NPL
National Physical Laboratory

Insights

Beyond measurement

Big data, climate modelling and a new era of measurement

Meeting Big Data's Biggest Challenges

The data revolution? We've seen nothing yet

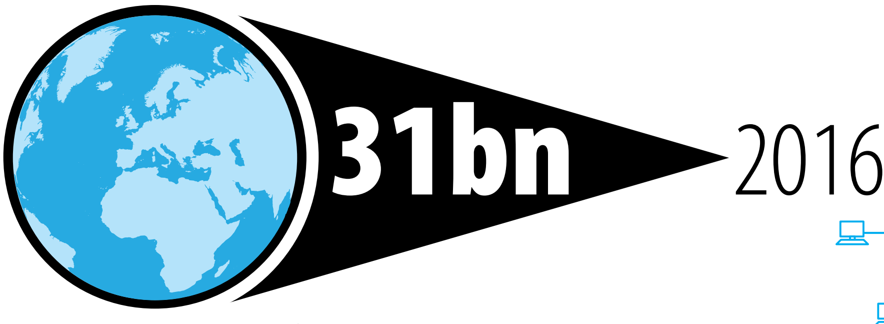
Contents

What is big data?

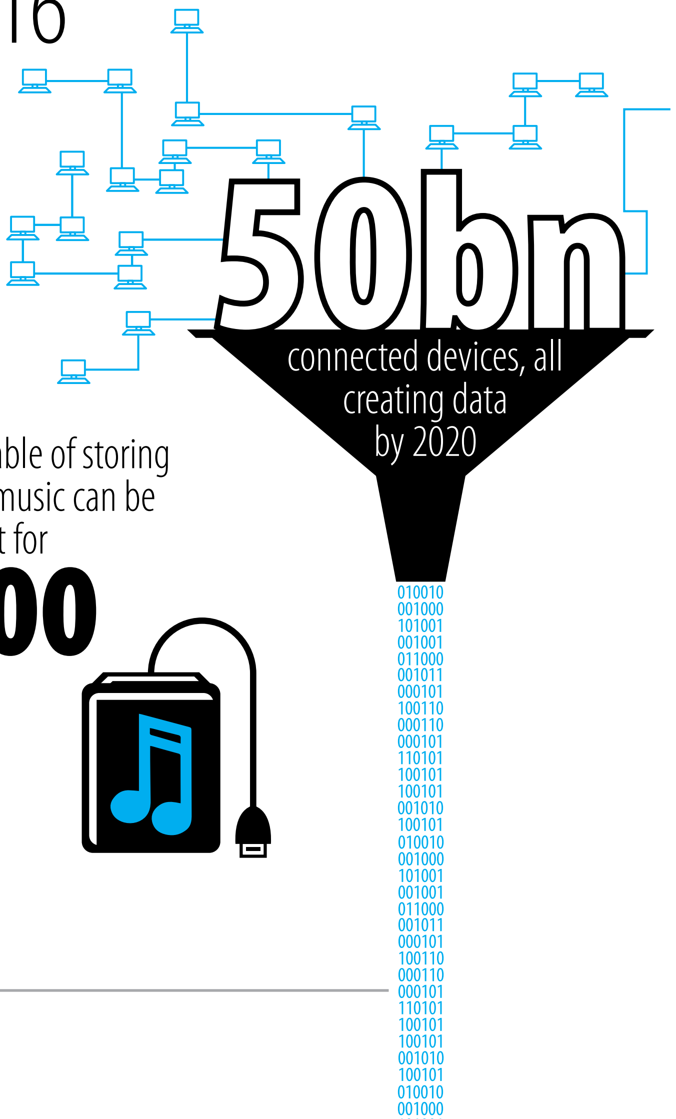
This issue of Insights from the National Physical Laboratory (NPL) is dedicated to big data, but what do we mean by that term?

In these pages you will see big data referred to as both a technology and a solution to various challenges. You will also see it described as a challenge in itself with new approaches and algorithms needed to solve it.

The difference is one of perspective – those in data provenance, computer science and end users will all have different views on what big data is, or could be. It is the role of a National Measurement Institute like NPL to span all of these views, and help partners in each sector to benefit from the potential of big data.

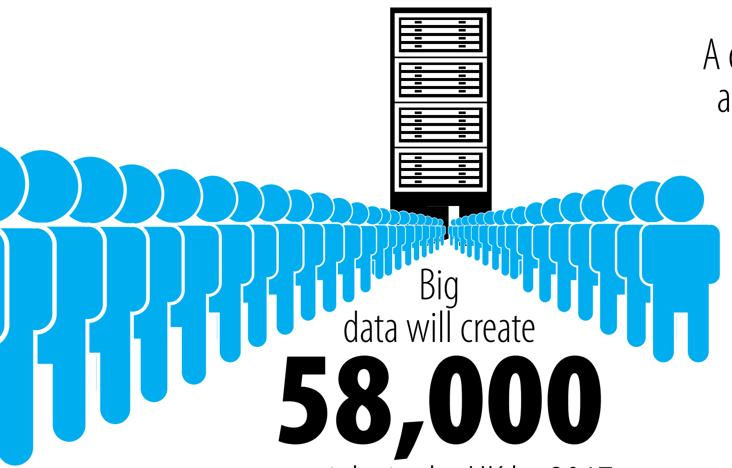


Global market for business data analysis is expected to be £31bn by 2016



A disk drive capable of storing all the world's music can be bought for

$600



Big data will create

58,000

new jobs in the UK by 2017

Foreword

Tim R. Dafforn is the Chief Scientific Adviser at the Department for Business, Innovation & Skills and Professor of Biotechnology at the University of Birmingham

Today we have a lot more information recorded and available to us than ever before, and thanks to the internet it is easier to record, collect and share it. The term 'big data' reflects the possibilities that this explosion of information can provide; it is about what we can learn from access to such huge bodies of data, about how we harness this information to improve our world.

'Big data' is already delivering benefits to the research community, industry, and society as a whole. Let's start by looking at research. Every day scientists at CERN processes one petabyte of information, which is equivalent to more than 200,000 DVDs from the experiments at the Large Hadron Collider (LHC). This data is then distributed across the world via a grid to provide more than 8,000 physicists near real-time access to LHC data helping to accelerate the research process. Big data also promises to help stimulate growth in industry - a 2012 report stated that big data could create 58,000 new jobs over five years, and contribute £216 billion to the UK economy, or 2.3% of GDP, over that time. Harnessing big data is also helping us to speed up the development of new drugs, improve law enforcement and tackle other societal issues.

In the UK, big data has already been classified as government-priority through its status as one of the eight great technologies of the future. Public funding in big data and algorithm research has been increasing through the Research Councils and through projects such as the £42 million investment in the Alan Turing Institute, which brings together leaders in advanced mathematics and computer science.

However, as with any data, we need to be confident of its accuracy. Big Data is defined by seven Vs: Volume, Velocity, Variety, Viscosity, Variability, Veracity and Volatility. The veracity criterion is key here, if we are to make decisions based on Big Data, then we need to ensure that it is unbiased and truly reliable. Therefore it is paramount that the standards are in place to ensure that the potential of big data is fully realised. That is where NPL can help, both by helping to define these new standards and by working with partners in industry and academia who are active in big data, and providing the highly accurate measurement science that can help them to develop solutions to the challenges they face.

How big data can help future proof our cities

Dr Geoff Darch is Head of Climate Change at Atkins and leads the company's Climate Futures team, which undertakes applied research and consultancy on the future impact of climate on infrastructure.

Today, more than half the world's population (3.5 billion people) lives in cities, a figure that is predicted to grow to 60% by 2030 and then 75% by 2050. The majority of this growth (95%) is expected to come from the developing world, often in regions that are dealing with the dual threat of huge urban population growth affecting resources and the impact of climate change. This includes areas such as Bangalore, whose population rose by almost 50% in a decade and Bangkok, whose economy lost $39billion through flooding.

The wider social and economic benefits of this growth in the UK are valued at

6.8 Billion

Alongside this is an explosion in data, providing researchers with an unparalleled wealth of information through sources such as satellites, mobile devices, sensors and social media. The question that policy makers are now asking is what this data can do to assist the dramatic growth of urban areas and the impacts of climate change in the coming decades – to effectively 'future proof' our cities.

Commentators are already analysing the possibilities big data can offer development. In 2012 the World Economic Forum published *Big Data, Big Impact: New Possibilities for International Development*.[1] The report identified four potential ways that big data could support international development, these were: faster outbreak tracking and response; improved understanding of crisis behaviour change; accurate mapping of service needs and the ability to predict demand and supply changes.

Also in 2012, Global Pulse, a UN initiative that uses digital data to support global development and humanitarian action, published *Big Data for Development: Challenges & Opportunities*.[2] This white paper identified the different types of data that was relevant to development work: passively collected transactional data, such as from phones; information such as social media, from which inferences can be made; data gathered from physical sensors like satellites or terrestrial sensors and crowd-sourced information.

Our aim at Atkins is to harness big data and use it to improve our cities, particularly through improving the infrastructure that supports our urban communities, enabling us to be more resilient and responsive to both our changing environment and needs of society. Through this we have developed *Future Proofing Cities*[3] – a report developed in partnership with the Department for International Development and University College London looking at the risks and opportunities for inclusive urban growth in developing countries. From this and our subsequent work including Future Proofing Indian Cities and Future Proofing Lagos we have identified six opportunities for using big data:

1. Quantifying vulnerability to risk. This will assess where communities are based in relation to natural hazards, such as flooding, and predict what adaptive capacity they have to respond such situations.

2. Identifying infrastructure deficits and damage. This can examine where (in relation to demand) there is missing or malfunctioning infrastructure.

3. Mapping city catchments. This could include identifying important locations such as where a city's food sources are.

4. Identifying the movements of people into and across a city, in relation to the types of transport available.

5. Developing climate-related thresholds for infrastructure demand and damage.

6. Understanding how people respond to natural disasters and how they might respond to risk reduction measures.

At Atkins we are working in partnership with specialists to further develop how we utilise big data to improve how we prepare our cities for future challenges. For example, we have recently contributed to the Association of Geographic Information's *2020 Foresight Report*.[4] This identified the critical role that the geospatial community will have to help the infrastructure sector understand, analyse and manage the masses of data available to it. Specifically, the geospatial community's expertise in data analytics can help tackle issues of data quality and security that will improve decision-making and planning.

Another example is the partnership we formed with mobile phone network EE, to create mobile data-enabled products and services that generate unique insights from big data. These products and services have the potential to change how the UK grows and improves its infrastructure by placing customers' needs at the heart of infrastructure decision-making.

Big data can enable us to better understand the metabolism of our cities. It can help us find any gaps in service provision and pinpoint infrastructure investments to the communities and that need it most - where the impact will be greatest. Big data can also help us avoid some issues with traditional data collection such as statistical sampling, cost and security.

However, big data is not a cure-all solution. As the Global Pulse white paper emphasises, there are two main factors that the success of big data in development hinges on:

"One is the level of institutional and financial support from public sector actors, and the willingness of private corporations and academic teams to collaborate with them, including by sharing data and technology and analytical tools. Two is the development and implementation of new norms and ontologies for the responsible use and sharing of Big Data for Development, backed by a new institutional architecture and new types of partnerships."

The swell of information available to us now can appear daunting, but the opportunities it offers in terms of development need to be grasped. By working with recognised experts across the private and public sector and combining big data with more traditional sources of information we can be better informed when answering questions about what our cities will need to be like in the future. There will be challenges along the way, but how we collectively respond to them will determine the shape, comfort and safety of our highly urbanised future communities.

Dr Geoff Darch is Head of Climate Change at Atkins and leads the company's Climate Futures team, which undertakes applied research and consultancy on the future impact of climate on infrastructure. He is a Chartered Scientist with a PhD in climate change impacts, a Guest Lecturer at the Climatic Research Unit, UEA, an Industrial Supervisor at the Centre for Urban Sustainability and Resilience, UCL, and is Chair of CIWEM's Climate Change Network.

[1] http://www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_Briefing_2012.pdf

[2] http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf

[3] http://www.futureproofingcities.com

[4] http://www.agi.org.uk/news/foresight-report

Harnessing unexploited satellite data for information with social and economic benefit

Robert Elliott, Space Business Development Manager at NPL, discusses how information combined from satellites and other sensors can be used for monitoring to improve decision making in a broad range of applications

When we talk about big data in Earth Observation (EO) we're often referring to the countless images that are captured by the satellites pointing at the Earth; images we're all familiar with owing to useful tools like Google Earth. But there's also a variety of data in invisible parts of the spectrum that isn't readily available yet: infrared, microwave or ultraviolet data for instance. Now there is a sharp increase in interest in turning all this data imagery into useful information to help make evidence-based decisions to reduce business operating costs, improve productivity and improve the wellbeing of people.

Such information can be used to show the ripeness of a field of wheat to tell a farmer when to harvest, or the extent of local flooding to aid emergency vehicle response, or the risk of a landslide next to a public road or rail track. The list of possible information products seems almost endless and new ones are being created all the time. Much of this interest has been sparked by investments such as the ESA and EU Copernicus programme, a constellation of satellites giving us a new source of data for the foreseeable future. However, there are challenges in delivering information such as this on demand or in real time, and efforts are ongoing to streamline access to data and create an ecosystem capable of delivering the valuable insights to support evidence based decision making. EO data, often combined with other forms of *in-situ* data has all the tell-tale signs of Big Data; it has Volume, Variety and Velocity. It also has Veracity (included in IBM's definition of Big Data), in other words, the inherent data uncertainty that comes with measurements of any kind.

This uncertainty in EO data was the reason that NPL and the University of Surrey joined forces and established the Global Sensing and Satellite (GloSS) Centre, to demonstrate data quality assurance in a new era of Earth Observation Information services in focused areas including Agriculture, Future Cities and Maritime Surveillance. The objective is to pull together the combined scientific excellence in data, sensors, metrology and satellite technologies and demonstrate the economic and societal benefits of quality-driven Earth observation solutions. This approach offers substantial economic and social gains in many areas:

Transport Infrastructure for Future Cities

Efficient asset management is required to meet increasingly stringent demands on safety and reliable performance of infrastructure. The Transport Infrastructure Management (TIM) Service is planned as a new generation of asset monitoring services that can be deployed across multinational organisations. The concept is based on the integration of location and mapping capabilities of space assets and ground local point sensor networks - delivering data and quality information in the most efficient manner.

A team, led by NPL, assessed the suitability of a 25 year service to cover the area of Greater London bounded by the M25 orbital motorway, including the roads, adjacent land, earthworks and structures. It will deliver a long term continuous monitoring of ground and structural surface displacement in 2D and 3D, in some cases down to centimetre level of information.

The large scale demonstration of the service for the whole of the London area is currently in discussion with the European Space Agency. The project outputs have also been of interest to water companies, rail operators, heritage buildings and insurance companies, with a host of focussed projects on these being delivered in 2016.

New crop information for farmers through sensor and satellite data innovation

The usefulness of Earth Observation data to support farmers and their vehicles is a well understood area and is frequently used in precision crop spraying, automation of large machinery and crop yield estimation. The GloSS centre takes a particular interest in how to generate entirely new types of quality information to farmers that will continue to support farming intensification, against a sector faced with reduced chemical resistance to pests and disease, low profit margins and new regulations banning the use of certain chemicals. Novel in-situ measurement sensors, such as those developed at NPL to measure hydration and sugar content of individual plants, may be used to validate new vegetation indices retrieved from satellite data, providing completely new insights into the status of a crop or support the decision making of automated machinery such as smart irrigation or robotic harvesting. The GloSS centre is also beginning to develop miniature multi- and hyperspectral sensors for use by farmers either on Unmanned Aerial Vehicles (UAVs) or on board low cost satellites, to provide data to aid crop and soil conditions in ways not yet developed.

Surveillance of the Maritime Environment

In 2014, NPL and the University of Surrey were involved in an air-based demonstration of SAR radar technology over the English Channel. A combination of innovation in multispectral imaging, data fusion and calibrated in-situ measurements helped to identify and distinguish oil spills in much finer detail, reducing the cost of surveillance and response to future oil spill hazards and clean-up operations. In future, new operational missions such as Sentinel-1 and NovaSAR, a UK satellite built by Surrey Satellite Technology Ltd, will provide long term operational services to the Maritime Sector such as these based on the demonstration work carried out by NPL and Surrey. The same SAR technology is also being increasingly used for ship detection and tracking, illegal logging, owing to its unique ability to operate day or night and can see through clouds.

The GloSS centre will continue to support the design and demonstration of quality-driven information services to help spark collaboration, innovation and public and private investment for academia, established organisations and value-adding start-up businesses.

©iStock.com/mediaphotos

Big data, climate modelling and a new era of measurement

Jane Burston, Head of Climate and Environment at NPL, discusses the emerging challenges presented by increasingly vast volumes of 'big data' from Earth Observation satellites.

With hundreds of satellites now providing constant data about the Earth at ever higher resolutions, the flow of data from space can now well and truly be defined as "big data". Over its first 50 years of history, Earth Observation delivered perhaps a few petabytes (1,000 terabytes) of data. Now, in the 21st century, we are experiencing a flood of data that will soon be requiring data processing and storage capacity at the exabyte (1 million terabytes) scale.

So much data is great for enabling more accurate long term predictions about the climate, helping to better inform government, industry and the public on the risks of climate change.

But storing, accessing and quality assuring all of this data is a huge challenge. One of Europe's newest satellites, Sentinel-2a, for example, was launched just a few months ago and will deliver over 600 terabytes of free, publically available data about Earth's climate in its first year alone. Scale this up to the hundreds of satellites currently operational and the problem becomes clear.

As the amount and usability of satellite data grows, so does the need to make it as accurate as possible. One of the core themes of NPL's Centre for Carbon Measurement is to lower the uncertainty of climate data all along the supply chain; from calibrating the satellites before they're launched to quality assuring the information delivered to end users.

Space is a harsh environment to do science. It's a vacuum close to absolute zero, has no ozone to protect against radiation from cosmic rays and involves a high-energy launch just to get there. As a result, even the most perfectly ground-calibrated satellite instrument will drift in its accuracy, resulting in uncertainties in the data often too high to confidently detect the long-term changes in climate.

To help overcome this problem, the Centre is working with the Committee on Earth Observation Satellites (CEOS) to establish a network of well-characterised reference sites on the ground to compare with the measurements taken overhead by the satellite.

Whilst these post-launch corrections help, they will always have to make assumptions about the effect of the atmosphere and local factors in their corrections, compromising their effectiveness. To overcome this, the Centre, with a wide consortium of partners, has designed and is leading the proposed satellite mission called TRUTHS (Traceable Radiometry Underpinning Terrestrial- and- Helio Studies). This will, if launched, mimic the laboratory calibration methods automatically in space,

allowing it to take measurements ten times more accurately than currently possible whilst also boosting the accuracy of other satellites.

But the key and truly unique benefit of TRUTHS is that it can transfer its calibration to other satellites by observing common targets and comparing the two measurements to understand how much the relative accuracy has drifted. This provides an opportunity to upgrade the accuracy of the whole optical Earth Observation system, effectively creating a 'standards laboratory in space' with just one satellite. This has great implications for tying together multiple data streams to create an absolute calibration, making the application of big data both easier and more robust.

As the era of 'big data' grows and the need for exabyte levels of storage and processing power becomes a reality, the need for accurate measurement and well-calibrated data will only grow with it. Our overall goal is to make Earth Observation and climate data as accurate as possible, so that the data produced can be applied anywhere and by anyone in the world, helping us understand our environment in unparalleled detail.

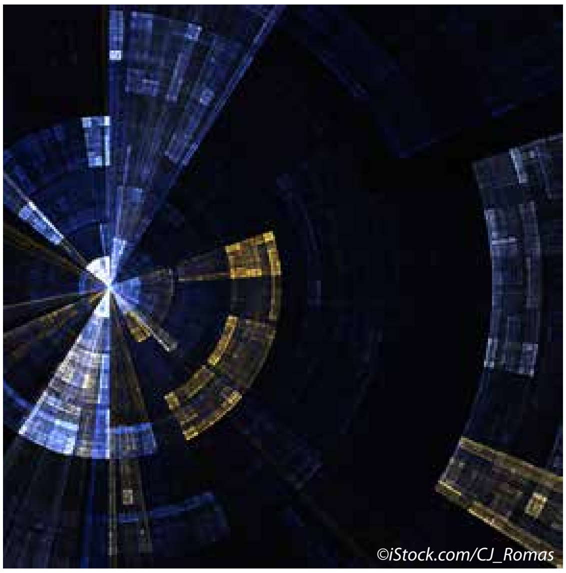
Jane Burston, Head of Climate and Environment, National Physical Lab

Meeting Big Data's Biggest Challenges

Big data can deliver huge benefits, matched only by the magnitude of the computing challenges involved in unleashing its full potential. Adrian Toland, Business Development Manager at the STFC Hartree Centre explains how they are building on the foundation of the Centre's close relationships with organisations like NPL to supply the firepower needed by industry and the research community.

The world is awash with data – an exponential flood of information of incredible scope and scale. For every organisation, the goal must be to ride the tide rather than get submerged by it. That means mining big data effectively, to identify trends and pinpoint new markets, to generate technical insights that accelerate and strengthen development of new products and services, to shed fresh light on behaviours and aspirations, and even to help address grand challenges such as energy security, food security, population migration and climate change.

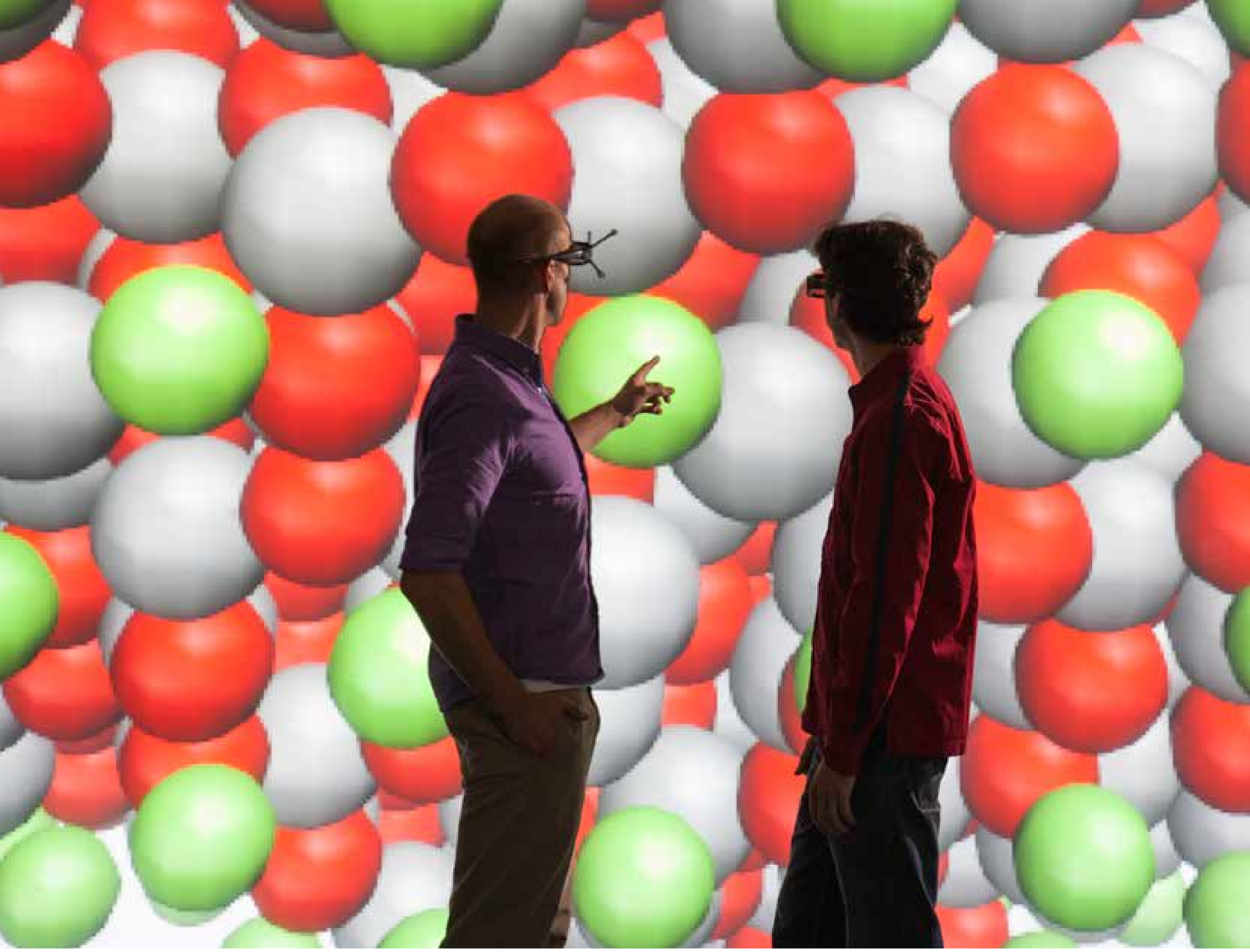
All of which is easier said than done, of course. Big data can be a perilous sea to navigate, largely due to three Vs: volume, variety and velocity. The sheer quantity now available can make it phenomenally hard to seize control and squeeze meaning from data – and it's predicted that the world will be generating an eye-watering 40 zettabytes by 2020. But the remarkable range of data poses a challenge too: internal, external and 'open'; structured, unstructured and semi-structured; at rest or streaming in real time. Meanwhile, the speed



©iStock.com/CJ_Romas

of data flows – from processes, networks, social media, mobile devices and the rest – keeps ratcheting up.

Taming big data, then, presents massive challenges – even before factoring in considerations like data safety, security and privacy. No wonder existing computing platforms and architectures, creaking under the weight, are doomed to become increasingly unfit for purpose. Moreover, for many organisations, it's simply unrealistic to develop and maintain the level of IT capability necessary to pinpoint useful information in data oceans, analyse it fast and efficiently, and then exploit it productively. Yet the price of falling short can be devastating, as they drown in data yet miss vital opportunities and look on helplessly as rivals and competitors successfully unlock the potential.



The Hartree – at the heart of the data revolution

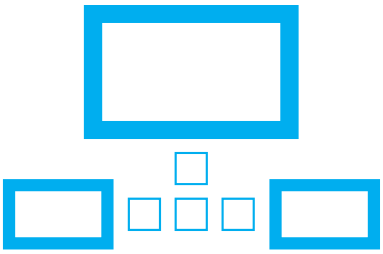
Part of the Science and Technology Facilities Council (STFC) and based at Daresbury Laboratory in Cheshire, the Hartree Centre has a 30 year track record in extending the frontiers of computing science. Backed by multi-million-pound UK Government investment and a strategic collaboration with IT giants IBM, we not only deliver world-leading research but also act as an industrial gateway to game-changing capabilities and facilities spanning high performance computing (HPC), data science, data analytics, and visualisation, simulation and modelling. As a core part of our mission, we offer specialism in big data, providing both businesses and research organisations with benchmark expertise in harnessing big data, in data-centric research and in related, fast-emerging fields such as cognitive computing, where intelligent machines learn from experience like humans do, and the Internet of Things.

Underpinning the Hartree's pedigree in the pioneering development and application of cutting-edge computing capabilities is an array of bilateral and multilateral partnerships typified by our productive, ongoing relationship with NPL. Dovetailing with wider links between NPL and STFC (including a Memorandum of Understanding with the ISIS pulsed neutron and muon source at STFC's Rutherford Appleton Laboratory), this relationship has, for instance, seen NPL and the Hartree join forces with IBM and Edinburgh University to devise a radical approach to materials simulation, combining affordability with a step change in accuracy and cutting the number of physical experiments needed to develop new materials.

The firm foundations that such fruitful collaborations help to provide ensure that the Hartree is well equipped to confront new and evolving challenges. With respect to big data, our relevant inventory of hardware, software and supporting capabilities includes world-class supercomputers and data analysis clusters,

©iStock.com/loveguli

hyperscale data storage, leading-edge code creation and optimisation skills, state of the art intelligent/predictive machines, plus top-end visualisation suites – everything needed to enable our customers and collaborators to identify valuable data streams, drill down into them, pinpoint useful information, link it together and view results accessibly and comprehensibly.



The UK's open data agenda has helped to create over

20,000

publicly available datasets via data.gov.uk, detailing transport use, our health, our society, education, government activity, the economy and the environment

A Question of Application

In the final analysis, such capabilities' value can only be judged by the difference they make – not just in terms of developing platforms and architectures that can meet tomorrow's needs but also in enabling all kinds of organisations to harness big data productively right now. Not surprisingly, more and more organisations across every conceivable sector are recognising big data's potential value in today's increasingly data-driven, knowledge-based economy, and its ability to provide evidence conducive to better decision making and ultimate success.

From helping bankers improve their understanding of financial markets and so more easily predict and prevent market volatility, to providing civil engineers with insights into the long-term behaviour of construction materials in projects around the world – the applications are virtually limitless. Take healthcare and the challenge of protecting humanity against future pandemics. Powerful big data capabilities can scan medical journals, research papers and a torrent of other data sources quickly and accurately – extracting and linking information about everything from laboratory tests, drug trials and vaccine efficacy to the incidence of different flu strains and the spread of animal diseases that may have potential to jump species and threaten mankind.

Here at the Hartree, we are already delivering results of precisely this wide-ranging and invaluable kind. To take just two examples, we have helped healthcare multinational GSK mine complex networks of information about gene-disease relationships to facilitate development of new drugs for incurable illnesses; and we have helped start-up company Democrata harness open data to predict the presence of ancient remains at proposed construction sites – all achieved alongside our ongoing HPC work with partners like Unilever to enable them to bring new, improved products to market more quickly and cost-effectively.

Big data is clearly one of the most valuable weapons available to meet current and emerging challenges, to drive new discoveries and to catalyse business success. As a recognised hotspot of expertise in this field, the Hartree Centre is whole-heartedly committed to ensuring that the potential is unleashed to the greatest possible benefit of the greatest number of people – today, tomorrow and for decades to come.

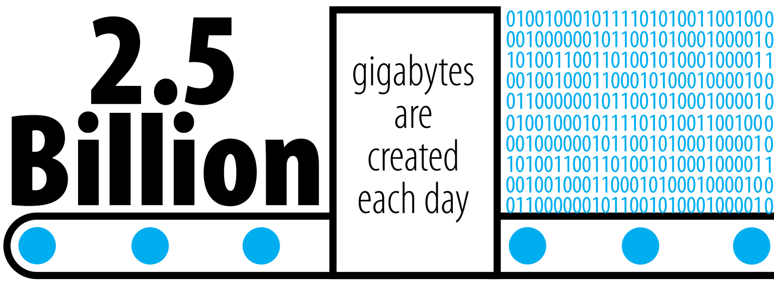
Big data for all: how SMEs can harness high performance computing

George Graham, Commercial Manager at EPCC, outlines how SMEs can make the most of the big data revolution through open source computing.

Industrial and scientific organisations have the ability to easily and rapidly collect and store vast amounts of domain related data. However the real challenge they face is how to generate business value from that data. How can they generate knowledge from data that will ultimately improve discovery and R&D processes; creating new structures and relationships, discovering deeper and more valuable insights, speeding up existing data analysis and generating important correlations, patterns and relationships. Without those derived benefits, large data sets have little or no value.

Deriving value however is not straightforward. Analysing large data sets is a complex activity. There are three dimensions of big data that create this complexity - volume, velocity and variability. Handling very large volumes of data, and indeed the rapid growth in data, presents significant challenges. Further, analytical results must be produced within acceptable time frames. The problem then is one of both Volume and Velocity, or, in other terms, one of both data and compute. For large amounts of data how can we derive efficient data analytics algorithms that can generate value from large data sets in reasonable timeframes? Often this means scaling up executions from lab workstations to High Performance Compute (HPC) platforms. Migrating to HPC platforms may further necessitate the re-engineering of analytical codes such that they can continue to run at scale. So in essence 'Big Data' needs 'Big Compute'. A further complexity is in the Variety of data types collected. Data can be collected from a large variety of sources, both internal and external, as part of R&D and discovery processes. New tools and techniques that can mine data sets consisting of multiple structure types, or indeed fundamentally unstructured raw data sets, then become a necessity.

Solving these issues is easier for some companies than for others. Big data analytics opportunities exist across a number of sectors including energy, oil and gas, life sciences, and financial services for example. And across Europe small to medium sized enterprises (SME) make up a large part of these sectors. Large organisations may have the financial resources to invest in new data and HPC infrastructures, software, tools and expertise.



However unlike large organisations SMEs have limited resources. They generally lack the mathematical, statistical and computational science skills required to take advantage of big data opportunities. A general lack of financial resources hinders adoption of the platforms, tools and skills that are required to generate value from big data.

A particular example of these challenges can be found in the life sciences sector. There are a broad range of big data uses in the sector, for example:

- Chemistry: bio-chemistry, molecular modelling, and protein folding
- Bio-engineering: agricultural engineering
- Genomics and proteomics: next generation sequencing
- Biology: molecular biology
- Pharmacology: pharmacokinetic/ pharmacodynamic (PK/PD) modelling
- Analytics: statistical analysis and bioinformatics
- Energy: biofuels

And of course these uses do not exist in isolation. Frequently workflows need to combine different types of information, such as data from genomics, proteomics, cellular signalling and clinical research. Thus there's the need to build large scale data infrastructures capable of handling multiple heterogeneous data types, integrated with large scale supercomputers and harnessing skills in biology, bioinformatics, statistics, mathematics, machine learning and computational science in order to generate valuable outcomes.

Whilst large organisations have the capability and resources to create these complex workflows, smaller companies do not. For SMEs the cost of acquiring large scale data and HPC infrastructures may be prohibitive and they generally lack the appropriate skills and know how. Whilst significant, these adoption barriers can nevertheless be lowered. We see two main ways of reducing adoption barriers. The first is through technology "servitisation", and the second is through the provision of structural assistance.

Replacing technology adoption through capital outlay, with access to technology through as-a-service models can deliver significant advantage to SMEs. Accessing cloud based compute and data processing services, for example, can be a feasible alternative to acquiring in-house hardware infrastructure. Pay-per use services such as Amazon EMR (Elastic Map Reduce) provide access to large scale data infrastructure and analytical platforms on an on-demand, pay-per-use basis. This can provide significant cost advantage over large capital outlays on in-house infrastructure. Similarly large scale compute can be accessed on a pay-per-use basis. Amazon Elastic Cloud (AWS/EC) and EPCC's Accelerator programme, for example, provide cost effective pay-per-use access to a broad range of HPC platforms.

A second way of lowering adoption barriers is through structural assistance. In the UK SME's can readily benefit from a wide range of publically financed resources. For example Innovate UK, and the locally focused Scottish Enterprise, provide a range of funding vehicles targeted at enabling impact from R&D programmes. In England a range of Catapult centres, such as the Cell Therapy Centre, the Digital Catapult, Medicines Technology Catapult and Precision Medicine Catapult provide ready access to resources, facilities and skills. The Scottish Innovation Centres such as The Data Lab, Digital Health Institute (DHI) and Division of Stratified Medicine provide access to both resources and funding for companies with impactful business cases. Across the UK the Farr Institute is rapidly developing engagement models that can assist both research and industrial users accelerate innovations in life sciences and health care by providing access to valuable patient data.

In summary, big data is complex. However a combination of technology "servitisation" and the existence of a vibrant sector ecosystem built from public-private collaborative partnerships will have a significant impact on accelerating the adoption of big data technologies by UK SMEs.

The data revolution? We've seen nothing yet

Professor John Bancroft, Director, NPL North of England

In its 2016 Visual Networking Index, Cisco predicts an 83 per cent growth in machine to machine (M2M) mobile traffic between now and 2020. Come 2020 such devices, like wearable devices and home monitors, will be generating over 2 million terabytes of information a month[1]. This is excluding all other traffic, mobile or non-mobile, such as smartphones, PCs or tablets. It requires little imagination to work out what direction our world is moving in, as IoT devices become the norm in all facets of life.

This gigantic reservoir of data can be used to solve societal problems, build billion dollar companies or simply make life a little easier by having a fridge that knows when you're out of milk. But these goals are only achievable if we know how accurate the data is.

Computer systems currently can't judge how good a piece of data is. As soon as information passes from the real world to the digital - via a sensor, for example - the system it moves into assumes it is reliable. In the real world that judgement would be made almost without thinking by our brains. Yet the infrastructure (i.e. the brain) to make this call simply doesn't exist in the digital world. We rely on the devices or processes that collect the information to be of sufficient quality, yet we also know that, for the vast majority in common usage, this is unlikely to be the case.

It is worth highlighting that the quality of the data itself is not the issue. Bad information can be just as informative as good. But we can't make that judgment at present, without knowing the situation in which that information was collected. Unless there is appropriate metadata that partners the information, a system that uses it also can't make that judgement and will assume it is reliable.

As data collection in every conceivable avenue of life grows - health tracking apps, pollution sensors, home environment monitors - so too are important decisions about our wellbeing made using this data. Understanding the reliability of this data is therefore crucial to our daily lives.

NPL is the home of metrology. It defines the high quality of standards by which industries, organisations and many things that we use day to day operate. It therefore makes sense for it to also define how the data that controls our world is understood, to put in place a standard protocol by which data and its quality can be marked. Giving the digital world the ability to understand the quality of the data it uses will dramatically increase the effectiveness of how it uses that information.

As the first Director of The Hartree Centre, one of the UK's foremost big data centres, I've seen the benefits that the information revolution has brought to organisations of all shapes and sizes. But this revolution has been tiny compared to what is on the horizon. The benefits of the amount of information we are about to generate could be vast, as long as we have the infrastructure in place to deal with it. This is my goal at NPL; as the UK's centre of excellence for metrology, my mission is to bring this to bear on big data and make sure that our data-centric world is of a benefit to all.

[1] *http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html#Trend_3_Measuring_Mobile_IoE*

National Physical Laboratory

NPL is a world-leading centre for the development and exploitation of measurement science, technology, related standards, and best practice in a diverse range of technical areas and market sectors. As the UK's National Measurement Institute, our capabilities underpin the UK National Measurement System (NMS), ensuring consistency and traceability of measurements in support of UK and overseas customer interests. We aim to provide world-class science and engineering with economic, social and environmental benefits to the UK.

National Physical Laboratory
Hampton Road
Teddington
Middlesex
United Kingdom
TW11 0LW

020 8977 3222
www.npl.co.uk