category and organisation, and, when paired with the findings presented in the figures, to provide greater clarity on why big data has (or has not) been adopted.

Barrier #1 – Financial Constraints and Lacking a Business Case for the Environment

Financial constraints were identified as a barrier. For a charity like BTO, finding income streams to analyse the amount of data the organisation has is not easy (A. Musgrove, *verbally*). WWF-UK also stated that limited financial resources pose a constraint to big data uptake (N. Gunn, *verbally*). For the IUCN, securing renewable income to support the on-going costs associated with data management, such as software development and maintenance fees, is a significant challenge (M. Sneary, *verbally*). Beyond costs, a business case for big data's use toward environmental efforts is an important consideration. Nijs of the Dutch government suggested that big data has launched in other sectors because it is linked to profits rather than the cost of sustaining the environment (*verbally*). Boatman added that there is no commerce in environmental data, lacking alignment with the capitalist model (*verbally*).

Barrier #2 – Lack of Skills and Training

Housing the necessary skills to employ big data was identified as a barrier by several organisations. This barrier again draws attention to the absence of the data scientist position in most organisations interviewed. NGOs see a skillset shortage either within their organisations or amongst NGOs in general (E. Fegraus, T. Cambridge, A. Musgrove, *verbally*). Hammer and Holm stated that the needed skills are often consumed by other sectors, such as retail, because data science skillsets are highly sought after and have therefore become expensive for organisations to employ (*verbally*). As a result, data analysis training was described as the most important element to launching big data initiatives going forward (R. Freeman, *verbally*).

Barrier #3 – Lack of Equality

Big data accessibility was listed as a barrier. WWF-UK raised the issues of accessibility and equality in the context of fisheries, stating that a mandate for new data standards could place developing fisheries in places like Africa and the Philippines at a disadvantage because they cannot afford the necessary upgrades (T. Cambridge, *verbally*). CI added that developing countries do not have the means to invest in data creation enablers such as weather and forest services, placing them at an information disadvantage (E. Fegraus, *verbally*). Wheeler noted that as income rises in developing countries so does the Willingness To Pay (WTP) for environmental amenities, but citizens might not have the information to know what is affecting their circumstances in order to take action regardless of income (*verbally*).

Barrier #4 – The Nature of Conservation and Cultural Legacies

Several organisations spoke of the nature of conservation as a hindrance to big data innovation. Gunn suggested that the methods that work for conservation often involve fieldwork and might not be effective for instilling climate change messages, which big data is more often applied to and vice versa (*verbally*). One interviewee noted that the conservation community often has an adversarial or combative outlook toward the corporate world, preventing partnerships that could leverage data driven conservation efforts (*anonymous*, *verbally*). Additionally, the environmental science community has an established way of doing its work, and perhaps it has settled into a certain comfort level (D. Wheeler, *verbally*). Big data initiatives might therefore be resisted in this space because they could change traditional methods (ibid).

Barrier #5 – Confidentiality, Security, and Disclosure

Several unique aspects of confidentiality and security were identified as barriers. For example, Nijs remarked that the Dutch government is careful with its maps of biodiversity (*verbally*). Maps are limited to a one by one kilometre grid size to protect endangered species (ibid). Hammer indicated that governments resisted GFW, especially in its early development (*verbally*). A government official in Indonesia questioned what right WRI had to utilize big data to look into the country (*verbally*). Wheeler added that, "small data threatens a few people, and big data threatens everybody" to describe the pushback from people whose "more narrowly defined interests are threatened" (*verbally*). Furthermore, Holm, in speaking for the open data movement, stated that concerns are raised by corporates who view data as proprietary and fear that its release poses risks to businesses' brands and profits depending on who is analysing the data and for what purpose (*verbally*). Boatman added that scientists and researchers often claim that their data is intellectual property (IP) and opening that data poses a potential breach to their IP (*verbally*).

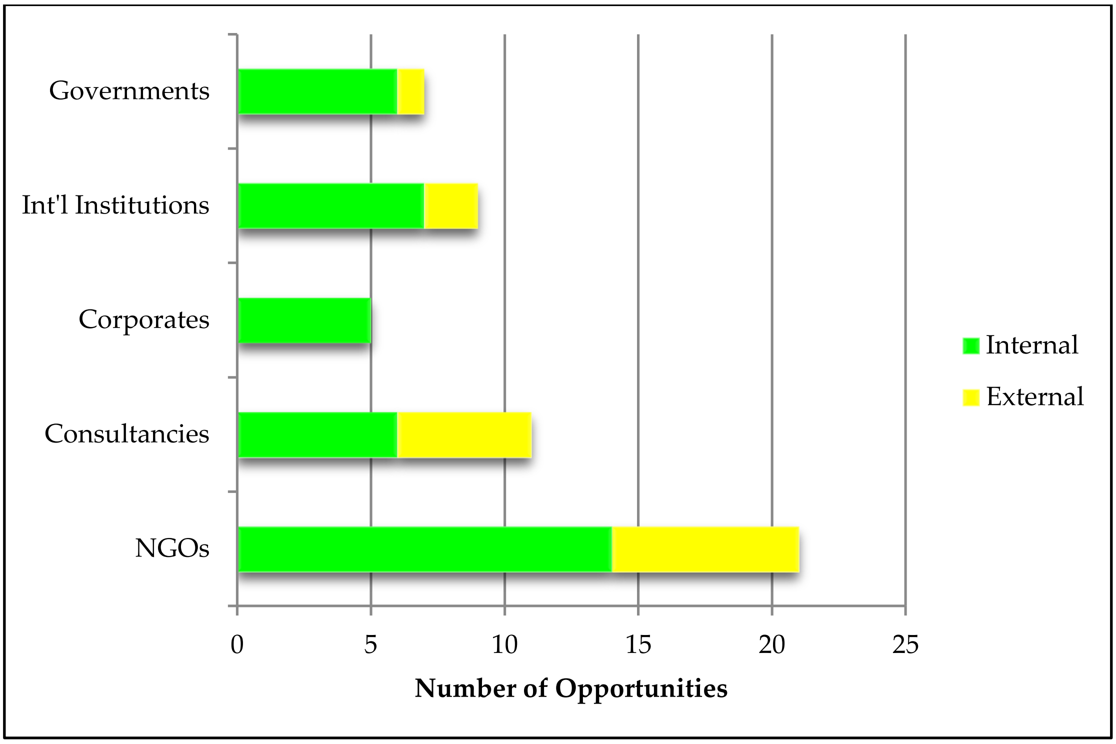
Additional Barriers to Big Data Utilisation for Environmental Sustainability Efforts

- **Tangibility** – Donors often require tangible products from donations, such as books or web sites, which while important to organisations, can leave a significant funding gap for on-going costs and investments that do not necessarily produce a tangible outcome (M. Sneary, *verbally*);
- **Data reusability** – Dataset management becomes critical to ensure that data is not lost or discarded but kept for potential future analysis. This has been a challenge for NGOs (E. Fegraus, R. Freeman, *verbally*). From an external perspective, Khokhar stated that data used by one party should be shared to generate solutions elsewhere (*verbally*);
- **Standards** – The lack or absence of standards for big data was raised in relation to issues around data uniformity, harmonization, and centralization on a global scale (T. Cambridge, *verbally*). Both Blundell and Armstrong of Anthesis also discussed the challenges associated with the inconsistency of data types across businesses, sectors, and countries (*verbally*);
- **Stakeholder interest** – Accenture's Rob Hayward works with the United Nations Global Compact (UNGC) on their triennial CEO Study on Sustainability. In the most recent study, Accenture interviewed and surveyed more than 1000 Chief Executive Officers (CEOs) across 103 countries and 27 industries to learn their views on a sustainable economy (Accenture.com). The study revealed that companies have become frustrated by a lack of reciprocity from consumers (R. Hayward, *verbally*). Broadly, the CEOs indicated that they have placed a high degree of emphasis on data collection and analysis to communicate sustainability performance to consumers (ibid). A companion study by Accenture and Havas Re:Purpose that surveyed 30,000 consumers in 20 countries revealed however, that only a third of consumers are regularly factoring in sustainability to their purchasing decisions (Accenture.com). Hayward suggested that this disconnect between corporates and consumers could plateau corporate big data developments toward sustainability efforts due to lowered incentive to invest in them (*verbally*);
- **Government engagement** – Hayward highlighted that businesses have expressed the need for government support on sustainability efforts to ensure a level playing field with businesses that are not investing in sustainability (ibid). BT's Better Future Forum identified that regulatory standards also need to better align big data privacy concerns with a level of acceptable access to personal information when it benefits social good (E. Anderson, *verbally*); and
- **Applicability** – Shea of LinkedIn, while gaining approval to be interviewed for this study, was challenged by a colleague, who suggested that big data is not relevant to sustainability (ibid). This raises the possibility that sustainability has been slow to coalesce with big data due to perceptions that the two fields lack applicability.

3.3.2 Opportunities for Big Data and Environmental Sustainability

Interviewees were also asked to provide insights into opportunities for big data uptake for environmental sustainability efforts. Figure 4 provides a summary of how many internal and external opportunities were identified by each category. Figure 5 provides a heat map of which organisations discussed different opportunities.

Figure 4. Number of internal versus external opportunities by category



NGOs mentioned the most opportunities with 21, which again could be explained by the fact that more interviews were conducted with them. It could also be attributed to the gap previously described between NGOs and other categories on the big data engagement scoring. The total of opportunities mentioned by NGOs is nearly double the next highest total at 11, which was generated by the consultancies. All categories identified more opportunities as internally relevant than externally relevant. This also holds true for all organisations except one, CI, who has progressed much further on big data adoption than the other NGOs interviewed.

I will now discuss in detail the opportunities raised by participants throughout the interviews. Greater detail on opportunities, when paired with the figures presented, provides an overall picture of how big data has been or could be adopted.

Figure 5. A Heat Map of big data adoption opportunities by organisation

Category	Organisation	Partnerships	Demonstrating Powerful Examples	Personalising the Environment	Emerging and Accessible Technology	Emerging Sources of Funding	Big Data in its Early Days	Merging of Sustainability and Strategy	Data as an Asset or Metric	Emerging Leaders and Innovators	Increased Transparency and Accountability	Citizen Science	Forums, Workshops, Hackathons
NGO	BTO	No mention	Internal Opportunity	No mention	No mention	No mention	No mention	No mention	No mention	No mention	No mention	No mention	No mention
NGO	CI	Internal Opportunity	External Opportunity	Internal Opportunity	No mention	External Opportunity	No mention	No mention	No mention	No mention	External Opportunity	No mention	No mention
NGO	WWF-UK	Internal Opportunity	Internal Opportunity	External Opportunity	External Opportunity	Internal Opportunity	No mention	Internal Opportunity	No mention	External Opportunity	Internal Opportunity	No mention	Internal Opportunity
NGO	ZSL	Internal Opportunity	No mention	Internal Opportunity	External Opportunity	Internal Opportunity	No mention	No mention	Internal Opportunity	No mention	No mention	No mention	Internal Opportunity
Consultancy	Accenture	External Opportunity	No mention	Internal Opportunity	No mention	No mention	No mention	Internal Opportunity	No mention	No mention	No mention	No mention	No mention
Consultancy	Anthesis	No mention	No mention	No mention	Internal Opportunity	No mention	No mention	No mention	No mention	No mention	External Opportunity	No mention	No mention
Consultancy	Volans	External Opportunity	External Opportunity	No mention	No mention	No mention	No mention	Internal Opportunity	Internal Opportunity	External Opportunity	Internal Opportunity	No mention	No mention
Corporate	BT Group	No mention	Internal Opportunity	Internal Opportunity	No mention	No mention	Internal Opportunity	No mention	No mention	No mention	No mention	No mention	No mention
Corporate	LinkedIn	No mention	Internal Opportunity	Internal Opportunity	No mention	No mention	No mention	No mention	No mention	No mention	No mention	No mention	No mention
Int'l Institution	IUCN	Internal Opportunity	Internal Opportunity	No mention	No mention	No mention	No mention	No mention	No mention	No mention	No mention	Internal Opportunity	No mention
Int'l Institution	World Bank	Internal Opportunity	No mention	No mention	No mention	No mention	No mention	No mention	No mention	No mention	No mention	No mention	No mention
Int'l Institution	WRI	Internal Opportunity	Internal Opportunity	No mention	External Opportunity	No mention	No mention	No mention	No mention	External Opportunity	No mention	No mention	Internal Opportunity
Government	Dutch	Internal Opportunity	No mention	No mention	No mention	No mention	Internal Opportunity	No mention	No mention	No mention	No mention	No mention	No mention
Government	US	Internal Opportunity	No mention	External Opportunity	Internal Opportunity	No mention	No mention	No mention	No mention	Internal Opportunity	No mention	No mention	Internal Opportunity

(grey)	No mention
(yellow)	External Opportunity
(green)	Internal Opportunity

Opportunities are allocated to different organisations based on verbal feedback during interviews. Grey squares do not indicate that the organisation disagrees with the opportunity, but rather that it did not get mentioned during the interview.

Opportunity #1 – Partnerships

Partnerships provide an opportunity to leverage big data use toward sustainability initiatives. The CI partnership with HP offered a win-win scenario, whereby CI leverages HP technology toward its conservation efforts, and HP receives both brand and product exposure (E. Fegraus, *verbally*). Volans added a unique perspective to partnerships. They worked with HP in 2009 to design its social innovation strategy. A component of it is an environmental focus, and that includes HP's partnership with CI (A. Feldman, *verbally*). The US government is also partnering across the world on open data access to increase the volume of data open to the public (J. Holm, *verbally*).

Opportunity #2 – Demonstrating Powerful Examples and Healthy Competition

Several organisations discussed the importance of learning from others and creating a healthy competitive atmosphere. For example, BTO could be learning from big data applications occurring externally (A. Musgrove, *verbally*). Additionally, the power of interactivity through engagement with others in the field, as well as pressure to keep up with the field is internalised at WRI (D. Wheeler, *verbally*).

Opportunity #3 – Personalising the Environment

Big data's applicability to human health is creating opportunities to generate greater interest in environmental sustainability. Gunn of WWF-UK built upon this concept by discussing an example from fisheries. Concerns about food poisoning led to the need to prove where the fish had come from to show where in the supply chain that problems arose (*verbally*). This had nothing to do with conservation but rather with health. However, this data now facilitates Marine Stewardship Council (MSC) certification, enabling the traceability of fish back to the crate or boat and making it a conservation tool (ibid). Personalizing the environment by linking it to human health assists with the movement of big data applicability to environmental sustainability. Devices, such as asthma inhalers, now measure environmental risk factors such as air quality, warning people when and where they might be at risk in addition to analysing that data to predict potential future issues (J. Holm, *verbally*).

Opportunity #4 – Emerging and Accessible Technology

Emerging and increasingly accessible technologies are creating opportunities for big data adoption. Environmental science technology is changing so that tools are becoming more accessible to the averagely trained ecologist, who can employ the tools rather than rely on computer scientists (R. Freeman, *verbally*). These tools and technologies are also becoming cheaper over time, increasing access and adoption (T. Cambridge, *verbally*). Wheeler stated that less reliance will be placed on local monitoring as a result of satellite based systems that analyse air quality and perhaps also water quality using similar techniques, and these will generate database volumes that grow quickly from terabytes to petabytes (*verbally*). From a government open data perspective, technologies have emerged to make big datasets available to the public, who would not have the access they do now ten years ago (M. Boatman, *verbally*).

Opportunity #5 – Emerging Sources of Funding for Innovation

Three NGOs stated that new sources of funding are opening for big data use. The rise of big data is drawing more sophisticated donors who are working in this space and asking questions as to how non-profits can use and manage large datasets (E. Fegraus, *verbally*). This is expected to aid in the uptake of technology in the conservation world (ibid). Financial awards for innovation are also available from organisations like Google, who recently awarded ZSL with an impact award for developing a new camera trapping system to deter poachers (R. Freeman, *verbally*).

Additional Opportunities for Big Data Adoption

- **Early days** –Big data is still in its early days in terms of development and applicability, and a heightened awareness throughout society is needed to enable the full potential of big data for social good (E. Anderson, *verbally*);
- **Data as an asset** – Data is becoming a metric and does not only create value, but also has inherent value as an asset (Simon, 2013). Traditionally, the metric for scientists has been number of citations, but this is beginning to shift as funding bodies and journals are learning that data is valuable in its own right (Kolker, 2013; R. Freeman, *verbally*);
- **Merging of sustainability and strategy** – The UN-Accenture CEO study revealed that 40 per cent of CEOs interviewed claimed that they are actively seeking to move from sustainability as a set function to

its integration into other functions, such as marketing, finance, and research and development (R. Hayward, *verbally*). Sustainability is no longer viewed in a silo, but instead as part of an organisation's global strategy (ibid). Gunn added that big data presents an opportunity to, "connect the dots" between critical functions such as fundraising, fieldwork, and community engagement (*verbally*). At present these functions are separate from each other at WWF-UK (ibid);

- **Leaders and innovators** – From a leadership standpoint, big data efforts grew within the World Bank initially due to pockets of people that pushed for innovation, which was supported by managers in key positions (D. Wheeler, *verbally*). The potential for increased big data use is also associated with generational turnover as senior leaders are retiring and being replaced by a generation that is more readily adopting new means of collecting and analysing information (M. Boatman, *verbally*);
- **Citizen science** – Citizen science, defined as, "scientific work undertaken by members of the general public, often in collaboration with or under the direction of professional scientists and scientific institutions (oed.com)," presents an opportunity for the enhancement of data's volume, variety, and velocity (M. Sneary, *verbally*). Data generated by citizen science presents analytical opportunities from unique methods such as mobile applications without the necessity to know what to look for in the data (ibid); and
- **Forums, workshops, and hackathons** – Bringing together a community of people interested in the development of big data applications for environmental sustainability through forums, workshops, and hackathons has been identified as an opportunity. Geo for Good, which is hosted by Google as an annual workshop to bring NGOs and not for profits together for the latest capabilities on Google's mapping platform (R. Freeman, *verbally*), is facilitating greater awareness of big data use. Hammer discussed the establishment of EcoHack, which brought together roughly 700 people this past year for a two-day hackathon for engineers, designers, and environmental scientists across four cities to discuss how technology can be applied to environmental sustainability (*verbally*).

Given the many barriers and opportunities discussed by these organisations, what is the way forward for them and others with environmental sustainability mandates? In the following section I synthesise findings and discuss how big data and environmental sustainability can coalesce to enhance the performance of the sustainability effort.

4 Discussion

4.1 The Need for Clarity

Several organisations claimed that big data requires clarified meaning and parameters (T. Cambridge, A. Armstrong, B. Blundell, E. Anderson, *verbally*). Indeed, this study reveals that the organisations working closest with nature, NGOs, are not speaking the same language in terms of how they describe their datasets and analytics. The disconnect also presents itself between categories. Given that partnerships were widely identified as an opportunity for big data adoption, it seems prudent to facilitate a common language. For example, an NGO that uses big data terminology might better engage with emerging sources of funding that could otherwise fly overhead.

Standards are also important for risk mitigation. Risks to privacy, security, equality, and even endangered species were raised as barriers to big data adoption. Parameters must be set to ensure not only that risks are mitigated, but also that reasons against big data uptake are addressed so it can be applied to its full potential. Many of the risks associated with big data could be mitigated, such as the pinpointing of endangered species' locations, which are already widely known by the entities that pose a threat to them (E. Fegraus, T. Cambridge, *verbally*). Big data risks should therefore be managed rather than used as a justification for inactivity (ibid).

The BT Better Future Forum confirmed the need for leadership in the establishment of big data standards. Without them it is difficult to understand how the collaborative partnerships needed to drive big data solutions to scale can form (E. Anderson, *verbally*). Given the government's experience in establishing open data parameters, extending those efforts to big data standardisation seems appropriate (MGI, 2011). Additionally, greater government outreach is needed for organisations looking to escalate their sustainability efforts through big data solutions (E. Anderson, M. Sneary, *verbally*). Concurrently, raised environmental performance standards through government regulation might reduce corporate fears that investing in data analytics for sustainability efforts will jeopardise competitiveness (R. Hayward, *verbally*).

4.2 The Relevance of Big Data to Environmental Sustainability Efforts

4.2.1 Adopt or Risk Irrelevancy

Does late or non-adoption of big data place organisations at a disadvantage? Simon made a sweeping statement that suggests this is indeed the case:

> The revolution is here, and it's high time that organisations of all sizes recognize it. ...Let the tinkering with Big Data begin across the board: public and private sectors, big and small companies, for-profits and non-profits. …The uninitiated, the skeptics, and the laggards who refuse to integrate data into their decision-making – and Big Data in particular – will only be left further and further behind (2013, 218).

If Simon's statement holds true, NGO scoring on the basic big data engagement indicators paints a bleak picture for the category. CI is the lone exception based on these criteria, but as discussed ZSL is incorporating data analytics that align with big data. ZSL also draws from Microsoft and Google to maintain its technological edge. BTO's work with data is core to its value proposition, but they have yet to on-board big data competencies, and WWF-UK is the least engaged at this stage with big data.

4.2.2 New Opportunities for Partnerships Based on Collaboration

The UN-Accenture CEO study highlighted an interesting shift. From 2007 to 2010 the studies revealed the declining influence of NGOs in informing and guiding corporate sustainability strategy, but in the 2013 study two-thirds of CEOs indicated that partnerships with NGOs are crucial to delivering sustainability (R. Hayward, *verbally*). 'Convergence' (ibid) entails mutual benefits to a partnership. Notably, CI engaged with HP to converge, and big data was the driving factor (E. Fegraus, *verbally*), illustrating that big data can be the enabler for convergence and environmental sustainability goal attainment.

The IUCN used to work with a service provider for software needs, but this relationship was of a transactional nature, which IUCN eventually found to be too expensive and of little reward (M. Sneary, *verbally*). WWF-UK has been in discussions with service providers that have approached the organisation, but WWF-UK questions whether or not they house enough data to justify the effort. Opportunely, the initiative exists on both sides, making big data a highly relevant medium to connect conservation goals with resource needs.

4.2.3 Big Data's Alignment with Conservation Goals and Values

Can big data solutions align with the nature of conservation efforts? Gunn noted that big data gains traction with sustainability issues that occur at a large scale, such as climate change (*verbally*). Hands-on conservation tends to be applied to changing local behaviours for environmental outcomes, and accordingly, fundraising is directed to localised conservation efforts rather than to big data solutions (ibid). Wheeler and WRI demonstrated however, that big data can indeed empower communities to generate change themselves by providing them with pinpoint, source information on environmental issues (D. Wheeler, *verbally*).

Are big data solutions aligned with conservation values? After giving a talk on the topic of big data and conservation, Paul Jepson, a Course Director within Oxford University's Department of Geography, was immediately asked, "What if this [technological application] isn't the reason we do conservation? What if we do it so that we can be outdoors in nature?" In other words, big data's alignment with conservation values was being challenged by the perception that technology acts as a divide between people and nature. Freeman suggested the opposite however, providing a values-driven reason for how big data aligns with conservation values:

> Obviously we still are impassioned by the natural world, but it's also true that there are now technologies that allow us to gather this data in the wild. …And that kind of data gathering in the wild, in the dark, in the rain, and on top of a mountain would not be possible without the technologies that are now available. And I think it's increasingly true that new technologies don't just move us away from real animals. They might give us a much richer understanding of them (*verbally*).

In summary, the range of reasons why big data should be relevant to organisations' sustainability efforts lends hope to its adoption, but they also raise a need for reciprocity from those involved in big data developments and efforts.

4.3 The Relevance of Sustainability to Organisations' Big Data Efforts

4.3.1 Revisiting Collaboration Opportunities

In addition to climate change threats and ecosystem health declines, what might trigger big data action toward sustainability efforts? Continuing along the line of reasoning from big data partnership opportunities, there is an emerging interest from the environmental field in big data that provides specialists with opportunities to build their businesses' brands while also showcasing their technologies. Compared with past decades, environmental sustainability is considered an emerging market (A. Feldman, *verbal*), and according to Freeman of ZSL, big data is very high on the environmental and ecological agenda (*verbally*). Notably, the NGO category was the most receptive to speaking on the topic of big data by measure of positive returns to interview requests. Five of the seven NGOs contacted offered interviews, signalling high interest in the topic. By far the lowest percentage of positive returns came from the corporate category however, indicating perhaps less interest in speaking on the topic despite having the greatest capacity to put big data into action.

4.3.2 Lending Purpose and Effectiveness to Corporate Sustainability Efforts

Big data is generating large benefits for corporate sustainability efforts as demonstrated by BT and LinkedIn's energy use reductions (E. Anderson, K. Shea, *verbally*). The UN-Accenture CEO study revealed that 84 per cent of CEOs believe that business should lead efforts to define and deliver sustainable development goals, but importantly, only 38 per cent of CEOs believe that they can accurately quantify the value of their sustainability efforts (accenture.com/ungcstudy). The combination of these results suggests broadly that CEOs feel a sense of ownership of sustainability initiatives, but improved means of performance measurement are required. As such, big data should indeed have precedence to enable corporate sustainability success.

Moreover, the relevance of environmental data to big data efforts appears to gain traction when it is personalized, owned, and used as a component of an issue rather than the central focus (R. Hayward, J. Holm, *verbally*). For example, the "quantified self" movement is based on people's ability to track and survey their own health (Dumbill and Mohin, 2013, 115), and this has been enabled by sensors and monitoring devices that measure and provide information to inform actions and behaviours (Barrett et al, 2013). The physical environment, namely air quality and pollution, has a direct impact on health, and therefore large amounts of environmental data are being merged with health data to generate healthcare solutions (ibid). These popular big data applications employ environmental data for the sake of human health, which is quickly personalised ahead of the surrounding environment.

Indeed, businesses perhaps misinterpreted how to best utilize big data for their sustainability efforts, directing it to 'outputs', such as per cent carbon reduction data for sustainability reports, rather than 'outcomes', such as air or water quality that connect environmental data to people's daily lives (R. Hayward, *verbally*). As such, companies that are leading sustainability efforts are doing so by embedding environmental outcomes into their success stories, and big data is playing a central role (P. Lacy, R. Hayward, *verbally*). The same applies to the use of environmental data to drive business success and profits. Originally, businesses were not generally applying environmental data for the sake of sustainability, but a clear case for efficiency and cost reduction can serve sustainability as well (R. Hayward, *verbally*).

4.4 How Big Data and Environmental Sustainability can Better Merge

Several shifts are happening in combination with some of the previously discussed opportunities associated with big data to enable a rapid merging with environmental sustainability interests and initiatives.

4.4.1 Emerging Business Models and the Circular Economy

The emergence of the 'circular economy' has been described as the single most important development to revolutionising environmental sustainability (P. Lacy, *verbally*). The circular economy entails a shift away from the classic linear business model, which often involves a product's creation, its use, and its discard to waste. The desired circular model, "advocates the need for a 'functional service' model in which manufacturers or retailers increasingly retain the ownership of their products and, where possible, act as service providers—selling the use of products, not their one-way consumption" (ellenmacarthurfoundation.org, 2013). Big data has a central role to play in conjunction with the digital revolution underway, and the uptake of these models could occur on a large scale, given that a third of CEO's in the 2013 UN-Accenture CEO study, "report that they are actively seeking to employ circular economy models" (48).

4.4.2 The Merging of Sustainability and Strategy

Further to the development of new business models that are employing big data for environmental sustainability efforts at a growing rate, corporate strategy is increasingly merging with environmental sustainability (R. Hayward, *verbally*). Energy and resource scarcity entail that companies must incorporate environmental sustainability into their strategy or risk continuity of operations and brand damage (ibid). Companies such as Unilever, BT Group, and Phillips are leaders in this regard, working ahead of government regulation to make environmental sustainability a central part of global strategy (P. Lacy, *verbally*). The merging of these two areas increases the likelihood that sustainability will be allocated more resources such as big data analytics tools and platforms.

5 Conclusion

In an article written for The Economist, Cukier (2010) drew upon advice from Mundie of Microsoft and Schmidt of Google in discussing how big data is likely to yield large benefits across different fields. Sitting on a presidential task force to reform American healthcare, the two men suggested:

> Look, if you really want to transform health care, you basically build a sort of health-care economy around the data that relate to people…. You would not just think of data as the 'exhaust' of providing health services, but rather they become a central asset in trying to figure out how you would improve every aspect of health care. It's a bit of an inversion (ibid).

Since that article was written in 2010, new literature continues to emerge that highlights how big data is being applied to revolutionise healthcare, indicating that the transformation is underway. My concern is that there has been little indication that environmental sustainability has a place on the big data agenda and vice versa.

I have presented findings that to varying degrees indicate that organisations across sectors have adopted big data in name and are shifting the thinking within the field from traditional data collection methods and approaches based on singular hypotheses to novel methods of collecting and mining data. Organisations like CI, WRI, BT, Anthesis, and the US government are making big data an integral part of their work and efforts toward environmental sustainability.

More work is required not only to better refine big data's meaning and standards, but also to generate further discussions that lend insight into how big data might be made applicable and available to organisations with different sustainability mandates. The opportunities raised to adopt big data however, demonstrate that these organisations are thinking about how to make big data central to the value they can bring to environmental sustainability efforts.

There is certainly cause for excitement. Through emerging trends, such as partnerships based on collaboration rather than transaction, outcomes rather than outputs driven applications, innovative circular business models, and the uniting of corporate strategy and sustainability, big data is positioned to become an integral element of environmental sustainability mandates.

I look ahead with enthusiasm to the future of environmental sustainability with big data further integrated and hope that the approach is a multi-disciplinary one between sectors, industries, and roles to drive the innovation forward on different fronts. Principally, I hope that this paper serves as a conversation starter, supplementing discussions on how the environment and data science merge to bring the sustainability of the natural world to the forefront of big data in action.

References

Accenture (2013) "The UN Global Compact-Accenture CEO Study on Sustainability 2013: Architects of a Better World". Retrieved from http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture-UN-Global-Compact-Acn-CEO-Study-Sustainability-2013.PDF.

Accenture (2014) About Accenture. Retrieved from http://www.accenture.com/us-en/company/Pages/index.aspx.

Accenture (2014) Big Data and Technology. Retrieved from http://www.accenture.com/us-en/strategy/analytics/big-data-technology/Pages/index.aspx.

Anthesis Group. About. Retrieved from http://anthesisgroup.com/about/.

Anthesis Group. Software & Systems. Retrieved from http://anthesisgroup.com/software/.

Barrett, M. A., Humblet, O., Hiatt, R. A., & Adler, N. E. (2013) "Big Data and Disease Prevention: From Quantified Self to Quantified Communities". Big Data. Vol. 1, No.3. Mary Ann Liebert, Inc.

British Trust for Ornithology. About BTO. Retrieved from http://www.bto.org/about-bto.

BT Group. About BT Group. Retrieved from http://www.btplc.com/thegroup/.

BT Group (2014) "Better future Forum – Highlights Report". Retrieved from http://betterfutureforum.bt.com/res/BFF2014highlights.pdf.

BT Group. Better Future Programme. Retrieved from http://www.btplc.com/Betterfuture/BetterBusiness/BetterFutureprogramme/index.htm.

Caldecott, B., Tilbury, J. & Carey, C. (2014) "Stranded Assets and Scenarios". Retrieved from http://www.smithschool.ox.ac.uk/research/stranded-assets/Stranded%20Assets%20and%20Scenarios%20-%20Discussion%20Paper.pdf.

Conservation International. About Us. Retrieved from http://www.conservation.org/about/Pages/default.aspx.

Conservation International (2013) Revolutionary Big Data Environmental Analytics Indicates Declining Animal Populations in Studied Tropical Forests. Retrieved from http://www.conservation.org/NewsRoom/pressreleases/Pages/Revolutionary-Big-Data-Environmental-Analysis-Indicates-Declining-Animal-Populations-in-Studied-Tropical-

Forests.aspx.

Crampton, J. W. et al. (2013) "Beyond the geotag: situating 'big data' and leveraging the potential of the geoweb", Cartography and Geographic Information Science, 40:2, 130-139, Retrieved from http://dx.doi.org/10.1080/15230406.2013.777137.

Cukier, K. (2010) "Data, Data Everywhere". The Economist. Retrieved from http://www.economist.com/node/15557443

CUREC (2011) Policy statement. Retrieved from https://www1.admin.ox.ac.uk/curec/policystatement/.

Data.gov. About Data.gov. Retrieved from http://www.data.gov/about.

Data.gov. Get Started. Retrieved from https://www.data.gov/.

Dumbill, E. (2013) "Making Sense of Big Data". Big Data. Vol. 1, No. 1. Mary Ann Liebert, Inc.

Dumbill, E. & Mohin, S. (2013) "Opportuninties at the Intersection of Health and Data". Big Data. Vol. 1, No. 3. Mary Ann Liebert, Inc.

Elkington, J. (2014) Get Ready For The Breakthrough Decade…. Volans. Retrieved from http://volans.com/2014/05/get-ready-for-the-breakthrough-decade/

Ellen MacArthur Foundation (2013) The Circular Model – An Overview. Retrieved from http://www.ellenmacarthurfoundation.org/circular-economy/circular-economy/the-circular-model-an-overview.

Graham, M. & Shelton, T. (2013) "Geography and the future of big data, big data and the future of geography". Dialogues in Human Geography. 255-261. Sage.

Haklay, M. et al. (2008) "Web Mapping 2.0: The Neogeography of the GeoWeb". Geography Compass. 2011–2039. Retrieved from ftp://s-she-junco.cits.rncan.gc.ca/pub/cartonat/Reference/VGI/ART-Hacklay,%20Singleton,%20Parker,%202008.pdf.

Haklay, M. (2013) "Neogeography and the delusion of democratisation". Retrieved from: http://discovery.ucl.ac.uk/1384805/1/Situating%20Neogeography%20Web.pdf.

Hewlett Packard. About HP Earth Insights. Retrieved from http://www8.hp.com/h71061/index.html#!panel=about.

IBM (2012) "Managing Big Data for Smart Grids and Smart Meters". Retrieved from http://www-935.ibm.com/services/multimedia/Managing_big_data_for_smart_grids_and_smart_meters.pdf

International Institute for Analytics (2013). The Rise of Analytics 3.0: How to compete in the data economy. Retrieved from http://iianalytics.com/analytics-resources/analytics-3.0.

IPCC (2013) "Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovern mental Panel on Climate Change" [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.

IUCN (2014). About IUCN. Retrieved from http://iucn.org/about/.

Kolker, E. et al (2013) "Towards more Transparent and Reproducible Omics Studies through a Common Metadata Checklist and Data Publications". Big Data. Vol. 1, No. 4. 196-201. Mary Ann Liebert, Inc.

Laney, D. (2001) "3D Data Management: Controlling Data Volume, Velocity, and Variety". Meta Group Inc. Retrieved from http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.

LinkedIn Press Centre (2014) About LinkedIn. Retrieved from http://press.linkedin.com/about.

Manyika, J. et al (2011) "Big Data: The next frontier for innovation, competition, and productivity". McKinsey Global Institute.

Mayer-Schönberger, V., & Cukier, K. (2013) "Big Data: A Revolution that will Transform how we Live, Work and Think". John Murray. London, UK.

Millennium Ecosystem Assessment (2005) "Ecosystems and Human Well-Being: Synthesis". Island Press, Washington DC. Retrieved from http://www.unep.org/maweb/documents/document.356.aspx.pdf.

Open Definition. The Open Definition. Retrieved from http://opendefinition.org.

Oxford English Dictionary (2014) Data. Retrieved from http://www.oed.com/view/Entry/296948?redirectedFrom=data#eid.

Oxford English Dictionary (2013) New words notes June 2014. Retrieved from http://public.oed.com/the-oed-today/recent-updates-to-the-oed/june-2014-update/new-words-notes-june-2014/.

Oxford Dictionaries (2014) Hackathon. Retrieved from http://www.oxforddictionaries.com/definition/english/hackathon.

Provost, F. & Fawcett, T. (2013) "Data Science for Business: What you need to know about data mining and data-analytic thinking". O'Reilly Media.

Simon, P. (2013) "Too Big to Ignore: The Business Case for Big Data". John Wiley & Sons, Inc. Hoboken, New Jersey.

Sowe, S. K. & Zettsu, K. (2014) "Curating Big Data Made Simple: Perspectives from scientific communities". Big Data. Vol. 2, No. 1. 23-33. Mary Ann Liebert, Inc.

Sui, D., Elwood, S., & Goodchild, M. (Eds.) (2012) "Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice". Springer. London, UK.

Sweet, L. E. & Moulaison, H. L. (2013) "Electronic Health Records Data and Metadata: Challenges for big data in the United States". Big Data. Vol. 1, No. 4. 245-251. Mary Ann Liebert, Inc.

The United States Environmental Protection Agency (2014) What is Sustainability? Retrieved from http://www.epa.gov/sustainability/basicinfo.htm.

The World Bank (2014) About. Retrieved from http://www.worldbank.org/en/about.

Tropical Ecology and Assessment Monitoring Network. About TEAM. Retrieved from http://www.teamnetwork.org/about.

Turner, A. J. (2006) "Introduction to Neogeography". O'Reilly Media, Inc. Retrieved from: http://pcmlp.socleg.ox.ac.uk/sites/pcmlp.socleg.ox.ac.uk/files/Introduction_to_Neogeography.pdf.

Volans (2013). What we Offer. http://volans.com/about/what-we-offer/.

Willcocks, L., Ventors, W. & Whitley (2011) "Meeting the Challenges of Cloud Computing". Accenture. Retrieved from http://www.accenture.com/us-en/outlook/Pages/outlook-online-2011-challenges-cloud-computing.aspx.

World Resources Institute. About WRI. Retrieved from http://www.wri.org/about.

World Resources Institute (2014) The Data Lab. Retrieved from http://datalab.wri.org.

World Resources Institute. Global Forest Watch. Retrieved from http://www.globalforestwatch.org/about/video.

World Resources Institute (2014) RELEASE: Global Forest Watch: Dynamic New Platform to Protect Forests Worldwide. Retrieved from http://www.wri.org/news/2014/02/release-global-forest-watch-dynamic-new-platform-protect-forests-worldwide.

WWF. About WWF. Retrieved from http://www.wwf.org.uk/about_wwf/.

WWF. Living Planet Report. Retrieved from http://wwf.panda.org/about_our_earth/all_publications/living_planet_report/.

ZSL. About the Zoological Society of London. Retrieved from http://www.zsl.org/about-us.

Smith School of Enterprise and the Environment
University of Oxford
South Parks Road
Oxford, OX1 3QY
United Kingdom

E enquiries@smithschool.ox.ac.uk
T +44 (0)1865 614963
www.smithschool.ox.ac.uk

SMITH SCHOOL OF ENTERPRISE AND THE ENVIRONMENT

UNIVERSITY OF OXFORD