

Harnessing data:

Data hubs

BETA





BETA version

The content in this section is available in a BETA version and is currently being reviewed.

Please send any comments and suggestions, additional resources and/or case studies that may help to improve or update this content to:

martin.ferguson@socitm.net





Opening-up data

Opening-up data is not just concerned with avoiding unnecessary security constraints and costs. Rather, it involves sharing information openly, internally and externally to improve its quality and value. Most data is not personal, sensitive or confidential, and much data security is unnecessary. It comes from a culture of ‘closed’ not ‘open’, where it is cheaper and easier to provide a blanket of security over all data rather than being more sophisticated in determining which data requires higher levels of security and protection. The culture extends to the workforce in which many feel they don’t need to worry about data security, because IT has sorted it all out – any breach is the fault of a failure of IT protection.

In practice however, UK public services are moving towards greater openness, and have regulatory and legal requirements to do so. The open data initiative began nearly a decade ago, and local government have arguably led the way, finding value in sharing data openly, allowing it to be used in ways beyond the original purpose for which it was captured, opening up data linkage possibilities across disparate organisations and systems.





Opening-up data

Whilst the original government aims of ‘armies of armchair auditors’ keeping public services in check was neither effective nor feasible, there have been a range of other benefits which public service organisations can list in their data strategies:

- Increasing customer and citizen insight
- Undertaking more sophisticated data matching to improve data quality, reduce data redundancy and reduce fraud
- Providing better customer services and anticipation of need or risk
- Improved management of resources, such as in being able to find underused buildings to share
- Better delivery of complex relationship services such as health and social care integration





Opening-up data

Opening-up data can also ensure that many versions of the same information can be exposed, both within an organisation but also across related services in other organisations. This can help remove redundant data and ensure decisions are taken using the most current and up to date data.

It also can fundamentally change the culture of an organisations in relations to data principles and use, moving beyond the traditional silos of data held protectively in departments or by professionals unwilling (sometimes with good reason) to share with the wider organisation, partner or the citizen whose data it is. It also helps to ensure that data and information are managed in their own right, rather than as appendages locked into propriety or bespoke systems with their own unique formats and standards.





Data hubs

The idea of pooling mountains of data and analysing to produce new insights is nothing new. Tesco began doing this with customer shopping habits in the 1980s but did not have the technical capability to analyse and use the data volumes. Today things are different. Powerful processing capabilities coupled with artificial intelligence engines and cloud infrastructure have transformed the possibilities into reality. 'Big Data' projects are being reborn.

But building a massive data warehouse and simply assuming everyone will use it in a uniform way could be a recipe for disaster of optimism over reality. Such a project can become unwieldy, costly and hard to realise value. Big data warehouses also do not always fit well with the model of public and private cloud systems.

What is emerging as a virtual data hub, often operating at a local level around the roles of health and local government, with data held across individual applications using a mixture of, standards, open source data management tools. Defining this is a data architecture is not easy, and CIOs need to collaborate closely with individual business teams to understand analytics agenda and requirements, and often data linkages are more important than creating data mountains in a data project.





Data hubs

Public services today are using predictive data analytics to start to mine data and solve the most intractable problems facing society. For example, Brent Community Protection team and IBM have carried out a proof of concept, building a predictive model to identify vulnerable young people most at risk of criminal exploitation, such as child sexual exploitation and gang drug running.

A growing network of connected devices in our public buildings, in our streets and even on our wrists is helping the public sector to deliver better services that are less intrusive and reduce the need for human intervention. The Cloud and AI are essential components to make sense of this stream of new data, using powerful and distributed cognitive engines to process structured and unstructured data, analysing and linking data according to patterns and connections, forming a big data virtual backbone. These algorithmic routines and logical components such as RPA (robotic process automation) are transforming the prospects of big data analytics for the public sector, predicting events and opportunities in every aspects of our daily lives.

Big data is often seen as the same as master data management – but they are not the same. Neither is the concept of data warehouse:

- “Big data” – typically, a project, defined around a specific function or purpose (such as customer service)
- “Data warehouse” – typically, about bringing all core data together into one large repository shared across the business
- “Master data management” – a concept for how core data is managed; whilst often centralised, some will also be in distributed systems





Data quality challenge – how to avoid ‘big bad data’

There is a need to live with data error – it cannot be eliminated altogether. This means focussing on the biggest risks and working to mitigate the effects of poor data quality , with timely intervention and strong policies and processes. Making responsibilities and data governance clear is the essential starting point to understanding the value of data and preventing preventable errors.

What is Quality Data?

“Quality data is useful data, and the reverse is true”

- Consistency in formats
- Properly maintained
- Timely
- Accurate and true
- Relevant metadata held
- Business rules for use documented
- Provenance known and retained
- GDPR compliant





Data quality challenge – how to avoid ‘big bad data’



Benefits of good data:

- Value increases in data mining activity for problem solving
- Data sharing is more effective and justifiable
- Process automation is only possible with good, clean data
- Customer relationship management systems and access authentication require good data
- Decision making is faster and with more confidence, empowering suppliers staff and customers.

Causes of poor data quality:

- Lack of ownership for the level of quality in data sets
- Specific data roles are not clear (SIRO, DPO, CDO, etc)
- Suppliers are not held accountable
- Sloppy data handling/sharing practices and weak accountability
- Information governance is missing or not directed from the board
- Poor processes for data management or MDM
- Culture does not value data
- Data standards lacking or inconsistently applied
- Risk management for data assets is immature, including security policies
- Digital maturity is low, such as how social media and new multi-cloud systems are used



Data mashups

The power of data lies in the ability to connect different data sets to create new insight. It is the basis of AI, which uses the intelligence in data to make predictive analysis and automated decision making. Today, sophisticated data tools can combine data from many sources - audio, video, RSS feeds and raw numeric or non-numeric data. These 'mashups' of disparate data sources are the basis of data mining tools which can trawl for new intelligence through enormous data volumes, looking for patterns and business value as well as for machine learning applications.

In the past much of this type of analysis depended on data warehousing - a large and carefully designed pool of data. Although there is still a place for selective data warehousing, perhaps in a CRM (customer relationships management system), many applications today do not depend on a single database managed on-premise. Instead, they make calls to multiple data files in a mix of hosted cloud, on-premise and open source data on the internet.

Clearly this depends on :

- Common data formats and definitions that ensure that data schemas permit appropriate data linkages to be made
- Methods for data collection
- Data analysis to establish the provenance and accuracy of data
- Appropriate tools that do the actual mashup.





Data mashups

Data mashup tools are not the sole domain of the information or IT professional. They also allow end-users to define their own data mashups and to visualise and analyse the resulting data blocks. Some will be concerned about this, worried that end-users can misuse data, create incorrect formulae and therefore make the wrong insights and conclusions. Business intelligence (BI) tools can be locked down, but this will arguably hold back the progress and learning necessary in a truly data-driven organisation. In any case, competent end-users will find ways of extracting data into spreadsheets if they choose.

Public service organisations must learn to grow data skills and trust the judgements of data specialists and end-users alike, learning openly when mistakes are inadvertently made. Better to concentrate on ensuring the quality of master/core data and support end-users in becoming intelligent data analysts.





Benefits derived from harnessing data



Discussions with some of those linked to this work have listed a range of benefits generated by their data projects.

- Quicker and better decision-making through the availability of a high-quality easy to retrieve data base
- Improving the response to customer citizens needs
- Developing commercial solutions to increase income generation
- Identifying data patterns, trends and new service opportunities
- Developing efficiency of internal processes such as automation and business process improvement
- Increasing the accuracy of data, especially through data matching techniques
- Better understanding of demand management, forecasting and trends
- Better understanding of citizens, and customers and communications with them
- Lower operational costs
- Improve risk management
- Optimize marketing and communication to citizens and customers
- Being able to simplify and consolidate business data interrogation tools and reporting
- Having a better understanding of the scale and potential benefit of total corporate data assets
- Improving data governance and information management cultures
- The ability to be able to capture, store, retrieve and analyse data in a flexible and agile fashion
- The ability to link data held across different data formats
- Improving the tracking and recording of data origins and movements between databases and organisations – data provenance
- Building more sophisticated analytical and predictive models, including supporting the basis of artificial intelligence development.



Additional resources

ODI resources:

- [Open data essentials eLearning](#)
- Open Data in a Day [training course](#)
- Most of our [tools](#) are designed to support open data publication and use
- Local government specific open data work - New Service Delivery Models [R&D project](#)
- [Checklist: How to design to scale](#) (as part of the [Data and Public Services Toolkit](#))





Additional resources (cont'd)

- Wise Council: Insights from the cutting edge of data-driven local government (Nesta-LGA) <https://www.nesta.org.uk/report/wise-council-insights-from-the-cutting-edge-of-data-driven-local-government/>
- Datavores of local government (Nesta) <https://www.nesta.org.uk/report/datavores-of-local-government/>
- London Office of Technology and Innovation (LOTI) <https://www.londoncouncils.gov.uk/who-we-are/london-office-technology-and-innovation-loti>
- Various resources on harnessing data (Eddie Copeland) <http://eddiecopeland.me/data/>
- Idea on a page (Eddie Copeland) https://docs.google.com/presentation/d/1MoHu7MUUpKUBHDmp9Phmh3dCf9KNKI9P3ROIR4g1yXIQ/edit#slide=id.g40e73117d2_0_0
- What's the ideal model for an Office of Data Analytics? (Eddie Copeland) <https://www.nesta.org.uk/blog/whats-the-ideal-model-for-an-office-of-data-analytics/>
- Big Data in the Big Apple (Policy Exchange) <https://policyexchange.org.uk/publication/big-data-in-the-big-apple-the-lessons-london-can-learn-from-new-yorks-data-driven-approach-to-smart-cities/>





Additional resources (cont'd)

- Big data – what does it mean to do it? (Socitm) <https://socitm.net/download/big-data-what-does-it-mean-to-do-it/>
- Realising the potential of big data (Socitm) <https://socitm.net/download/realising-the-potential-of-big-data-analytics/>
- Big data or big hype (Govtech) <https://www.govtech.com/library/papers/Big-Data-or-Big-Hype-99664.html>
- Digital twins (Deloitte) <https://www2.deloitte.com/us/en/insights/focus/tech-trends/2020/digital-twin-applications-bridging-the-physical-and-digital.html>
- 5 steps to transform your IT organisation with AI (Splunk) https://www.splunk.com/en_us/form/5-steps-to-transform-your-it-organization-with-ai.html





Case studies

- Open data projects (Data Mill North) <https://datamillnorth.org/products/>
- Collaboration projects between the public sector, industry and academia (The Data Lab) <https://www.thedatalab.com/case-studies/>
- Edinburgh open data portal (Edinburgh Council) <https://edinburghopendata.info>
- Barking and Dagenham: using data and technology to improve people's lives (Socitm) <https://socitm.net/download/report-barking-and-dagenham/>
- Mudlarking in the social ecology of cities (Socitm) <https://socitm.net/download/mudlarking-in-the-social-ecology-of-cities-breaking-the-public-policy-impasse-2/>
- Council and fire and rescue authority data (LG Inform) <https://lginform.local.gov.uk>
- Unlocking the value of London's public sector data (LOTI) https://www.linkedin.com/posts/eddiecopeland_data-london-localgov-activity-6604634899624284160-ov0Y
- City Tools: shining a light on the technologies powering London's boroughs (LOTI) <https://www.linkedin.com/pulse/city-tools-shining-light-technologies-powering-londons-eddie-copeland/?trackingId=CqWI90drRqybY9XiOp%2FUQA%3D%3D>
- Worcestershire Office of Data Analytics (WODA) <https://connectedlocalgovernment.tv/post/sharing-births-marriages-deaths-data-to-improve-services-and-reduce-fraud>
- Using data in the service of humanity (DataKind) <https://www.datakind.org/chapters/datakind-uk>
- Vulnerable children (Mastodon C) <https://www.mastodonc.com/case-studies/>





Case studies

- Suffolk - <https://www.ukauthority.com/articles/suffolk-authorities-launch-data-analytics-operation/>
- West Midlands Combined Authority - <https://www.ukauthority.com/articles/west-midlands-to-get-office-for-data-analytics/>
- Barking and Dagenham - Social Progress Index Source (Pye Nyunt) <https://www.youtube.com/watch?v=P9TqISoSqFs>)
- Calderdale (Solace) <http://www.solace.org.uk/knowledge/articles/2019-03-01-open-data-in-calderdale/>
- Community Safety Strategy (Brent LB) <http://democracy.brent.gov.uk/documents/s71358/07a.%20Community%20Safety%20Strategy%20towards%20a%20Safer%20Brent%202018-21.pdf>
- Predicting gang exploitation (The Guardian) <https://www.theguardian.com/society/2018/sep/17/data-on-thousands-of-children-used-to-predict-risk-of-gang-exploitation>
- Predicting child abuse (The Guardian) <https://www.theguardian.com/society/2018/sep/16/councils-use-377000-peoples-data-in-efforts-to-predict-child-abuse>
- Innovation in the public sector (Civtech, Scotland) <https://civtech.atlassian.net/wiki/spaces/CC/overview?mode=global>
- Civic hacking initiative (Code the City, Scotland) <https://codethecity.org>
- New approaches to community pain points (MAVHACK, Australia) <https://www.mavdigital.com/portfolio/mav-hack/>
- [Take your city to the next level by leveraging data](#) (MCE)

