

Three Lessons From 100 Years of Data Management

Digital data management is 60 years old,¹ yet only 3 percent of organizational data meets basic quality standards,² and three-quarters of employees mistrust such data.³ Furthermore, more than 70 percent of employees have access to data they should not have,⁴ representing the potential for security breaches of catastrophic proportions. In many cases, no one is checking whether data are fit for the intended purpose, creating data-based misrepresentations that could compromise an enterprise's ability to make the right decisions or meet its objectives or could lead to regulatory penalties for using data the enterprise had no legal right to use.

Low levels of public trust in data practices are a global phenomenon caused in part by high-profile failures to protect personal data from abuse, giving rise to calls for changes in data-driven systems and the structures that enable them.⁵ One high-profile private-sector example of privacy abuse is the Facebook–Cambridge Analytica scandal, wherein the data of tens of millions of users were abused. Facebook agreed to settle the ensuing lawsuit, thereby avoiding the need to answer questions about its alleged cover-up of the data breach.⁶

The public sector has not been spared from data-driven mistrust. Only 49 percent of Britons trust the government to store their data. Similar levels of distrust have been identified in Australia and the United States, driven by headlines reporting the loss (security issues), misuse (privacy issues) and inaccuracy (data quality issues) of personal data.⁷ A survey in Japan found that 80 percent of people lost trust in the government's economic indicators after faulty wage data were released. New Zealanders' trust was eroded due to the low response rate to the 2018 census.⁸ These problems are exacerbated in the public sector because trust between citizen and state is more complex than trust between customer and retailer. Most people in the United States believe the government withholds important information it could safely release to the public.⁹ In addition, when publicly funded institutions do not make data available to the

public, it feeds distrust.¹⁰ If information is provided to the public in an effort to rebuild trust, it is important that the information be both useful to and usable by citizens, as in cases in Brazil, Sri Lanka and Uruguay.¹¹

Misinformation, disinformation and conspiracy theories are products of distrust, and badly behaving institutions enhance the attraction of these products to segments of the population that feel unrepresented.¹² The increase in populism is also an outcome of the decline in trust in the public sector.¹³ Ninety-three percent of enterprises believe that collecting, managing, storing and analyzing data must improve,¹⁴ confirming that major efforts are needed to address the global state of data management.



GUY PEARCE | CGEIT, CDPSE

Has an academic background in computer science and commerce and has served in strategic leadership, IT governance and enterprise governance capacities. He has been active in digital transformation since 1999, focusing on the people and process integration of emerging technology into organizations to ensure effective adoption. Pearce maintains a deep interest in data and their disciplines that accelerated with the launch of his high school data start-up many years ago. He was awarded the 2019 ISACA® Michael Cangemi Best Book/Author award for contributions to IT governance, and he consults in digital transformation, data and IT.

Recent Data Management History

To know where data management should be headed, it is important to understand its past. The first human use of data occurred 20,000 years ago;¹⁵ but the focus here is on the generations of data environments and data problems starting in the late 19th century (**figure 1**).

The role of the database remained largely stable between 1890 and 1960.¹⁶ However, increasingly complex data management challenges have been weakly addressed over time, causing an expanding gap between data management and data technology across data management generations. A data management generation is a combination of data technology and data management paradigm shifts. Although trust is a function of competence (capability and reliability) and intent (humanity and transparency),¹⁷ only data management competence is considered here.

First Generation: Paper Records

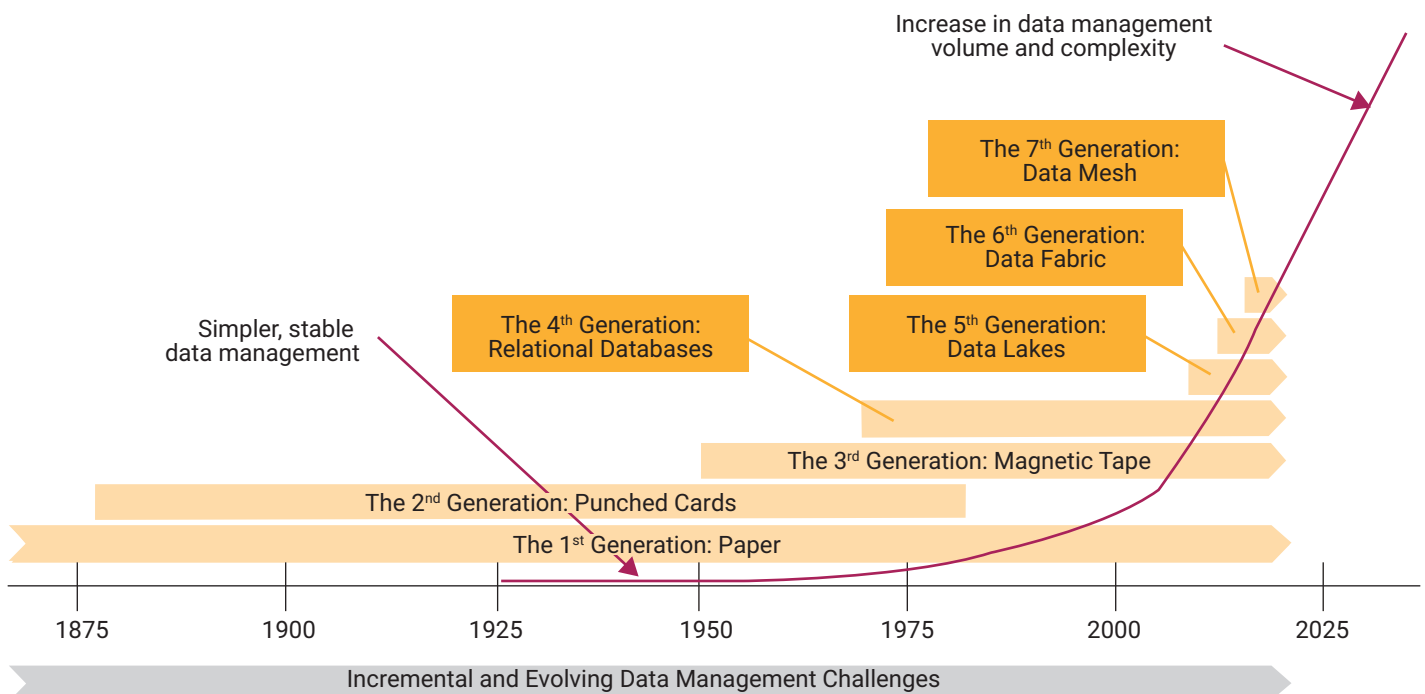
Before computers, paper records such as forms and ledgers served as databases (data storage).¹⁸

Data management by means of paper records was successful. For example, in the early 20th century, the Metropolitan Life Insurance Company accurately and reliably managed 10 million insurance policies with file cards and paper forms.¹⁹

This success was attributable to defining terms and having people perform exact procedures—algorithms—executed in analog. Commodore Grace Hopper, a pioneer in computer programming, subsequently incorporated the algorithms' verb "perform" and noun "procedure" into the Common Business Oriented Language (COBOL) programming language many decades later as a way to make it accessible to business users. She also structured COBOL's arithmetic instructions in the same way as paper-era procedures.²⁰

Another example of the effectiveness of paper database management occurred in January 1936, when the US Congress voted to pay bonuses promised to veterans of World War I. In spite of the complex calculations that had to be performed on paper, more than half of the US\$1.7 billion allocation had been paid before the end of 1936. By comparison, the recent US healthcare IT rollout was less efficient.²¹

FIGURE 1
Seven Generations of Data Management



Slower data management in the precomputer era was arguably more effective than data management today.²² In fact, 18 percent of accountants still use paper ledgers, continuing a legacy of accurate analog recordkeeping that began in the 13th century.²³

Second Generation: Punched Cards

The 1880 US census manually enumerated 50 million people, and it took seven years to publish the results. Even then, many responses remained uncounted.²⁴ Herman Hollerith (whose enterprise became IBM) was looking for a better way to compile statistical information for the 1890 US census and decided on punched cards, which originated as a technology in the textile industry in 1801.²⁵ A punched card is a rectangular paper-based artifact through which holes are punched in predefined places to represent data, allowing a full data set to be represented by multiple punched cards. Using punched card technology, his enterprise published the census results in only three years, establishing its improvement over manual counting.²⁶

Punched cards transformed data from a human-readable form to a computer-readable form. During the 1890 census, each card was stamped with a number corresponding to a family name, and an index maintained elsewhere. If the index was lost, there would be no way to search the data;²⁷ therefore, disaster recovery and business continuity were vulnerabilities. As in the paper era, discipline and organization were vital data management attributes in the punched card era.²⁸

Punched cards were the last generation in which data took a physical form. Their use peaked in 1967, at 200 billion cards per year, but they had become technically obsolete by the mid-1980s.²⁹ However, they were still used for US presidential election ballots as recently as 2000.³⁰

Third Generation: Magnetic Tape

Magnetic tape played a role in audio recordings starting in 1928.³¹ The first use of magnetic tape in computing occurred in 1951. Hard drives (magnetic discs) emerged six years later but did not become commonplace until the 1970s.^{32, 33}

Today, the demand for magnetic tape is stronger than ever. Amazon, Google and Microsoft all use new-

Today, tape storage is more energy efficient and reliable than hard drives, and more secure because the air gap—if a tape is not mounted—means that its data cannot be accessed or modified.

generation tape technology as part of their storage architectures.³⁴ Both IBM and Sony released new tape storage devices in the last five years.³⁵ Even the cloud depends on tape backups.³⁶ Although the amount of data being recorded is increasing by 30 to 40 percent per year, the capacity of hard drives is increasing at less than half that rate.³⁷ Today, tape storage is more energy efficient and reliable than hard drives, and more secure because the air gap—if a tape is not mounted—means that its data cannot be accessed or modified. Tape storage costs only one-sixth of hard drive storage and is getting less expensive.³⁸

The major disadvantages of magnetic tape are the high initial cost of specialized equipment, susceptibility to physical and environmental damage, and the sequential (vs. random access) search mechanism.³⁹ The major data management challenges of the magnetic tape era are storage organization, storage costs, data validation, and the need to balance the costs of greater data accuracy and the costs of delaying the pursuit of accuracy.⁴⁰

Exploration of the privacy consequences of personal data collection began during the magnetic tape generation of data management. For example, a lesson plan for an introductory computer class asked students:

*Would you want other members of the class to see your responses?... What should be done with [them]? Would you feel better if [they were] kept locked up by the teacher? Or would you rather see [them] burned? If [they were] destroyed and we decided later that we'd like to cross tabulate some results or add the results of two or three classes together to get a better overall average we wouldn't be able to. Then what?*⁴¹

More than 45 years later, privacy problems in education technology persist.⁴²

The failure to establish a single view of the truth introduced data-driven operational risk, giving rise to the data silo problem that continues to plague data management.

Fourth Generation: Data Warehouses and Data Silos

The general-purpose relational database was developed in the 1970s when data were measured in human-created megabytes rather than today's machine-generated exabyte streams. It is a useful technology but cannot fulfill operational and analytical needs simultaneously. In turn, Structured Query Language (SQL), the data definition, query, manipulation and control language for the database, was only standardized in 1986.⁴³

Bill Inmon, a US computer scientist and the father of data warehousing, advocated top-down, enterprise-scope, normalized entities and relationships (avoiding redundancy) in the 1980s, while Ralph Kimball, a US innovator, writer, educator and an original architect of data warehousing, advocated bottom-up, business unit-scope, and denormalized fact and dimension tables (reducing joins) in the 1990s.⁴⁴ The popularity of the disciplines created by these pioneers continues today within their specific technical scopes.

It has been proposed that the general-purpose relational database, "while attempting to be a 'one-size-fits-all' solution, in fact, excels at nothing."⁴⁵ Similarly, the so-called single view of the truth is seldom achieved with this technology, as recent polls suggest that 30 percent of enterprises have six or more versions of the truth (data warehouses) in play, duplicating data, cost and effort.⁴⁶ The failure to establish a single view of the truth introduced data-driven operational risk, giving rise to the data silo problem that continues to plague data management.

Redundancy and duplication are two of the main reasons for out-of-control data warehouse costs:

The sheer design of the data warehouse architecture—the [extract, transform, load] ETL processes inherent to data warehousing, the entire replication of the data from the

*transactional systems into the data warehouse, the [online analytical processing] OLAP cubes created to optimize reporting on data warehouses and the several layers of personalized and optimized data copies created—is responsible for creating multiple copies of data at different layers of the data warehouse, which creates unnecessary redundancies and duplication of data and efforts and exponentially increases the infrastructure and maintenance costs over time.*⁴⁷

This generation identified a major problem: Data are often dirty (i.e., inaccurate). Dirty data can be missing data, incorrect data and nonstandard representations of the same data.⁴⁸ In addition, duplicated data often tell different stories, compromising trust.

The height of academic interest in the data quality problem—defining data quality and its dimensions—occurred in the 1990s and early 2000s.⁴⁹ The value of clean data has only recently received greater attention.⁵⁰

Fifth Generation: Data Lakes

Data lake technology originated in 2010.⁵¹ It facilitates trends in big data, machine-generated data, the cloud, and optimization for analytics and end user self-service. It is still maturing. The imposition of a schema, the inability to manage unstructured data and the scaling challenges of data storage in relational databases are gone with big data and data lake technology. Ironically, relational functionality was retrofitted onto data lake technology to serve diverse use cases for data not facilitated by this technology.⁵² The data lake is envisioned as an end to the data silo problem of the previous generation.⁵³

However, any data can be stored in a data lake, and this introduces several issues, such as whether all stored data are needed. Having too much data creates unnecessary complexity and introduces organizational and regulatory risk factors derived from the storage of data regardless of their origins (third-party data).⁵⁴ This problem leads to data lakes migrating from data catchalls to more goal-oriented storage, driving their evolution to data lakehouses.⁵⁵

IT research analysts have begun to declare the failure of data lakes.⁵⁶ There have been many reports of costly failures of execution that turned data lakes

into data swamps—"large data lakes filled with raw, uncurated and siloed data."⁵⁷ Data management has become headline news in this generation, given the assertion that "if a data lake isn't properly managed from conception, it will turn into a 'data swamp,' or a lake with low-quality, poorly cataloged data that can't be easily accessed."⁵⁸ Data lakes pose significant challenges for data governance, data lineage and deficiencies in operational metadata management, and concerns related to privacy, security, access control and even storage costs.^{59, 60}

The onset of this generation turned the once relatively stable world of data management into something much more dynamic and volatile.⁶¹

Today's data management issues have been a long time in the making, and they have yet to be fully addressed.

Sixth Generation: Data Fabrics

A data fabric aims to simplify data access and facilitate data consumption.⁶² Addressing the operational metadata deficiencies of the fifth generation, this generation (originating in concept in 2015⁶³) focuses on data catalogs for finding and understanding data, data hubs for managing lake and warehouse silos, and a data fabric tool set that supports the DataOps paradigm.⁶⁴ Business or operational metadata are the keys to the data fabric; they provide information supporting the stored data, such as their meaning, how they are used, where they come from, their quality, who stewards the data, and their class and category. Though it is still maturing, the primary goal of the data fabric is data orchestration, which combines architecture and technology to facilitate the management of different types and sources of data onsite and in the cloud. It aims to ease the complexities of managing many kinds of data using multiple database management systems deployed across different platforms.⁶⁵

Data virtualization in the data fabric has changed the way some data are managed. However, nonpersistence—the fact that the data are processed in memory (volatile storage) rather than being stored or persisted (nonvolatile storage)—presents challenges for large, complex workloads and data integration.⁶⁶

Data fabric aims to facilitate better data management, mostly by automation.^{67, 68} However, it can introduce new security risk to an enterprise, in spite of centralized and consistent governance and security processes.^{69, 70} Because the data fabric masks the physical location of distributed sources (in the data virtualization layer),⁷¹ the data lineage challenge of the previous generation has not been solved because data visibility is obfuscated.

Seventh Generation: Data Mesh

Data mesh architecture (originating in 2019) introduces a different approach to data governance. The primary differences are federated (vs. centralized) governance, the extent of automation (vs. human intervention) and the measurement of governance success based on data consumption rather than input (e.g., data tables).⁷² It holds much promise for the future of data management.

A data mesh is centered on the creation of data products. Data products can be defined as groups of related data that serve a particular operational need and in which:

- Data product owners (domain owners) are empowered to make decisions about their data.
- Data product owners enforce those decisions by sharing rather than copying data.
- Data sharing across the enterprise is visible (transparency).⁷³

Furthermore, a data product must be discoverable, addressable, self-describing, interoperable, trustworthy and secure. Trustworthiness refers specifically to acting on regular and automated data quality checks.⁷⁴

The data mesh consists of four principles: domain ownership, data as a product, federated computational governance and self-service.⁷⁵ The dynamics of the data mesh involve more than a different approach to data governance as that term is commonly understood by data professionals. The keyword is "computational," meaning that data governance includes consideration of the computational resources of the data ecosystem.⁷⁶ In this way, governance in the data mesh finally bridges the gap between data governance and IT governance.

Comodore Hopper was on to something when she made COBOL accessible to businesses to facilitate



LOOKING FOR MORE?

- Read *Ensuring Privacy Through Effective Data Management*. www.isaca.org/ensuring-privacy-data-management
- Learn more about, discuss and collaborate on information and cybersecurity in ISACA's Online Forums. <https://engage.isaca.org/onlineforums>

business self-service. It has taken 60 years, but self-service is back with the data mesh.

Three Executive Lessons About Data Management

Effective data management does not happen by accident. Today's data management issues have been a long time in the making, and they have yet to be fully addressed. Executive leaders, given their accountability to their chief executive officers, their presidents and, ultimately, their boards, can use the lessons learned from the past to ensure well-managed data. There are three main lessons regarding data management as a discipline, data management skills, and the data management fundamentals of data risk, data quality and metadata:

- 1. **Data management is a discipline, not something to be done off the side of a desk**—The last time the word “discipline” was used in a data management context was during the second generation. This is unfortunate, as data management is now much more a discipline than a mere job description or task, especially when it is not supported by policies, standards, processes and documentation (a foundation of data governance) or by roles, responsibilities and accountabilities (another foundation of data governance) that show a clear segregation of duties to eliminate conflicts of interest. For example, the people providing access to sensitive data and using sensitive data must be different.

There is no longer an excuse for casual data management, given so many standards to support a disciplined approach. These include The Open

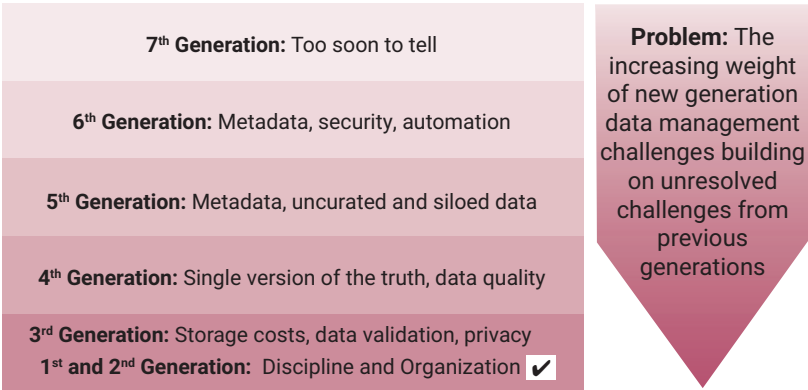
Group Architecture Framework (TOGAF) and The Zachman Framework for overall architecture, COBIT® and International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) 38500 for data platform, the US National Institute of Standards and Technology (NIST) Special Publication (SP) 800-53 and ISO/IEC 27000 for data security, the Information Technology Infrastructure Library (ITIL) for managing service desk issues (responsiveness to end user data problems is key), ISO 8000 for managing data quality and master data, ISO/IEC 11179 for managing metadata, and ISO 31000 for risk management.

The impact of any change in data collection, data storage, data movement or data access should be expressed in security and privacy terms as part of the regular course of business.

Security (including access control) and privacy are key disciplines in any data architecture, and there are many ways to stay on top of these disciplines today. There is thus little to no reason to ignore security and privacy as the bedrock of today's data architectures. Yet, it still happens, in spite of, for example, data privacy being a talking point for at least 45 years. The impact of any change in data collection, data storage, data movement or data access should be expressed in security and privacy terms as part of the regular course of business.

- 2. **Data management skills are changing quickly**—Storage first emerged as a data management issue in the third generation, followed by data architecture in the fourth generation, scalable storage and data diversity in the fifth generation, metadata and interoperability in the sixth generation, and data products and federated data governance in the seventh generation. Data management challenges will continue to diversify. The last three generations emerged very quickly, necessitating continuous and rapid learning. Data management challenges cannot be effectively managed with an earlier generation's

FIGURE 2
Key Data Management Challenges Across Generations



techniques. If there was ever a hesitation to keep up with developments in the data space, its rapid evolution in the last decade should be inspiration for a change in behavior.

Since each new generation of data management emerged to solve problems of the previous generation, being ignorant of how advances in data management technology can solve data challenges introduces data sustainability risk to the organization. There is no choice but to keep reading, be an active part of the global IT and data communities, and gradually introduce new learnings into the organization. The longer the delay, the more overwhelming the volume of information will be to absorb, and the greater the chance of the redundancy of an older data generation's skill set (**figure 2**).

- 3. Sustain a focus on the foundations: data quality, metadata and data risk management**—Data quality and operational metadata have emerged as primary themes of data management in the interests of ensuring fit-for-purpose data. Therefore, it is no surprise that addressing metadata and data quality are foundational elements of full-fledged data management capabilities. Some primary metadata activities include creating a business glossary, creating a data catalog, classifying critical data elements and being able to map back to the production source of those critical data elements. The scope of data quality activities—data cleansing—is usually limited to an organization's critical data elements because the scope of an organization's entire asset is simply too big. It would also be a waste of time, given how much data in an organization are not used. In addition, the discipline of data risk management is instrumental to organizational resilience.⁷⁷ In this case, the primary risk focus areas are to ensure the sustainability of the data asset, that the data are fit-for-purpose, that sensitive data are secure (least privilege with a clearly defined purpose), and that data are accessible by the people who need them.

It is easy to be overwhelmed by growing data management requirements, but not all data are equally important, whether from a discipline perspective, a skills perspective or a foundational data management perspective. It has been estimated that only one-quarter to two-thirds of data are actually used.^{78, 79, 80} The process of identifying critical data is facilitated by the enterprise architecture function (which maps

The facts that only a tiny fraction of organizational data meet basic quality standards and the majority of employees mistrust them highlights the urgent need for better data management practices.

organizational processes and data flows between procedures), which is the sum of business architecture, information architecture, application architecture, data architecture and IT architecture. Data elements that are not part of an enterprise's core processes can often be de-prioritized.

Conclusion

Over time, technological capabilities for managing growing data portfolios have increased, but two key data management capabilities have been lost that may be the root cause of the data challenges organizations currently face: the discipline and organizational attributes of the first and second generations of data management. Low levels of public trust in data practices and occurrences of data misuse might be less common if discipline and organization could be sustained as pillars of data management.

Statistics on data mistrust and abuse should alert organizations to prioritize data quality and security as part of rebuilding public trust in their organizations. The facts that only a tiny fraction of organizational data meet basic quality standards and the majority of employees mistrust them highlights the urgent need for better data management practices. The potential for catastrophic security breaches caused by unauthorized access to data underscores the importance of implementing robust security protocols. All of these activities demand discipline and organization to be effective and sustainable.

Organizations must ensure that data are fit for their intended purposes and are being used legally or they risk damaging their reputation, losing the trust of their employees and customers and potentially facing significant financial penalties. There is no better time to recognize the role that data management plays in achieving organizational objectives and, thus, to invest in the necessary resources and expertise to manage data effectively.

Endnotes

- 1 Foote, K. D.; "A Brief History of Data Management," Dataversity, 19 February 2022, <https://www.dataversity.net/brief-history-data-management/>
- 2 Nagle, T.; "Only Three Percent of Companies' Data Meets Basic Quality Standards," *Harvard Business Review*, 11 September 2017, <https://hbr.org/2017/09/only-3-of-companies-data-meets-basic-quality-standards>
- 3 COSOL, "The High Cost of Low Quality Data," <https://cosol.global/the-high-cost-of-low-quality-data>
- 4 DalleMule, L.; T. H. Davenport; "What's Your Data Strategy?" *Harvard Business Review*, May–June 2017, <https://hbr.org/2017/05/whats-your-data-strategy>
- 5 Hartman, T.; H. Kennedy; R. Steedman; R. Jones; "Public Perceptions of Good Data Management: Findings From a UK-Based Survey," *Big Data and Society*, 2020, <https://journals.sagepub.com/doi/pdf/10.1177/2053951720935616>
- 6 Townsend, M.; "Facebook–Cambridge Analytica Data Breach Lawsuit Ends in 11th Hour Settlement," *The Guardian*, 27 August 2022, <https://amp-theguardian-com.cdn.ampproject.org/c/s/amp.theguardian.com/technology/2022/aug/27/facebook-cambridge-analytica-data-breach-lawsuit-ends-in-11th-hour-settlement>
- 7 Chawda, V.; "Building Trust in Government's Use of Data," KPMG, June 2018, <https://home.kpmg/xx/en/home/insights/2018/06/building-trust-in-governments-use-of-data.html>
- 8 Paris21, "Building Trust in Data," https://paris21.org/sites/default/files/inline-files/CRF_BackgroundNote_0.pdf
- 9 Rainie, L.; A. Perrin; "Key Findings About Americans' Declining Trust in Government and Each Other," Pew Research Center, 22 July 2019, <https://www.pewresearch.org/fact-tank/2019/07/22/key-findings-about-americans-declining-trust-in-government-and-each-other/>
- 10 Cohen, R.; "Breaking Down Public Trust," University of Michigan, Ann Arbor, Michigan, USA, 10 June 2021, <https://fordschool.umich.edu/news/2021/rebuilding-trust-in-government-democracy>
- 11 Pradhan, S.; "An Open Government Approach to Rebuilding Citizen Trust," Open Government Partnership, <https://www.opengovpartnership.org/trust/an-open-government-approach-to-rebuilding-citizen-trust/>
- 12 Op cit Cohen
- 13 Murtin, F.; "Trust and Its Determinants: Evidence From the Trustlab Experiment," Organisation for Economic Co-operation and Development (OECD), 25 June 2018, https://www.yann-algan.com/wp-content/uploads/2020/12/Murtin-et-al-2018_Trust-and-its-Determinants_OECD.pdf
- 14 Staff, "Eighty-Three Percent of IT Leaders Are Not Fully Satisfied With Their Data Warehousing Initiatives," *Businesswire*, 5 August 2020, <https://www.businesswire.com/news/home/20200805005182/en/83-of-IT-Leaders-are-Not-Fully-Satisfied-with-their-Data-Warehousing-Initiatives-According-to-New-Research-from-SnapLogic>
- 15 El Shatby, S.; "The History of Data: From Ancient Times to Modern Day," 365 Data Science, 1 June 2022, <https://365datascience.com/trending/history-of-data/>
- 16 Driscoll, K.; "From Punched Cards to 'Big Data': A Social History of Database Populism," *Communication+1*, vol. 1 iss. 1, 2012, <https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1006&context=cpo>
- 17 Eggers, W. D.; J. Knight; B. Chew; R. J. Krawiec; "Rebuilding Trust in Government," Deloitte, 9 March 2021, <https://www2.deloitte.com/us/en/insights/industry/public-sector/building-trust-in-government.html>
- 18 Gutteridge, L.; "Systems Before Computers had Better Database Technology," *Codeburst*, 7 June 2019, <https://codeburst.io/https-medium-com-databases-bfore-computers-6d57be7db9c5>
- 19 *Ibid.*
- 20 *Ibid.*
- 21 Gutteridge, L.; "The Great Generation Was Really Good at Enterprise Systems—Today We Suck," *Codeburst*, 1 November 2018, <https://codeburst.io/the-great-generation-was-really-good-at-enterprise-systems-we-suck-e0ddb051735>
- 22 *Ibid.*
- 23 Austin, A.; "One in Five Accountants Still Using Paper Ledgers," *Accountancy Daily*, 13 October 2017, <https://www.accountancydaily.co/one-five-accountants-still-using-paper-ledgers>
- 24 Armstrong, D.; "The Social Life of Data Points: Antecedents of Digital Technologies," *Social Studies of Science*, vol. 49, iss. 1, 2019, p. 102–117, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6902902/>
- 25 Computer History Museum, "1801: Punched Cards Control Jacquard Loom,"

<https://www.computerhistory.org/storageengine/punched-cards-control-jacquard-loom/>

- 26 *Op cit* Armstrong
- 27 *Op cit* Driscoll
- 28 Grinyer, P. H.; "What Were Punch Cards and How Did They Change Business," British Computer Society, 24 November 2021, <https://www.bcs.org/articles-opinion-and-research/what-were-punch-cards-and-how-did-they-change-business/>
- 29 Scharf, C.; "Where Would We Be Without the Paper Punch Card?" *Slate*, 23 June 2021, <https://slate.com/technology/2021/06/caleb-scharf-ascent-of-information-punch-cards.html>
- 30 John, D. W.; "Punched Cards: A Brief Illustrated Technical History," University of Iowa Department of Computer Science, Iowa City, Iowa, USA, 23 October 2012, <https://homepage.divms.uiowa.edu/~jones/cards/history.html>
- 31 Iron Mountain, "The History of Magnetic Tape and Computing: A 65-Year-Old Marriage Continues to Evolve," <https://www.ironmountain.com/resources/general-articles/t/the-history-of-magnetic-tape-and-computing-a-65-year-old-marriage-continues-to-evolve>
- 32 Computer History Museum, "Magnetic Tape," <https://www.computerhistory.org/revolution/memory-storage/8/258>
- 33 *Op cit* Gutteridge 2019
- 34 *Op cit* Iron Mountain
- 35 *Op cit* Computer History Museum, "Magnetic Tape"
- 36 Decker, S.; "The Future of the Cloud Depends on Magnetic Tape," Bloomberg, 17 October 2018, <https://www.bloomberg.com/news/articles/2018-10-17/the-future-of-the-cloud-depends-on-magnetic-tape#xj4y7vzkg>
- 37 Lantz, M.; "Why the Future of Data Storage Is (Still) Magnetic Tape," *IEEE Spectrum*, 28 August 2018, <https://spectrum.ieee.org/why-the-future-of-data-storage-is-still-magnetic-tape>
- 38 *Ibid.*
- 39 Comp Sci Station, "Magnetic Tape Storage: Advantages and Disadvantages," 12 April 2018, <https://compscistation.com/magnetic-tape-storage-advantages-and-disadvantages/>
- 40 Cooper, B. E.; "Computing Aspects of Data Management," *Journal of the Royal Statistical Society*, vol. 21, iss. 1, 1972, p. 65–75, https://www.jstor.org/stable/2346608?read-now=1&oauth_
- 41 *Op cit* Driscoll
- 42 LinkedIn, "Google Transfer Analysis for Student in Denmark," 15 August 2022, <https://www.linkedin.com/feed/update/urn:li:activity:6965298694191521794/>
- 43 *Op cit* Driscoll
- 44 Ekanayake, I.; "Inmon vs. Kimball—The Great Data Warehousing Debate," *Medium*, 7 February 2021, <https://medium.com/cloudzone/inmon-vs-kimball-the-great-data-warehousing-debate-78c57f0b5e0e>
- 45 *Op cit* Driscoll
- 46 Wells, D.; "The Continuing Evolution of Data Management," Eckerson Group, 11 February 2019, <https://www.eckerson.com/articles/the-continuing-evolution-of-data-management>
- 47 Kodikal, P.; "Five Limitations of Data Warehouses in Today's World of Infinite Data," Dremio, 16 August 2021, <https://www.dremio.com/blog/5-limitations-of-data-warehouses-in-today-s-world-of-infinite-data/>
- 48 Kim, W.; B.-J. Choi; E.-K. Hong; S.-K. Kim; D. Lee; "A Taxonomy of Dirty Data," *Data Mining and Knowledge Discovery*, vol. 7, iss. 1, 2003, p. 81–99, https://www.researchgate.net/publication/220451798_A_Taxonomy_of_Dirty_Data
- 49 Keller, S.; G. Korkmaz; M. Orr; A. Schroeder; S. Shipp; "The Evolution of Data Quality: Understanding the Transdisciplinary Origins of Data Quality Concepts and Approaches," *Annual Reviews*, 6 January 2017, <https://www.annualreviews.org/doi/10.1146/annurev-statistics-060116-054114>
- 50 *Op cit* Kim et al.
- 51 Foote, K. D.; "A Brief History of Data Lakes," Dataversity, 2 July 2020, <https://www.dataversity.net/brief-history-data-lakes/>
- 52 Russom, P.; "The Evolution of Data Lake Architectures," Transforming Data With Intelligence (TDWI), 29 May 2020, <https://tdwi.org/articles/2020/05/29/arch-all-evolution-of-data-lake-architectures.aspx>
- 53 i-SCOOP, "Data Lakes and the Data Lake Market: The What, Why and How," <https://www.i-scoop.eu/big-data-action-value-context/data-lakes/>
- 54 Kidd, C.; "Data Storage Explained: Data Lake vs Warehouse vs Database," BMC Blogs, 28 October 2020, <https://www.bmc.com/blogs/data-lake-vs-data-warehouse-vs-database-whats-the-difference/>
- 55 *Op cit* i-SCOOP
- 56 *Ibid.*

- 57 Drayson, M.; A. Bashir; *The Evolution of Data Management: A Practitioners Perspective*, Nippon Telegraph and Telephone (NTT), Japan, <https://www.dimensiondata.com/-/media/ntt/global/solutions/intelligent-business/intelligent-business-landing-page/evolution-of-data-management-ebook.pdf>
- 58 Edwards, J.; "Five Things You Need to Know About Data Lakes," *Network Computing*, 31 May 2019, <https://www.networkcomputing.com/data-centers/5-things-you-need-know-about-data-lakes>
- 59 *Op cit* Wells
- 60 Banafa, A.; "What Is a Data Lake," BBVA OpenMind, 9 June 2021, <https://www.bbvaopenmind.com/en/technology/digital-world/data-lake-an-opportunity-or-a-dream-for-big-data/>
- 61 *Op cit* Wells
- 62 IBM, "What Is a Data Fabric?" <https://www.ibm.com/analytics/data-fabric>
- 63 Power, L.; "Unifying Data Management With Data Fabrics," *IT Business Edge*, 17 June 2022, <https://www.itbusinessedge.com/storage/data-fabrics/>
- 64 *Op cit* Wells
- 65 *Ibid.*
- 66 *Op cit* Drayson and Bashir
- 67 Talend, "What Is Data Fabric?" <https://www.talend.com/resources/what-is-data-fabric/>
- 68 Raza, M.; "Data Fabric Explained: Concepts, Capabilities and Value Props," BMC Blogs, 9 November 2021, <https://www.bmc.com/blogs/data-fabric/>
- 69 Tibco, "What Is Data Fabric?" <https://www.tibco.com/reference-center/what-is-data-fabric>
- 70 *Op cit* Raza
- 71 Swayer, S.; "Data Fabric Architecture: Advantages and Disadvantages," *IT Pro Today*, 16 August 2021, <https://www.itprotoday.com/analytics-and-reporting/data-fabric-architecture-advantages-and-disadvantages>
- 72 Dehghani, Z.; "Data Mesh Principles and Logical Architecture," MartinFowler, 3 December 2020, <https://martinfowler.com/articles/data-mesh-principles.html>
- 73 J. P. Morgan Chase, "Evolution of Data Mesh Architecture Can Drive Significant Value in Modern Enterprise," <https://www.jpmorgan.com/technology/technology-blog/evolution-of-data-mesh-architecture>
- 74 Rigol, X. G.; "Data as a Product vs. Data Products. What Are the Differences?" *Towards Data Science*, 8 July 2021, <https://towardsdatascience.com/data-as-a-product-vs-data-products-what-are-the-differences-b43ddb0f123>
- 75 Perrin, J.-G.; "The Next Generation of Data Platforms Is the Data Mesh," *Medium*, 3 August 2022, <https://medium.com/paypal-tech/the-next-generation-of-data-platforms-is-the-data-mesh-b7df4b825522>
- 76 *Ibid.*
- 77 Pearce, G.; "Data Resilience Is Data Risk Management," *ISACA® Journal*, vol. 3, 2021, <https://www.isaca.org/archives>
- 78 Jha, P.; "Top Three Data Challenges Companies Faced in 2020," *Spiceworks*, 25 November 2020, <https://www.spiceworks.com/tech/data-management/news/top-3-data-challenges-companies-faced-in-2020/>
- 79 *Op cit* DalleMule and Davenport
- 80 IBM, "Build Your Data Architecture," <https://www.ibm.com/resources/the-data-differentiator/data-architecture>