

# DATA ANALYTICS



## - Preparing and exploiting data for value

### Purpose

This Data Analytics (DA) position paper presents an overview in summary form of key literature (academic and practitioner) that has informed the shape and scope of the Data Analytics (DA) Critical Capability (CC) of IT-CMF in the digital business context.

### Relevance and Importance of Data Analytics

A data analytics capability provides an organization with the tools, technologies, processes, and approaches to examine its data and help it understand what is happening within the organization and its business ecosystem. Operations, quality, marketing, finance, logistics, engineering, and product or service design teams all benefit from understanding what is happening in terms of resource usage and expected levels of activity regarding each product or service in each market segment that is supplied by different parts of the organization [1]. Data should form the basis for routine business decisions and the organization should select key metrics to drive its operations effectively. Key performance indicators that reflect alignment or progress towards the organization's strategy are essential to maintain a long-term focus and direction.

Data analytics supports identifying correlations between data, and the corresponding meta-data, to assist in problem solving and the identification of what may be causing any issues. For example, pattern recognition techniques can be a very powerful tool in managing both inventory levels and production/service capacity. Further, statistical process control methods are very useful in manufacturing [2] to ensure that tolerances are always met. For example, control limits are often combined with drift analysis to identify a machine or process that is potentially out of control before control limits are breached. The monitoring of process variance and methods like six-sigma [3], [4] can help operations get to a near zero level of defects or errors.

Combining operational data with long-term strategic objectives and goals provides a knowledge base upon which long term planning can be completed. Data analytics can be applied to understand current and potential future risks. Concepts like predictive analytics [5]–[7] that can be built using combinations or correlations, pattern recognitions, and probabilistic approaches provide benefits in problem solving and planning. A Harvard Business Review Quick Pulse survey found that 58% of organizations experience better customer retention and loyalty where tailored real-time analytics are used [8]. Some 80% of companies expect that real-time customer analytics will be important in the years to 2020. The ability to increase customer spend through combining internal data with external sources such as total market size, census data, weather, transport systems data, and so forth can provide context and understanding not achievable using internal data alone.

The survival of any trading organization is based on supporting its customers profitably and in developing new products, services, and markets for future exploitation by its customers. Data analytics can help organizations to uncover insights. These insights potentially help identify new business opportunities and may lead to more efficient operations, higher profits, and more satisfied customers. Focused analytics data reviews and discussions by cross functional teams can enhance ideation as part of a structured innovative process.

## Historical Developments in Data Analytics

The Shu Jing [9] [10] (The Book of Documents circa 400 [BCE](#)) relates that:

”The legendary Emperor Yao (circa 2,000 BCE) commanded Xi and He, in reverent accordance with their observation of the wide heavens, to calculate and delineate the movements and appearances of the sun, the moon, the stars, and the zodiacal spaces, and to deliver respectfully the seasons to the people”.

Emperor Yao essentially requests that the positions of stars, planets, and the moon be observed, and data be collected. The data should be analysed with a view to clearly informing ordinary people of a way to tell the time of the year based on the position of these stellar and solar system objects. The seasons *were to be delivered respectfully to the people*, implying that the results were to be in a language and form that they could understand. Thus, the gathering of data for analysis to solve problems has been practiced for at least four thousand years. It took the United States seven years to tabulate its 1880 census data. The use of tabulating machines in the 1890 census reduced the calculation effort to 18 months at a fraction of the cost [11].

Computer files, databases (hierarchical and relational), data warehouses, and spreadsheets are all widely used today to sort, pivot, tabulate, and process data for analytical purposes. Specialized and general statistical packages offer an array of possible solutions to our data analytical computational needs. Semi structured data can be parsed and processed in parallel with structured data using appropriate parsing techniques. Unstructured data like text can be analysed using appropriate tools. Word processors today check spelling, punctuation, and grammar style, as it relates to the document being drafted, and analyse headers and footers for a variety of media and media sizes. Some word processors will complete this analysis in real-time as the document is typed and auto-corrections can be applied either automatically or interactively. Statistical process control systems can also be considered as real-time analytics. Analytical tools can be used with audio and visual data for many purposes (e.g. forensics, recognition, or medical diagnosis).

Turnkey solutions are offered by many solutions vendors today that have incorporated analytics appropriate to the solutions domain. Examples of common solutions with integrated analytics are Customer/Stakeholder Relationship Management (CRM) and issue tracking systems. Sophisticated factory image scanning equipment or medical image processing equipment is also often supplied with analytics capabilities that are essential features of the equipment and solutions provided.

## Key Data Analytics Insights

Much of the research conducted in recent years in information management and data analytics has concentrated on big data, as it has been associated with the operations management concepts of volume, variety, and velocity. Big data solutions fragmented the analytics market and allowed many new entrants to develop markets around solutions to the volume, variety, and velocity problems that stretched the limits of existing solutions [5], [12]–[20].

Structured Query Language (SQL) JOIN clauses, for example, can be demanding on IT resources and the way they are written can have a considerable impact on performance. The potential to cause problems in production systems is amplified when dealing with high volumes of data or highly complex data that necessitates multiple JOINS [21], [22]. Thus, the complexity of data in some cases has led to unique solutions and techniques.

Technical specialist areas supported by AI and artificial vision have emerged. Microsoft and Google have invested heavily, for example, in automating retina scan analysis for routine diabetes monitoring [23]. A small kit takes the scan at home, and the user uploads it to an analysis server with accuracy rates of 95%, which exceed that of humans who have an 87% accuracy rate. Data analytics for diagnostic purposes is not new. Visual inspections are possible with artificial vision that can extend the visible spectrum to include infrared and ultraviolet [24]–[26]. Thus, we can now capture data and information that humans cannot perceive, and the analysis of these data sets can be automated to provide actionable insights.

Our ability to analyse sound has improved exponentially. Today we can analyse music and produce instrument specific scores from that analysis (albeit limited at this time 2018). We can also identify individual voices, with a high level of accuracy, by comparing audio recordings. The results of sound analysis have been accepted as evidential in many jurisdictions.

Modern robotics are touch aware. Today humans can work in cooperation with robotics safely, without the need for cages and protective shielding [27]–[30].

Thus, today's transducers allow us to digitize and make available to machines, sight, hearing, and touch. The senses of taste and smell have a considerable way to go. However, research funding around smell suggests that artificial noses that are far more sensitive than humans or even dogs will be available sooner rather than later. Sensor saturation issues can be addressed using the inhale, sample, and exhale to flush sensors for the next sample [31].

Thus, digitization is fuelling and enabling the collection and analysis of even more data types through new transducers. A similar revolution is occurring with actuators enabling the remote or automated control of machines and environment controls in our buildings, homes, and vehicles.

Marketing departments were early users of analytics in organizations. Profiled sales data allows an organization to analyse its sales penetrations by gender, age group, product, region, and so forth.

Combining these internal data sets with census data can be powerful in aiding analytics and understanding. These activities are enabled by having the correct meta data to facilitate the analysis. Product or service portfolio managers analyse the features in products or services, and the demand for those features and their cost relative to what customers are willing to pay. The optimum mix of features in each product or service can be determined from such analysis. Analytics aid in understanding the demographics of one's customer base and the effectiveness of the products and services being offered to customers. These are examples of powerful insights that an organization can attain from its data sets and readily available external data sets. Identifying correlations between data sets is a powerful aid to understanding and may help with diagnosis. Experimentation should be encouraged to determine causations. Pattern recognition can help with prediction where the patterns are recurring.

To summarize, data analytics aids understanding of current activities, assists with problem solving, and enables prediction to plan future activities.

## Managing Data Analytics

### Building a data analytics capability

To develop a data analytics capability an organization should develop a strategy for its analytics function. Management must provide a supportive organization structure and enabling governance criteria in which the capability can be developed. For an organization to benefit, its data analytics activity must lead to useful exploitation of analytical data. Developing a culture that values data and analytical insights in decision-making and problem solving is therefore essential. Successful data analytics needs to be budgeted for and adequately funded. With these foundational steps, an organization will set the direction for analytics and enable it with an appropriate organization structure, appropriate governance, and adequate funding. A data driven culture will ensure that the data analytics capability will grow and flourish.

In developing any capability, people, processes, and technologies need to be considered. The myriad of technologies available makes identifying the correct mix of solutions both problematic and complex. Before making any long-term commitments on resources and funding, an organization should raise the awareness of its people/decision makers to an appropriate level of competence. The use of reputable consultants and trusted vendors may help accelerate this process. Training should be provided, and employees should be afforded the time to learn. Data analytics experimentation should be encouraged with an understanding that some experiments will fail. Gartner [32] expects that the current shortage of data scientists will be short lived. This is based on the premise that organizations are more likely to use vendor supplied domain specific analytics features. These vendor-supplied features will address the essential analytics needs for business management purposes and reduce the demand for specific expertise. In terms of processes, analytics activities can consume high levels of compute and IT resources and consequently the business operations need to be effectively protected. A data

architecture and the provisioning of data for analytics may be the best way to enable analytics while protecting the business operations. Ideally, the chosen processes should be lean and agile.

## Data analytics success

Successful data analytics can be identified as follows: by the value delivered, by the data analytics capability available to the organization, and by evidence of continuous improvement in the provisioning of a data analytics service. Evidence of each of these three success factors is outlined below.

### Data Analytics Value

- The organization repeatedly achieves significant data analytical goals and objectives that are bias free and deliver significant value.
- Data analytics costs (including data procurement and provisioning costs) are low and data analytics tooling and resource utilization is high.
- Business operations are not adversely affected by data analytics activities.

### Data Analytics Capability

- The organization can analyse all of the required data types (e.g. unstructured, semi-structured, and structured) using appropriate meta-data and methods, when and where needed, in a cost-effective manner.
- The capability can scale up or down with minimal impact on costs. Data analytics variable costs are in line with the macro-economic numbers of the organization.
- All individuals and business units have a clear understanding of the appropriate metrics and the insights gained from data analytics, as they pertain to their department and their individual contribution to the organization.

### Data Analytics Improvement

- A programme of continuous improvement leverages the latest research, both academic and industry based, vendor advocacy, professional recommendations, and stakeholder feedback from across the business ecosystem.
- The organization's reputation is enhanced by its effective use of data and analytical insights in decision-making.
- A data-driven culture is present throughout the organization, where decisions are based on facts rather than opinions.

## Avoiding data analytics failures

Research [33], [34] has identified recurring reasons for data analytics failures:

<b>Problem</b>	<b>Cause</b>	<b>Preventative actions</b>
Poor choices for processes, technologies, or analytical methods are made.	Low levels of data analytics maturity.	Raise awareness to a sufficient level before long term choices are made and/or use reliable consultants or trusted vendors to advise on the available options.
Cultural resistance to data-based decision-making and the necessary sharing of data.	Difficulties influencing cultural change.	Set a clear strategy for data analytics and base the data analytics organization on collaboration, cooperation, and problem solving with an appointed data analytics leader.
The capability may be limited in terms of complexity or volume of assignments.	Shortage of skills.	Develop the organization's own people and use consultants to fill urgent gaps to aid learning initiatives and increase capacity.
The analytics function lacks direction and focus.	Unclear business value or objectives not defined.	The management team must define goals and objectives for the analytics function having informed itself of the possibilities that analytics can deliver.
Analytics results are in obvious disagreement with other reports.	Poor quality data or data taken from different time or geographical regions are misinterpreted.	Establish cross-functional teams to ensure that data quality standards are set and maintained. Educate users on various views of data and what can and should be compared and what cannot be compared.
Costly data cleansing exercises are not effective.	Root causes of data quality issues are not being identified and corrected.	Work with all data processors to ensure data quality checks are pushed as close as is practical to the point of capture or entry. Encourage a 'get it right first time, and every time' culture.
People disengage from analytics data gathering or data analytics results usage.	Ethics and legal compliance issues.	Get ethics and legal compliance right. Ethics and legal compliance on analytics are important to maintaining the organization's reputation as people will disengage from activities in which they feel uncomfortable.

Organizational culture can accelerate the application of analytics, amplify its power, and steer companies away from risky outcomes [35].

## Overview of the Data Analytics Capability in IT-CMF

The IT-CMF [36] provides a comprehensive set of Critical Capabilities (CCs) to support the management of information technology for business value realization. A number of these capabilities relate to the concept of data management, including Enterprise Architecture Management (EAM), Service Provisioning (SRP), Technical Infrastructure Management (TIM), Enterprise Information Management (EIM), Information Security Management (ISM), and Personal Data Protection (PDP). The Data Analytics (DA) CC is comprised of five capability categories focused on people, technology, process, foundational activities to create the capability, and executing or operating activities once the capability is established. Each category has a series of related capability building blocks. It is closely related to the above-mentioned capabilities and must ensure that these capabilities are made aware of the activities, plans, and objectives of data analytics. The DA capability in turn must be guided by:

- The architectural guidance and implementation roadmaps from Enterprise Architecture Management (EAM)
- The current and future services offered on the service catalogue, which is managed by Service Provisioning (SRP)
- Infrastructure design capabilities and limitations for computing, networks, storage, and security as deployed by Technical Infrastructure Management (TIM)
- The access criteria and controls, structure, relationships, and meta-data available to support data analytics in Enterprise Information Management (EIM)
- Security classifications and related security requirements including anonymization, encryption, retention, and disposal guidance as managed by Information Security Management (ISM)
- Policies, rules, and guidance on authorized uses and appropriate management (including privacy impact analysis requirements and separation of roles) of personal and 'personal sensitive' data in Personal Data Protection (PDP).

## Conclusions

Data analytics is rapidly becoming a competitive necessity, and real-time analytics is becoming increasingly important, particularly in stakeholder relationship management. Developing a data analytics capability takes time, and so it should be comprehensively planned, be executed as a part of the overall business strategy, and its use should be championed by highlighting and promoting any early successes. The data analytics domain is changing at an ever-increasing pace, and so the required expertise may need to be acquired and kept up-to-date. Finally, a programme of continuous improvement in the management of data analytics is crucial for its long-term effectiveness.

## Research Methods

A structured search and review of the academic, vendor, and industry literature was conducted. The literature search focused on the following: establishing a data analytics capability, operating a data



analytics capability, the influence of big data and the internet of things, pointers to the causes of success or failure of data analytics, and caveats or issues demanding care in the use of data analytics. As the material was developed, iterative feedback from data analytics users and external academic reviewers guided the shaping and refinement of the completed artifacts.

## References

- [1] M. McDonald and H. Wilson, *Marketing plans - How to prepare them, how to use them*, 8th ed. The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, U.K.: John Wiley & Sons, Ltd, 2016.
- [2] N. Slack, A. Brandon-Jones, and R. Johnston, *Operations management*, 7th ed. Harlow CM20 2JE, U.K.: Pearson Education Ltd, 2013.
- [3] T. Pyzdek and P.A. Keller, *The six sigma handbook*. McGraw-Hill Education New York, 2014.
- [4] A.M. Joglekar, *Statistical methods for six sigma in R&D and manufacturing*. Hoboken, NJ, U.S.A.: John Wiley and Sons Inc., 2003.
- [5] G. Wang, A. Gunasekaran, E.W.T. Ngai, and T. Papadopoulos, 'Big data analytics in logistics and supply chain management: Certain investigations for research and applications', *Int. J. Prod. Econ.*, vol. 176, pp. 98–110, Jun. 2016.
- [6] T. Schoenherr and C. Speier-Pero, 'Data science, predictive analytics, and big data in supply chain management: Current state and future potential', *J. Bus. Logist.*, vol. 36, no. 1, pp. 120–132, 2015.
- [7] B. Marr, *Big data: Using SMART big data, analytics and metrics to make better decisions and improve performance*. John Wiley & Sons, 2015.
- [8] W. Raj, M. Malik, and A. Kanioura, 'Real-time analytics - The key to unlocking customer insights and driving the customer experience', *Harvard Business Review*, MC208640518, 2018.
- [9] Confucius, *Shu Jing - book of documents*. Richardson-Prachai Solutions Ltd, 34 Egertom Gate, Milton Keynes, MK5 7HH, U.K.: Jiahu Books, 2013.
- [10] C. Rovelli, *The first scientist - Anaximander and his legacy*. 904 Edgewood Road, Yardley, Pennsylvania 19067: Westholme Publishing, LLC, 2011.
- [11] M. Miyazaki, 'A brief history of data analytics', FlyData, 2015.
- [12] Y. Wang, L. Kung, and T.A. Byrd, 'Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations', *Technol. Forecast. Soc. Change*, vol. 126, pp. 3–13, Jan. 2018.
- [13] D. Stodder, 'TDWI checklist report: Strategies for improving big data quality for BI and analytics - Applying data quality methods and technologies to big data challenges that fit your organization's objectives', TDWI, Checklist Report, 2018.



- [14] S. Wang, J. Wan, D. Zhang, D. Li, and C. Zhang, 'Towards smart factory for industry 4.0: a self-organized multi-agent system with big data based feedback and coordination', *Comput. Netw.*, vol. 101, pp. 158–168, Jun. 2016.
- [15] M.M. Rathore, A. Ahmad, A. Paul, and S. Rho, 'Urban planning and building smart cities based on the Internet of Things using big data analytics', *Comput. Netw.*, vol. 101, pp. 63–80, Jun. 2016.
- [16] M. Salehan and D.J. Kim, 'Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics', *Decis. Support Syst.*, vol. 81, pp. 30–40, Jan. 2016.
- [17] Y. Sun, H. Song, A.J. Jara, and R. Bie, 'Internet of Things and big data analytics for smart and connected communities', *IEEE Access*, vol. 4, pp. 766–773, 2016.
- [18] J.S. Rumsfeld, K.E. Joynt, and T.M. Maddox, 'Big data analytics to improve cardiovascular care: promise and challenges', *Nat. Rev. Cardiol.*, vol. 13, no. 6, p.350, 2016.
- [19] D. Singh and C.K. Reddy, 'A survey on platforms for big data analytics', *J. Big Data*, vol. 2, no. 1, Dec. 2015.
- [20] M.M. Najafabadi, F. Villanustre, T.M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, 'Deep learning applications and challenges in big data analytics', *J. Big Data*, vol. 2, no. 1, Dec. 2015.
- [21] Sisense, '8 Ways to fine-tune your SQL queries (for production databases)'. Sisense, Jun. 2017.
- [22] K. Ousterhout, R. Rasti, S. Ratnasamy, S. Shenker, and B.G. Chun, 'Making sense of performance in data analytics frameworks', p.16, 2015.
- [23] M. Boran, 'AI retinal scanning offers vision of the future for care of diabetics', *The Irish Times*, Dublin, 16 Aug. 2018.
- [24] I. Yusuf, T. Fung, S.N. Peirson, and C.K. Patel, 'Artificial iris implants for aniridia and oculocutaneous albinism: Does near infra-red light transmission risk retinal phototoxicity?', *Invest. Ophthalmol. Vis. Sci.*, vol. 55, no. 13, pp. 3786–3786, 2014.
- [25] C.S. Nandi, B. Tudu, and C. Koley, 'Computer vision based mango fruit grading system', presented at the International Conference on Innovative Engineering Technologies (ICIET'2014), 2014, pp. 1–5.
- [26] Q. Wu, L. Xie, and H. Xu, 'Determination of toxigenic fungi and aflatoxins in nuts and dried fruits using imaging and spectroscopic techniques', *Food Chem.*, 2018.
- [27] K. MacLean, 'Taking haptic design from research to practice', Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces, 2017, pp. 1–11.
- [28] J. Banks, 'The human touch: practical and ethical implications of putting AI and robotics to work for patients', *IEEE Pulse*, vol. 9, no. 3, pp. 15–18, 2018.
- [29] J. O'Neill, J. Lu, R. Dockter, and T. Kowalewski, 'Practical, stretchable smart skin sensors for contact-aware robots in safe and collaborative interactions', IEEE International Conference on Robotics and Automation (ICRA), 2015, pp. 624–629.

- [30] U.B. Rongala, A. Mazzone, D. Camboni, M.C. Carrozza, and C.M. Oddo, 'Neuromorphic artificial sense of touch: bridging robotics and neuroscience', in *Robotics Research*, Springer, 2018, pp. 617–630.
- [31] B.L. Villarreal and J.L. Gordillo, 'Bioinspired smell sensor: Nostril model and design', *IEEEASME Trans. Mechatron.*, vol. 21, no. 2, Apr. 2016.
- [32] G. Herschel *et al.*, 'Predicts 2018: analytics and BI strategy', Gartner, G00341404, Mar. 2018.
- [33] O. Fleming, T. Fountaine, N. Henke, and T. Saleh, 'Ten red flags signaling your analytics program will fail', McKinsey & Company, McKinsey Quarterly, May. 2018.
- [34] J. Heizenberg and A.D. Duncan, '5 pitfalls to avoid when designing an effective data and analysis organization', Gartner, G00367016, Jul. 2018.
- [35] A. Díaz, K. Rowshankish, and T. Saleh, 'Data culture where organization meets analytics', McKinsey & Company, Interview McKinsey Quarterly, Sep. 2018.
- [36] M. Curley *et al.*, *IT Capability Maturity Framework (IT-CMF) - The body of knowledge guide*, 2nd ed. Netherlands: Van Haren Publishing, Zaltbommel, [www.vanharen.net](http://www.vanharen.net), 2016.

## Contributing Author

Conor O'Brien, Researcher, Innovation Value Institute.

Gerard Conway, Researcher, Innovation Value Institute

## About IVI

The Innovation Value Institute (IVI) is a multi-disciplinary research and education establishment co-founded by Maynooth University and Intel Corporation. IVI researches and develops management frameworks to assist business and IT executives deliver digitally enabled business innovation. IVI is supported by a global consortium of likeminded peers drawn from a community of public and private sector organizations, academia, analysts, professional associations, independent software vendors, and professional services organizations. Together, this consortium promotes an open ecosystem of research, education, advisory support, international networking, and communities-of-practice. IVI is supported through Enterprise Ireland's and IDA's Technology Centre programme.

## Contact IVI

For more information on this capability, IT-CMF and other IT management topics, or on becoming a member of IVI's international research consortium, please visit [www.ivi.ie](http://www.ivi.ie) or contact us at: [ivi@nuim.ie](mailto:ivi@nuim.ie) or +353 (0)1 708 6931.



Innovation Value Institute, IVI, IT Capability Maturity Framework, and IT-CMF are trademarks of the Innovation Value Institute. Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this document, and the Institute was aware of a trademark claim, the designations have been printed with initial capital letters or all in capital letters.