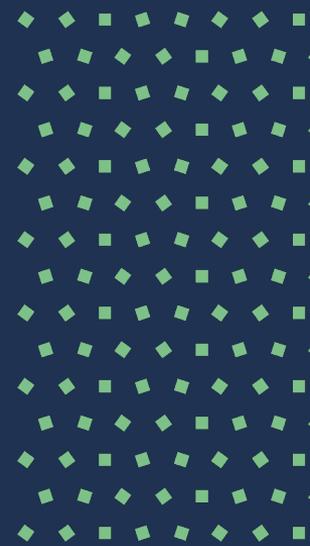


Harnessing data:

Data handling tools

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





Tools, technologies and techniques

Data has become a growing target for software suppliers, helping organisations to manage the growing data volumes, to interrogate, to share and to use the data to construct new insight. Many tools come automatically with ERP suites, such as ORACLE and SAP, or are in-built and connected with office suites such as Google and Microsoft. But this does not include the wide range of data tools designed for a specific purpose (such as customer data and CRM) or sector (such as planning and design), nor the growing range of opensource data analysis tools. Furthermore, this is a fast moving area, with new tools addressing areas such as VR data, AI deployment, cyber, master data management and automated systems testing, IoT data, big data, statistical analyses.

But which tools should a public sector organisation invest in? There are many free tools, and the most sophisticated ‘state of the art’ tools are often eye-wateringly expensive. We don’t give a list here (it would be out of data and contentious before it was read), but we do give insight into the sort of tools available and how they can add value.





Tools, technologies and techniques



Which tools you select depends on:

- Your data maturity position – how ready are you to exploit expensive tools that may require reorganising how you use and analyse data? Do you have the capacity and skills?
- Your budget – how much value can you extract and how much investment are you willing and able to make?
- Where you are on your data journey – You may need to start by using tools to organise and to clean your data, dealing with data quality issues before attempting complex data mashups and business intelligence reporting
- Do you know where your data is? – It may be spread across a range of systems and remote cloud hosting platforms. That might require some data mining and aggregation tools first to get a control of your data assets.
- How you are using data – If you are already in a partnership for data sharing, such as multi-agency working between health and social care, you may need more sophisticated tools that can target data tracking and secure sharing.
- Have you had a data incident? – Maybe a data breach or near miss, paper or electronic. This can point to the need to change culture, behaviours but also the technology you deploy and how it is used.



Tools, technologies and techniques



Some of the better-known tools for large scale data projects (e.g. big data), for example, around in 2019 include:

Statwing Hadoop Quoble HPC Cassandra MongoDB, Apache Storm, CouchDB, Statwing, Flink, Pentaho, Hive, Rapidminer, Cloudera, DataCleaner, Openrefine, Talend, Apache SAMOA, Neo4j, Teradata, Tableau.

Technologies that are increasingly deployed for data analysis work. Such tools include for example:

- Statistical analysis software (e.g. SPSS, SAS)
- Spreadsheets (e.g. Excel) for working with relatively small datasets
- Relational database systems (e.g. MySQL)
- Visualization tools (e.g. Tableau)
- Big data tools (e.g. Hadoop, Hive, Pig, and Impala)
- ML tools (e.g. TensorFlow, Caffe, MxNet, Torch)
- Data archive and retrieval tools



Data tools

There are too many available data management tools on the market to list comprehensively, but a selection of target business applications include:

- Data loss prevention – detecting data ‘leakage’ from all network end points
- Distributed data base tools allowing data and reports to be distributed geographically yet securely controlled for ad hoc query, indexing, and aggregation in real-time.
- Mashup and integration of data from multiple data sources, both structured and unstructured
- Encrypting strong protection tools to share data safely, including data transfer with authentication of sender and recipient
- Data flow monitoring tools supporting complex and dynamic systems to track risk
- Data matching tools to create business insight and how to standardise data formats
- Data testing tools that help with data quality, error detection, ‘dark data’, and reduce the overheads of testing
- Data mining tools that identify and bring to the surface sensitive and personal data helping with GDPR compliance and good data husbandry
- Role-based security tools to support the implementation of automated access, based on responsibilities
- AI investigative tools that support audit trails and journal analysis to ensure traceable decisions were taken by people or machines
- Fraud detection and data frantic tools, both general and specialist
- Data archiving tools that help to ensure low-cost storage, and safe and secure archiving that is still retrievable
- GDPR compliance testing tools, identification and matching of personal data
- Data risk analysis tools, based on the usage patterns and sharing history
- Data interrogation and reporting tools for performance management and business delivery, for both specialists and generalists in the organisation
- Data maintenance tools, from web updating to simple records maintenance with automated checks and data linkages for validation and authenticated change control





Data tools

Managing unstructured data

Most public organisations know how to handle structured data. It may be held in a range of spreadsheets and systems across the organisation, but it can be organised because it has a structured format. Unstructured data however is more challenging. It's embedded in disparate sources such as webpages, emails, documents, social media and more. It's often freeform and in text (rather than numbers), making it difficult for traditional IT systems to analyse it. So its value goes untapped.

Today, most organisations have far more unstructured data than structured data, so it is a problem to be tackled in a data strategy. A human can do this manually, usually by starting with the business problems to be solved, reviewing the various relevant data sources and then extracting and formatting the unstructured data so that it is useable and validated. But there are now a variety of tools available that can help, and do so much faster, typically based on AI technology and, for example, Natural Language Processing (NLP).

“58% of councils are using business intelligence tools, and over 70% are prioritising data analytics for service improvement, with nearly half planning to invest in new tools”

Computer weekly 2018

55% of an organization's data is “dark” according to new global research for Splunk by TRUE Global Intelligence.





Additional resources

- Data Handling Guidelines (NLAWARP) <https://www.nlawarp.net/wp-content/uploads/2018/04/GDPR-V5-Final.pdf>
- The state of dark data (Splunk) https://www.splunk.com/en_us/form/the-state-of-dark-data.html

