THINK BIG:
BRITAIN'S DATA HEARTBEAT

wanDISCO

DATA IS THE NEW OIL OF THE INTERNET

MEGLENA KUNEVA
EUROPEAN CONSUMER COMMISSIONER

FOREWORD

The term Big Data can be somewhat vague – with many people using it as shorthand to cover the broad church of emerging technologies that also include analytics, storage, cloud computing and business intelligence.

In actuality, Big Data refers to when an organisation produces data at such a rate, variety and scale that it exceeds the processing capacity of conventional technology. While this once might have caused nothing but headaches, cost-effective approaches are emerging that makes taming this data deluge a viable undertaking.

Everything today produces data: from social networks, web server logs, traffic sensors, satellite photography and audio streams, to web pages, GPS trails, telemetry from vehicles and financial market reports.

It has long been posited that within these large data sets lie patterns of information that can be used to provide unparalleled insights into the way society operates – everything from tracking the spread of disease to predicting the voting habits of entire populations.

The emergence of Big Data has brought with it a new focus on agility and curiosity, with successful strategies revolving around experimentation and exploration, testing new ideas in a way that is far removed from the statistician's traditional approach.

The size of samples now available to analysts is massively improving the accuracy of forecasting and as such is one of the main draws of Big Data. Due to the sheer volume of the data being produced, thousands of individual variables can be taken into account – but it demands a new type of storage that can scale affordably. Storing such a large volume of information using traditional data warehouses would see costs spiral over time – with the volume of data only set to increase.

The speed at which data flows into an organisation has followed a similar pattern to that of volume, with the growing prevalence of the Internet and mobile technology seeing huge amounts of information being produced at an ever-increasing rate.

Today, for example, retailers are able to compile highly accurate profiles of customer experience, tracking everything from dwell time and individual click-throughs to final sales.

It is the ability to process this data in real time that is seeing retailers make highly accurate recommendations to customers based on their real-time shopping habits. But this demands technology that can crunch the information as soon as it trickles in, rather than leaving it to be analysed at a later date.

In many ways, however, Big Data rests on being able to collate digital information from a variety of sources. Rarely does data come neatly structured and whenever different browsers or feeds send through information, you can guarantee things will get messy.

The truth is Big Data's strengths only come to the fore when you combine information from many different outlets and analyse them collectively to identify patterns. For disasters such as the 2014 Ebola crisis, this meant being able to access health clinic reports, media updates, social media posts, information from public workers on the ground, transactional data from retailers and pharmacies, and travel ticket purchases alongside helpline data.

Crunching data on this scale necessitated an ability to handle a variety of information quickly, collating and analysing it in its native form as soon as it was produced all over the globe – a process known as multi-data centre ingest.

Here at WANdisco, we believe the UK stands to be one of the biggest winners in the Big Data space, in particular its public and third sectors. But while Big Data analytics are often championed by the private sector, its potential use by Government and charities has long been underexplored.

This report is a rallying call to Government and the Third Sector. It sets out specific areas where Big Data can make a tangible difference, and spelling out how effective strategies will not only save untold fortunes but open up new avenues of discovery.

In **Think Big: Britain's Data Heartbeat**, we aim to show that the UK stands on the verge of breakthroughs on a global scale. It is our belief that effective data analysis strategies will redefine the way in which society interacts with technology. Every imaginable decision-making process will become more efficient, more powerful and more lasting.

Welcome to the world of Big Data.

DAVID RICHARDS
CEO, PRESIDENT & CO-FOUNDER, WANDISCO

BIG DATA
BY NUMBERS

BY 2017 THERE WILL BE A TOTAL OF 30 BILLION INTERNET CONNECTED DEVICES IN THE WORLD

GARTNER 2014

THERE WILL BE OVER 40 TRILLION GIGABYTES ON EARTH BY 2020 – THAT'S 5,200 GIGABYTES FOR EVERY PERSON

IDC & EMC 2012

THE MARKET FOR BIG DATA TECHNOLOGY AND SERVICES WILL REACH $23.8 BILLION BY 2016

IDC 2013

OVER THE LAST 10 YEARS THE DIGITAL SHARE OF THE WORLD'S STORED INFORMATION HAS INCREASED FROM 25% TO OVER 98%

IBM 2013

THE BIG DATA MARKET WILL BENEFIT THE UK ECONOMY BY £216 BILLION AND CREATE 58,000 NEW JOBS BEFORE 2017

UK GOVERNMENT

THE BIG DATA OPPORTUNITY IN THE UK PUBLIC SECTOR COULD BE WORTH BETWEEN £16 BILLION AND £33 BILLION A YEAR

POLICY EXCHANGE

APPLICATION DOWNLOADS ARE INCREASING AT A RATE OF 50% YEAR- ON-YEAR – 85 BILLION IN 2014

APPLE 2014

BIG DATA JOBS ARE FORECAST TO INCREASE 92% BY 2017

SAS 2013

CHAPTER 01

01 SMARTER GOVERNMENT

Government IT doesn't mean what it used to. In the past, computers were merely a substitute for the pen and paper to carry out well-known back office processes – little more than 'faster paper,' in the words of tech expert and head of strategy at Silicon Valley Data Science, Edd Dumbill.

Fast forward to 2014 and an independent review called on the establishment of an expert technology ethics body, one that would help address complex challenges that are becoming increasingly frequent in a government driven by tech.

"The Government has to recognise that technology is no longer a subsection of a department, but part of every sector," stated the independent review, Making Digital Government Work for Everyone, written by an independent panel of more than 20 advisers and volunteers. "Every department in every local authority is going to be hoping to improve services while reducing cost and the only thing that is going to make that possible is technology – there is nothing else that enables engagement, productivity and reduces cost."

Today's Government has embraced technology like no other, with digital services at the heart of plans to improve public services. With the UK's Civil Service made up of 24 ministerial departments, 22 non-ministerial departments, and a further 345 government agencies and public bodies, employing a total of 447,000, the reach of the digital revolution is vast.

The data deluge is by no means a phenomenon restricted to the private sector. And faced with the pressure of slashing costs, the potential for the Civil Service to utilise Big Data is huge.

Buzzwords such as 'joined-up government' and 'efficiency savings' have been bandied around for years, but now have the opportunity to make real progress.

Joining up public sector sources can make government more efficient, save money, identify fraud and help public bodies better serve their citizens. And with industries becoming rapidly digitised, adapting to the data-driven world is just as time sensitive for the Civil Service as any other sector.

JOINING UP PUBLIC SECTOR DATA SOURCES CAN MAKE GOVERNMENT MORE EFFICIENT, SAVE MONEY, IDENTIFY FRAUD AND HELP PUBLIC BODIES BETTER SERVE THEIR CITIZENS.

CLAIRE VYVYAN
EXECUTIVE DIRECTOR AND
GENERAL MANAGER OF PUBLIC SECTOR AT DELL UK

DATA CAN ENABLE GOVERNMENT TO DO EXISTING THINGS MORE CHEAPLY, DO EXISTING THINGS BETTER AND DO NEW THINGS WE DON'T CURRENTLY DO.

TOM HEATH
HEAD OF RESEARCH AT THE OPEN DATA INSTITUTE

In a recent comment article for The Times, Matt Ridley praised the UK government's digital service (GDS) as one of the current administration's success stories. Championed by Francis Maude, as Minister for the Cabinet Office, supported by Baroness Lane-Fox and run by Mike Bracken, the GDS is reshaping the way the public sector implements and designs large IT projects to ensure that cost and time overruns are a thing of the past.

Whereas previous strategies relied upon bloated documents that tried to specify the needs of end users before a single line of code had been written, Francis Maude and Mike Bracken are teaching the Civil Service to start small, fail fast and – most importantly – learn to adapt in response to digital feedback from users in order to evolve processes along the way.

Online services provided by the public sector now begin with a discovery phase, lasting six to 12 weeks, before an 'alpha' prototype of a working software service is produced, eventually followed by a 'beta' that's only available to a private audience of specialist users. Only once rigorously tested is this opened up to the public and only much later is the old service turned off.

At all times during this process, feedback is taken on board from a diverse array of users, allowing the service to be updated, modified and honed. This is the same approach taken by the many IT services we use today, from Facebook to BBC News.

This data-driven strategy has already led to vast savings on behalf of the taxpayer as contracts with external IT providers are shorter and smaller – an estimated £4bn a year will have been saved by the end of 2020.

The biggest victory for the Civil Service will be to use data analytics to better understand the services they are providing, evaluating what works, identifying early interventions and improving service design.

But a look further afield suggests that the Civil Service will face many similar problems that have blighted the private sector's attempts to go digital. Here are our recommendations for a data-driven government:

INVEST IN SKILLS

Ensuring that staff have the adequate skills needed to make the most of the technology is paramount. Policymakers might well be dazzled by the possibilities presented by data solutions, but more work could be done to ensure their staff have a solid grounding in statistics in order to properly articulate the question and interrogate the conclusions. Training programmes that promote digital and data skills – in addition to basic scientific literacy – must be placed into the core Civil Service competency framework.

RECRUIT MORE DATA SCIENTISTS

It is vital that government departments make use of more statisticians, analysts, and data scientists. IT departments classically serve a vital role, but they're generally not populated with great analysts, often threatened by government businesses having coders and programmers working on and making sense of their data.

INCENTIVISE USE OF DATA ANALYTICS

Resistance to change is a common occurrence in most companies, while an aversion to using new technology for data-driven decision-making is often the largest obstacle facing wholesale adoption. Staff should be incentivised to increase their use of data and analytics, including payment-by-results.

INTRODUCING **BIG DATA**

HADOOP

Apache Hadoop is the enterprise framework of Big Data that is used both to store and process incredibly large data sets.

The underlying technology was originally invented by Google to index all the rich textual and unstructured information they were collating, and present results to users in a meaningful way.

Yahoo developed Apache Hadoop as an enterprise platform incorporating this technology, allowing companies to run extensive analytics of both structured and unstructured data – information that doesn't fit nicely into tables. In the same way Google indexed users' search behaviour, Hadoop lets organisations learn more about customers and consumers.

The system is designed to run on a large number of machines that don't share any memory or disks. When an organisation loads its data into Hadoop, the software separates that information into pieces that it then spreads across different servers – meaning there is no one place where you manage all of your data. Because there are multiple copy stores, data stored on a server that goes offline or dies can be automatically replicated from a known good copy.

CLOUDERA

Cloudera was founded in 2008 to provide the first enterprise-ready implementation of Apache Hadoop, set up by three engineers from Google, Yahoo and Facebook – three businesses that were among the trailblazers of Hadoop.

Based in Palo Alto, California, Cloudera has enjoyed rapid growth since first receiving some $5 million worth of funding in 2009. The company has expanded to become one of the leading providers of Big Data solutions, used by a diverse range of companies and organisations including Expedia, BT, Western Union, Nokia and eBay.

In March 2014, the firm received a $740 million investment from Intel, with its 15% stake valuing the firm at $4.1 billion. Cloudera is one of a handful of Silicon Valley start-ups to receive multi-billion dollar valuations from investors pre IPO.

By replacing its own Big Data project with Cloudera's solutions, Intel provided a strong indication that Hadoop is here to stay.

HORTONWORKS

Hortonworks does one thing: building, managing and implementing Hadoop. The company has devoted itself to working within the open source space, reaching out to customers through the existing products of its partners such as Microsoft, Teradata and SAP. It houses the largest collection of Hadoop "committers" – the name given to those who add code to the Hadoop mainframe.

Founded in June 2011, the firm's dedication to open source has seen it quickly established as one of the major players within the Hadoop community.

The company was set up by 24 engineers who were involved in the original development of Hadoop at Yahoo and, backed by Benchmark Capital, the firm raised $25 million in November of that year.

Named after Horton the Elephant of the Horton Hears a Who! book, the company was described by Forrester Research as the "technology leader and ecosystem builder of the entire Hadoop industry".

FIVE REASONS WHY YOUR HADOOP NEEDS WANDISCO

01

Continuous availability with Maximum Performance and Scalability

WANdisco's patented technology removes the bottleneck of a single active NameNode in a single data centre, balances workload across data centres, and enables LAN-speed read/write at any location without downtime or data loss – even during Hadoop upgrades and maintenance.

02

100% Use of Compute Resources

Active-active architecture eliminates read-only backup servers by making all servers fully readable and writable, so you can take full advantage of the hardware at each location. For even greater efficiency, features such as selective replication and asymmetric block replication allow different hardware footprints in each location.

03

Cluster Zoning

Delegate your most resource-intensive data load and in-memory applications to high-spec servers while running less critical batch applications on commodity servers. Maintain quality of service for all users and eliminate the need for costly hardware throughout the cluster.

04

Multi-Data Centre Ingest

Ingest data at any number of locations simultaneously, automatically replicate where you choose, and analyse from anywhere with no single point of failure. No admin overhead, no loss of data.

05

Selective replication with global access

WANdisco makes Hadoop data centre-aware so you can perform global, company-wide roll-up analysis from anywhere regardless of how your data is distributed.

CHAPTER 02

02 BETTER OUTCOMES IN HEALTHCARE

> There is perhaps no greater indication of humanity's achievements than the advancements made by modern medicine.

From advancements in sterilisation through to the discovery of anaesthetics, X-Rays and Penicillin, progress has been marked by discoveries that are typically defined by 'Eureka' moments, achievements reliant on inspiration to take the medical world forward.

But this paradigm could all change, with many predicting that Big Data will usher in a new era for medical science, one that relies less on individual brilliance and more on the collective potential of analytics and algorithms.

Healthcare has long been held up as one of Big Data's early benefactors. The data deluge of volume, variety and velocity fuelled by the rise of intelligent and interconnected devices has not escaped the medical world and there is huge potential for innovation.

Everything from heart monitors and ventilators, to medicine dispensers and thermometers now send out a seemingly infinite amount of data. The same is true of our smart devices, many of which can also function as mobile health centres capable of checking our condition every minute.

Within all the bleeps and electronic noise are trends that are informing the next generation of pharmacists, doctors and nurses responsible for our health. And thanks to data analytics, crunching the vast amounts of electronic data that's emitted and collated by the second is already providing hugely significant insights and valuable findings.

Almost everything that determines our health – from our individual genetic coding to our particular retail habits – is becoming knowable. With all this information at its fingertips, the medical profession is capable of spotting patterns of disease, gauging the efficacy of treatments and identifying links between causes and symptoms.

This is where the UK has a massive advantage. Dating back to the 1940s, the NHS possesses a library of information that could provide doctors with unparalleled insights. Using this information to spot patterns of disease, gauge the efficacy of treatments and identify links between causes and symptoms is all within reach. The more data available to be analysed, the better the medical insights and the more lives that can be saved.

WE ARE MOVING FROM A WORLD WHERE WE TREAT ILLNESSES TO ONE WHERE WE PREDICT AND PREVENT THEM.

MATTHEW WALL
BUSINESS REPORTER, BBC NEWS

THE NHS IS A RELENTLESS PRODUCER OF DATA – BE IT PATIENT DATA, PERFORMANCE DATA, CLINICAL DATA, OR ADMINISTRATIVE DATA [...] IN MAKING GOOD USE OF ITS DATA, THE NHS CAN BE MORE EFFICIENT AND WILL BE BETTER PLACED TO MEET ITS OWN TARGETS.

DAVID DOWNING
DIRECTOR OF HEALTH, SAS UK

The British government's commitment to Big Data has featured significant investment in the health sector. In addition to the Care. data initiative, Westminster has also unveiled its backing for a new Big Data institute in partnership with Oxford University that will see the creation of a research centre dedicated to using digital technology to drastically improve the sector.

The Li Ka Shing Centre for Health Information and Discovery, which will form part of Oxford University's Medical Sciences Division, will be used by researchers to advance genomic and genetic medicine. They hope to accurately identify good drug targets for industry to pursue, using biology and automation to speed up the early stages of drug discovery for cancer, diabetes, dementia and psychiatric conditions.

The Institute has the specific brief of using things like patient records, DNA sequencing, clinical trials, medical imaging and national registries to enable health care agencies to use this information more effectively.

Another government-backed project, a partnership between Genomics England and the US biotechnology firm Illumina, seeks to map 100,000 people's genomes by 2017. The aim of the initiative is to use Big Data and genetics to develop personalised medicine, to ensure that patients receive the most appropriate treatment.

Personalised medicine has been hailed as being the most significant medical development of the last 50 years, described by Peter Johnson, chief clinician at charity Cancer Research UK, as "the most exciting change in cancer treatment since the invention of chemotherapy."

Current treatments, based as they are on type and stage, fail many cancer sufferers, and the hope is that a more sophisticated understanding of genetics will stem this tide. Cancer researchers are aiming to use the DNA of an individual's cancer cells, rather than the patient's inherited DNA, for the process. This requires the genetic sequencing of the 3 billion components of the cancer's DNA, a technique reliant on Big Data platforms such as Hadoop to be completed effectively.

Ultimately, it is hoped health providers will be able to offer bespoke treatments on a patient-by-patient basis.

Some of the most exciting work today is taking place through the collaboration between Silicon Valley giants and not-for-profits.

Intel recently announced that it is working with the Michael J. Fox Foundation for Parkinson's Research on a new pilot initiative that is aimed at using data mined from wearable devices to detect patterns in the progression of the disease. The data will be collected on Intel's software platform, and analysts will examine it to look for markers of Parkinson's that cannot be perceived by the naked eye.

The potential to collect and analyse data from thousands of individuals on measurable features of Parkinson's, such as slowness of movement, tremor and sleep quality, could enable researchers to assemble a better picture of the clinical progression of Parkinson's and track its relationship to molecular changes.

Wearables can unobtrusively gather and transmit objective, experiential data in real time, 24 hours a day, seven days a week. With this approach, researchers could go from looking at a very small number of data points and pencil-and-paper patient diaries to analysing hundreds of readings per second from thousands of patients. Attaining a critical mass of data will let data scientists detect patterns and make new discoveries.

Cheap blood glucose level testers, from companies such as Sanofi Aventis, can now be added to smartphones with the subsequent data recorded and collated for analysis. It is hoped that the insight they provide will be able to halt an obesity epidemic that means far more children are now succumbing to type-2 diabetes.

The data can be combined with other information from exercise and diet apps to build up a more complete picture of patient lifestyles.

But bringing all these disparate data sets together and standardising them is a big challenge. Many expect the UK to arrive at a scenario whereby the NHS has completely digitalised healthcare records but a lot of this data is in physically separated databases, with about 400 in the average hospital.

In addition to providing tailored medicine and advancing preventative methods, Big Data analytics can help squeeze out costly inefficiencies in the medical system.

As populations grow and people live longer, healthcare costs are growing to unsustainable levels, with the NHS tasked with finding annual savings of £20bn by 2015.

By looking at doctors' prescriptions for cholesterol-reducing medicine across the country, Prescribing Analytics, a joint venture comprising NHS doctors, academics and tech start-ups, found that more than £200m a year could be saved if doctors switched from branded to generic versions of drugs. Other research has found these savings could amount to £1.4bn if the same approach were applied to a range of community-prescribed drugs.

Beyond the UK, hospitals are making better use of data to reduce waiting times in emergency rooms, track patient movements and moderate X-ray dosages as part of a strategy to manage resources more effectively.

Access to vast amounts of data is enabling companies to forecast which demographics are likely to cost the most to treat in the future for conditions such as asthma or diabetes. This is helping healthcare providers to intervene earlier and redesign their services to cope with the massive increase in healthcare demand.

CASE STUDY

UC IRVINE HEALTH

UC Irvine Health comprises the clinical, medical education and research enterprises of the University of California, Irvine. UC Irvine Medical Center was listed among America's Best Hospitals for the 13th consecutive year in 2013 by U.S. News & World Report.

The hospital is one of the world's first medical institutions to embrace the opportunity provided by Big Data, unveiling a radical scheme that is helping to reduce the number of deaths caused by medical error.

WANdisco is helping the hospital use Big Data platform Hadoop to digitally collate, store and analyse all data relating to its patients' conditions in real time. Electronic signals sent out by equipment such as heart monitors, ventilators or wearable devices can now be monitored whether the patient is in the hospital, at home or on the move.

UCI can now process accurate pattern-set recognitions, use algorithms to monitor patient recovery for non-linear complications, and build predictive-modelling systems to minimise deaths caused by medical error. This is drastically improving the level of care offered to patients with doctors alerted as soon as vital signs cross a key threshold. It has also helped ease the burden on doctors and nurses whose heavy patient loads prevent round-the-clock observation.

The technology makes it much easier for doctors and nurses to manage diseases such as heart failure and sepsis, a potentially fatal whole-body inflammation caused by severe infection, which progress algorithmically rather than in a linear manner. In these cases, a patient's condition can deteriorate drastically in a matter of minutes.

In order to commit to Big Data in this way, hospitals need to know that nothing will fail – they need round the clock availability and reliability of this information every second of every day.

CHAPTER 03

03 SAVING LIVES

Humanitarian organisations are facing an increasingly challenging future. Greater numbers of people are requiring aid, there are more people displaced for longer periods and natural and man-made scenarios are increasingly severe or complex. However, there continues to be a significant gap between the requirement for humanitarian intervention and the funding available for it.

Whilst donations have actually increased since 2008, so too has the cost of providing aid, leading to a growing necessity for humanitarian organisations to streamline their practices and deliver assistance as efficiently as possible. This is where Big Data comes in.

Big Data analytics is all about combining information from many different sources and analysing them collectively to identify patterns. For a crisis such as the Ebola outbreak, this means accessing health clinic reports, media updates, social media posts, information from public workers on the ground, transactional data from retailers and pharmacies, and travel ticket purchases alongside helpline data.

Being able to crunch data on this scale necessitates being able to do so quickly, collating and analysing it in its native form as soon as it is produced across the globe — a process known as multi-centre ingest. Combined with the vast quantities of public information already accessible via the Internet, Big Data can help ensure those working in hazardous environments are able to stay on top of ever-changing situations.

Only by crunching data on this scale can we truly determine in a timely manner whether containment policies, education campaigns and other preventative treatments are proving effective. Big Data is already transforming humanitarian responses across the world – particularly in two areas that have radically improved the precision of aid delivery:

ENHANCED EARLY WARNING

By being better able to predict the onset of a potential crisis, organisations are able to more accurately allocate funds, make preparations and mobilise resources. The 6.02 magnitude earthquake in Napa in August 2014 was predicted by the UC Berkley Shake Alert analysis technology. The prediction only came ten seconds before the event, but as such innovations improve it is realistic to envisage a time when the effects of crises could be significantly mitigated by early warning systems.

REAL-TIME AWARENESS

Receiving an on-going understanding of operational conditions gives humanitarian organisations the insight required to focus efforts and ensure the safety of staff. As sensor technology improves and hardware can be deployed which constantly monitor a broad range of conditions, large quantities of data can be remotely analysed and updates fed to frontline staff.

The response and reconstruction that followed the 7.0 magnitude Haiti earthquake was the first notable humanitarian response to incorporate Big Data analytics.

FINDING WAYS TO MAKE BIG DATA USEFUL TO HUMANITARIAN DECISION-MAKERS IS ONE OF THE GREAT CHALLENGES, AND OPPORTUNITIES, OF THE NETWORK AGE.

OCHA'S HUMANITARIANISM IN THE NETWORK AGE REPORT

A live crisis map of the 2010 earthquake was created using thousands of social media posts on platforms such as Facebook and Twitter. This map showed a number of variables including where victims were buried under collapsed buildings and where medical aid was needed. Within days, this map also included information gathered from an SMS number specifically devised for Haitians in need of assistance. This map proved crucial for first responders with hundreds of lives saved using the information charted on this map.

The difficulty is creating a network of sensors in areas that have been affected by major natural or manmade crises. A lack of technology or manpower has led to situations where no information is available from disaster-struck regions, or the information that is being generated is too disparate to efficiently log and manage.

One solution has been to track and analyse content from the media. As part of the UN's Global Pulse initiative, researchers used eight years of unstructured news data from French media to discover whether it was possible to map thematic shifts in media attention against world food security. Techniques such as data mining, text analysis and semantic clustering provided a clear correlation between emerging trends in food security and media articles that could be monitored in real time.

However, social media has emerged as the most effective way of crowdsourcing evidence during a crisis. In Forbes magazine, Patrick Meier, director of social innovation at the Qatar Computing Research Institute, explains the technology they are focusing on improving: "We are developing Twitter 'classifiers,' algorithms that can automatically identify relevant and informative tweets during crises. Individual classifiers will automatically capture eyewitness reports, infrastructure-damage assessments, casualties, humanitarian needs, offers of help, and so forth."

This combination of machine learning with effective data analysis strategies is revolutionising the response to humanitarian crises. Organisations from the private and public sector across the world are investing significant resources into R&D around this specific area.

Indeed, it is now accepted that Big Data should form a key component of the future strategy of both large and small humanitarian focused organisations. Governments and non-governmental organisations should consider how they can utilise data science to enrich their understanding of events and communities in order to make their actions more timely and effective. Only by being able to make a step change in the ability to predict, plan and execute in response to the world's humanitarian crises will the needs of beneficiary communities be met more effectively.

GOVERNMENTS AND OTHER ORGANISATIONS HAVE VALUABLE OPEN DATA THAT COULD HELP IN RELIEF EFFORTS – ABOUT ROADS, AIRPORTS, SCHOOLS, MEDICAL FACILITIES AND POPULATIONS. SUCH INFORMATION CAN HELP TO DRIVE DATA-DRIVEN DECISIONS DURING TIMES OF UNCERTAINTY.

JEANNE HOLM
AMERICA'S DATA.GOV INITIATIVE

THE BIG DATA MARKET IS FORECAST TO BE WORTH $32.1 BILLION IN 2015

RISING BY 49.5% TO $48 BILLION IN 2016 AND 66% TO $53.4 BILLION BY 2017

WIKIBON

CASE STUDY

UNICEF: A DATA-DRIVEN ORGANISATION

The United Nations Children's Fund (Unicef) was established by the United Nations in 1946 with the goal of providing long-term humanitarian and development assistance to children in developing countries.

Unicef works in more than 190 countries across the globe, protecting children from danger and helping to build a world where the rights of every child are realised. The organisation today is just one of a number of groups using integrated data technologies to overcome the barriers of time and distance.

Working with some of the poorest communities around the world and responding to emergencies, access to up-to-date information can mean the difference between life and death. In collaboration with a key set of partners, Unicef Innovation is designing and implementing programs, products and services for underserved communities. The ultimate goal is to strengthen essential services through real-time data and tracking and to deliver life-enhancing content that empowers underserved populations to access information and share their voice on matters concerning everything from government policy, health and sanitation to education, identity, security, and beyond.

Technology has gone from playing a relatively minor role in Unicef's operations to becoming an important part of the charity's strategic outreach. In 2007, Unicef's Ugandan office established a unique Technology for Development unit with the brief of working with government and regional partners to better develop bespoke solutions to local problems. Set up in collaboration with the government of Uganda and other development organisations, its DevTrac initiative was one of the early success stories, providing a online, transparent reporting on public services, in a way that merges real-time reporting with other data collected in the field. This tool is now open-source and available to any government or service delivery organisation to manage and publish the ongoing impact of their work.

This information-led approach enables greatly increased accuracy of decision-making processes at both a regional and national level, allowing governments to better determine allocation of resources.

A frequent problem encountered by both relief and government workers is that the information they work with can quickly become out of date – and nowhere is this more evident than in the tracking of infectious diseases. Through the use of mobile technology, Unicef is using SMS data from thousands of health facilities – produced by hundreds of thousands of health workers – to generate health reports in real time, providing governments with real-time information on registration of pregnancy and birth, electronic records of vaccinations and infant growth, tracking of disease outbreaks and medication stock levels.

In Rwanda, Unicef and its partners aim is for no woman to die during pregnancy. A successful initial pilot to register and track pregnant women, has now been extended nationwide and enables community healthcare workers to use SMS technology to support mothers and babies during pregnancy and within the crucial first 1,000 days of a child's life.

In Zambia a Unicef initiative is using these tactics to tackle the HIV epidemic, where early detection of the disease can have a significant impact on survival rates. Project Mwana uses mobile technology to reduce the time it takes to transport samples from test centres in rural communities for analysis in specialist centres. In these communities, distance and poor transport links between rural villages and the laboratories often represents the main delay that prevents an early diagnosis. Survival rates are up to 75% higher for HIV-positive newborns that are diagnosed and begin treatment within their first 12 weeks of life. Previously, the whole process could take as long as 66 days in total, but returning test results via SMS decreases the turnaround time by as much as 50% with the return journey completed almost instantaneously.

CHAPTER 04

04 A STRONGER DEFENCE

> Defence agencies across the world have invested significant budgetary resource into surveillance hardware, machine learning devices and sensors but are not yet harnessing the vast quantities of data these devices gather to see what insights can be revealed by them.

In 2013, Ben Rooney, European technology correspondent for the Wall Street Journal wrote: "The U.K. and other military forces are failing to manage the flow of data modern war-fighting equipment generate, which could result in operations failing." The results of the report he was covering, Big Data for Defence and Security from The Royal United Services Institute, gave damning evidence that the military services are being swamped, rather than enhanced, by data.

The defence sector has a particularly challenging set of requirements that need to be met by data analysis: whereas most businesses have localised architecture, defence services require databases which collect information from all over the world. The proliferation of data-enabled hardware means that numerous servers are asked to collect enormous quantities of data - without automatic and efficient analysis strategies, this creates a data lake with no tangible benefits.

The geographical diversity of sensors is also challenging, particularly as many of them are located in hostile environments. Traditional data collection methods are reliant on single points of data storage – a problem if that server explodes meaning the data is lost forever.

Although the defence sector has traditionally developed technology which is then successfully commercialised, in the Big Data space, the flow of development has been reversed.

The proliferation of Hadoop services has perfectly suited the needs of the defence industry – multi-centre ingest allows databases to receive information for unlimited numbers of sensors without the risk of downtime, whilst specific products such as WANdisco's Non-stop Namenode have solved the problem of server destruction. Rather than losing data, the information is remembered and gathered by a different server on that particular network.

BIG DATA HOLDS GREAT POTENTIAL FOR THE DEFENCE AND SECURITY SECTOR BUT [THE MOD] MUST NOT FALL INTO THE TRAP OF PROCURING ONLY BESPOKE SOFTWARE SOLUTIONS IF IT IS TO EXPLOIT THE TECHNOLOGY IN A TIMELY MANNER.

AIR CHIEF MARSHAL SIR STUART PEACH
VICE CHIEF OF DEFENCE STAFF

However, it is the possibility of real time analysis of data that represents the fundamental breakthrough for this sector. For agencies with the specific objective of identifying and neutralising security threats, this allows data scientists within the military to pinpoint outliers and anomalies in real time, rather than retrospectively.

Many global defence agencies are currently in the trial phases of modern data analysis strategies and the on-going data arms-race will be accelerated by the introduction of data scientists equipped with the technology to understand the information at their fingertips. Software has caught up with hardware.

That is not to say the sector is now free of challenges surrounding Big Data. Concerns around privacy and data protection remain, with an inability to analyse highly targeted meta-data leading to defence agencies having to reject data sets for fear of jeopardising anonymity. The technology exists to avoid this problem, but is relatively under-deployed.

The other problem is one of personnel. Whereas the private sector, which developed Big Data, has started to recognise the need for data scientists within their teams, this is still a new art in the defence industry.

The reality is that the wars of the future will be fought just as intensely by data scientists as soldiers, and countries need to prepare themselves accordingly. In modern combat scenarios a data scientist helping interpret and analyse data could save many more lives than a hundred troops on the ground.

The volume of data produced by the burgeoning number of unmanned systems – a medium altitude, long endurance system such as Reaper MQ9 or Watchkeeper, for example – can collect the equivalent of 20 to 40 laptops' worth of data per sortie. Much of this information is only retrospectively analysed.

SOURCE
WWW.TECHRADAR.COM/NEWS/WORLD-OF-TECH/CAN-THE-UK-BE-A-BIG-DATA-LEADER-IN-THE-MILITARY--1257275

IMAGE ESO/JOSÉ FRANCISCO SALGADO

CASE STUDY

GOVERNMENT USE CASE

WANdisco is working with some of the largest and most influential Government agencies in the Western world. The way in which these organisations deploy Big Data will transform our understanding of society.

The staggering potential of Big Data for Government is the depth of understanding that can be derived from the ability to find answers to questions you didn't even know existed.

WANdisco is working with a number of Government agencies that are currently overwhelmed by the sheer volume of information. With traditional databases, it is hard enough to store this data let alone run the queries that provide valuable insights.

One specific agency we work with has access to hospitals and transport hubs together with the entire power grid and social media networks. The sheer quantity of information generated by these resources is astronomical and constantly growing.

The key problem with so much data is the costs associated with managing these vast quantities of data. Whilst the gravity of this issue varies across Government departments, it is a universal concern.

The other major problem is uniformity. For example, flight departures and manifests are entirely different files to tweets and likes on Facebook. Until WANdisco, no solution existed to save and analyse these disparate types of data at the same time.

Reliability has also been a major factor holding back Governments from utilising Big Data. All the insights in the world are useless if they might suddenly disappear or become inaccessible. This is a particular issue for global operations in hostile environments.

Finally, whilst the access to so many forms of information offers Government agencies enormous opportunity, it also brings with it significant responsibility. Concerns around privacy are paramount and the technology previously did not exist to differentiate between what is relevant and what is not.

WANdisco is in the process of deploying its products and services to Government agencies in a way that will enable them to not only store vast quantities of data in the same place, but to make that resulting 'data lake' simple to question and query.

In the process, WANdisco is doing no less than unlocking the secrets of society.

For Government agencies involved in matters of life and death, being able to monitor anomalies in healthcare or public safety is vital. A data strategy built on real-time analysis turns the breadth of Government resources into a life-saving device, allowing incredible insights into unusual behaviour across the country.

In one particular instance, a regulatory body was struggling to deal with the representatives of major organisations misrepresenting their products and encouraging the sale of potentially dangerous medication. The agency had very little means of tracking and tackling such criminal behaviour. With Big Data, they can instantly monitor patterns of sales to determine where criminality is occurring and save time and resources pursuing a highly targeted and accurate disciplinary approach.

The patented technology behind WANdisco means that organisations can now rely on 100% access to their information, regardless of where it was submitted or stored. For Government agencies where access to its data is vital, this is a breakthrough that has opened up effective data analytics for the public sector.

Governments around the world are sitting on vast mountains of the world's most precious commodity – data. In the US, we are beginning to see an understanding of just how much data is worth and what can be done with it. WANdisco is proud to be the bank facilitating the trade of this new currency. Now the Big Data community must make the British Government realise the value of the data at its fingertips.

CONCLUSION

The British Government is investing in the modernisation of its own administrative systems and recognises the potential of the country's Big Data sector through initiatives such as the Alan Turing Institute for the education of data analysts. However, more needs to be done.

Governments around the world are already trialling Big Data techniques which go further, are more ambitious and will have a far greater impact on their citizens, than anything currently being attempted in the UK.

Big Data can obviously save money and improve speed in terms of governmental efficiency, but it also has the potential to drive innovation, direct research and save lives. This is where the British Government has a unique role to play in bringing together its own resources and the private sector.

When Bill Clinton announced a $1bn investment in the Human Genome Project in 2000, he was widely criticised. However, that investment allowed the US Government to engage with the leading minds in the private sector and a $100bn industry was created as a result. The US remains the world's leading innovator in genetics and that research has directly led to innumerable scientific breakthroughs.

Big Data has exactly the same potential. We do not advocate an entire open data platform for official data, but giving the right individuals and the right companies access to the right information will accelerate Britain's fight to become the globally dominant force in data analytics. No other country has the databases to generate the insights we have.

Concerns around privacy and anonymity persist. But we have the technology to analyse very specific data points and leave anything that identifies people untouched. The debate should not be around why we would endanger data protection for the growth of industry, but why we would possibly endanger the next era of disease prevention, sociological insight and disaster management.

The British Government must be the focal point around which the Big Data industry gathers. Its responsibility is to combine the business, social and moral arguments around the development of Big Data strategies and communicate them to the technology community.

The potential of Big Data across all sectors of business and society is such that a 100% commitment to the technology could become the Government's greatest achievement.

- £42 million committed to The Alan Turing Institute

  GEORGE OSBORNE MP
  MARCH 2014

- £113 million to be invested in a Big Data centre in Hartree, Daresbury

  GEORGE OSBORNE MP
  DECEMBER 2014

- 1 million+ data sets made open by Governments worldwide as of 2013

  MCKINSEY GLOBAL INSTITUTE OPEN DATA REPORT
  OCTOBER 2013

- €2.5bn EU investment in Big Data to fuel the European economy

  NEELIE KROES
  OCTOBER 2014

- Up to £66bn potential saving for the NHS by improving efficiencies through data analytics

  VOLTERRA AND EMC SUSTAINING UNIVERSAL HEALTH REPORT
  SEPTEMBER 2014

PREDICTIONS & RECOMMENDATIONS

Britain has an enviable reputation for technological innovation and with the right focus and investment, Big Data will become one of our greatest success stories.

In *Think Big: Britain's Data Heartbeat* we have mapped out what needs to be done for the UK to capitalise – here are our key findings.

01

Data is the most valuable asset currently available in the UK and its value continues to grow.

02

Government should work with the private sector to explore Big Data solutions that improve efficiency and performance in public services. The benefits must be communicated to the public to ensure mainstream buy in to data solutions.

03

The private sector should work with Government experts to generate a clear roadmap for a data-enabled future in Britain.

04

Data scientists should be well remunerated within Government. By demonstrating the value of experts in this emerging technology, Government can drive demand for the workers best equipped to drive innovation.

05

Government and the private sector need to share the aspiration that Britain can become a world leader in Big Data. By fuelling industry and maintaining our enviable public services, we should be proud to present Big Data as the technology on which our country is built.

ABOUT WANDISCO

WANdisco (Wide Area Network Distributed Computing) is a provider of enterprise-ready, non-stop software solutions that enable globally distributed organizations to meet today's data challenges of secure storage, scalability and availability. WANdisco's products are differentiated by the company's patented, active-active data replication technology, serving crucial high availability (HA) requirements, including Hadoop Big Data and Application Lifecycle Management (ALM).

Our customers include a host of Fortune 1000 companies such as Hewlett Packard, Intel, John Deere, European Southern Observatory, Barclays Capital, Walmart, GE, Cisco and Nokia.

wanDISCO

WWW.WANDISCO.COM