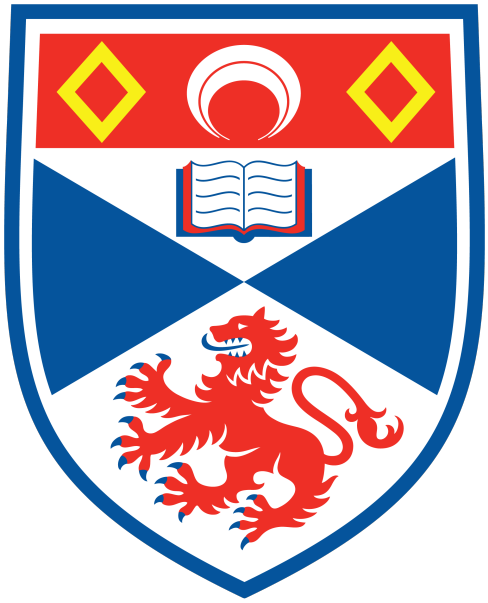
The Evolution of Data Management Systems

– from pre-digital to digital database management systems

Topic 1: Before Databases

Group A - Authors ordered by matriculation number:
200029960 210016568 210018181 210020818



University of St Andrews

IS5102 - Database Management Systems
Academic Year 2021/2, Semester 1 - Practical 3

Lecturer: Alexander Konovalov

Word Count: 4834

Due Date: November 23, 2021

Contents

I. Abstract - 210020818

In 2021, digital database management systems are an integral part of the software architecture which underpin any organisation's software systems. In the modern era database management systems have relational and non-relational implementations, and organisations now consider database efficiency as a key component of an applications overall performance. With the huge reliance on digital database management systems today, the question of how an organisation could organise and manage data at scale prior to their inception is often asked. There is considerable intrigue about how some of the pre-digital data management systems worked, and whether there were any similarities and differences compared to modern day implementations.

II. Introduction - 200029960

Pre-digital database management systems refers to a system of recording, organizing, and indexing data on paper before the advent of computers.

The conceptualisation of a database was developed and implemented long before the digital age. Pre-digital data management systems saw data stored in journals, ledgers, catalogue cards, and endless archives. We will explore pre-digital systems through the library, banking, and healthcare industries and compare their practices to our modern database management systems. Each example posits key issues surrounding this era of database management, such as security, accuracy, long-term preservation, and the classification of documents which we study in further detail to understand how pre-digital database management systems truly worked.

III. Pre-Digital Data Management Systems

i. Libraries - 200029960

Prior to the digitalisation of library data management systems, a physical paper-based catalogue card system operated across libraries worldwide. For centuries, card catalogues were an essential way in which libraries catalogued and indexed their books. Each book had its own card, which held essential information such as author, title, subject matter, and additional indexes such as its physical location within the library. Each card was then physically held in a drawer organised alphabetically for ease of access[1].

The importance of sorting and categorising information was just as crucial for the physical catalogue system as it is within today's modern data management systems. In the 1870s, Melvil Dewey, an American librarian, revolutionised how library systems operated. The Dewey Decimal System (DDS) vastly surpassed previous 'fixed location systems', which saw a fixed storage location for each book until the next re-categorisation of the library. The DDS organises the contents of a library based on the grouping of books into ten categories, with each category assigned 100 numbers. Each category is then divided further to deliver more specified subject areas[2]. Dewey set a promise that the number's meanings were forever connected and there would

only be new numbers added for subjects. Coupling his promise with the ability to systematise a shelf arrangement, the DSS solved the most significant issues within the industry at the time - lack of standardisation and ability to classify books subject logically and physically[1], which saw mass adoption throughout the USA.

Comparing the library pre-digital data management system to that of a modern digital system has some very evident differences. The digital library removes the physical boundary that a pre-digital library poses; it can effectively operate from a microcomputer with a large memory compared to the thousands of units that hold tens of thousands of card-catalogues[3]. Due to this, the modern database management system holds all the benefits of the pre-digital system, but with the further benefit of being much more flexible with access being possible from many users for the same record at the same time in comparison to one user per resource and a waitlist initiated for further interested parties.

Furthermore, the digital database management system will demand greater accuracy in punctuation, spelling, coding, and even word arrangement as records become standardised. Any errors are magnified in a digital database management system as they can lead to records becoming unretrievable[4] which could be considered a drawback for the modern-day system. However, if the system has rules in place, such as fields that cannot be null, overall data quality will improve due to the number of discrepancies reduced. The improved data quality opens further advantages of today's data management systems, such as the user's ability to search differing fields. Users can broaden or narrow their search by utilising Boolean operators such as OR, AND, and NOT within digital library systems such as the online public access catalogue (OPAC) that Morrupisi and Mooko[5] analyse. Keyword searching, Boolean, and even the ability to filter their search by document type create a very different and user-friendly system to the card catalogue.

Within the structure of a card catalogue system, there are various distinct entries for a single system. The modern catalogue has a master file that retains a one entry structure for each item; which has a number of indexes as access points within the master file. Although both types of catalogue systems differ in their process of record creation, manipulation, and retrieval of information, there are some similarities. Both systems employ similar hierarchical categories; many digital library data management systems follow the Dewey Decimal System when categorising and allowing users to browse via this method. The first digital version was released in 2006 and is now widely used and referred to as 'WebDewey'[6].

Overall, technology has allowed libraries to use modern digital data management systems to eliminate restrictions on opening hours in a multi-user environment, improving searching and reducing the cost and physical space required.

ii. Banking - 210020818

Prior to the inception of modern data management systems banking purposes had to be fulfilled by more traditional means. Banks would keep detailed paper records for every account along with information about the account holder. Information about customers and their banking records were stored in large filing cabinets and sorted using a suitable system chosen at the organisations discretion. This file system formed the central record keeping system for the institution and would be used by employees regularly to perform operational tasks. Choosing an efficient filing system helped the banking company store data systematically and efficiently. It also helped employees with access to the filing system to retrieve the information they required easily.

The process of developing a suitable filing system involves determining a list of categories to store all of the information within. A category in this context is a group of documents which belongs together. When establishing categories an organisation would look at it in a similar way

to how we develop ER models today. It is important to look at the list critically and determine whether categories should be broken up further into sub categories or whether they should be combined. This logic was implemented to ensure fast storage/retrieval of documents, it should not be the case that there are too many categories making it difficult for someone to decide which category information will belong. Similarly there should not be too few categories so that finding an item becomes laborious. Once the categories have been determined a filing index is then drawn up so that everyone can understand the system being used. This index is called the filing key. Within each category files are sorted by a filing rule i.e. alphabetically, chronologically. When a file is removed from the system, a protocol must be put in place to ensure the file is not lost or not returned. File out books were used to document who borrowed a file and when it was taken/replaced. File out cards were used in the filing cabinet to sit in place of the borrowed file, they also contained the details logged in the file out book [7].

Large paper trails and documents manipulated by hand were used to fulfil customer needs and ensure the integrity of the system. Functions we ask of our banks today such as depositing funds, making withdrawals and transferring between accounts were performed by completing various forms and moving signed/stamped documents between stakeholders. The following documents were required to make this pre-digital management system work effectively:

1. Ledger – Every bank would have a ledger which is permanent summary of all accounts recorded in journals. The Ledger lists individual transactions across all accounts by date. Double-entry accounting is used with all debit and credit entries being transcribed in separate columns, along with any metadata describing the transaction e.g. account numbers, date and time. The Ledger balance would be calculated by the bank at the end of each working day.

2. Passbook – A passbook is a physical document owned by an account holder. It is a written log of all bank transactions made against an individual's account. These transactions would include: Debits, Credits, Loans and Fixed Deposits [8]. Its typical purpose is to allow a customer to keep track of their account, and it is required to perform any transaction whilst in a branch of the bank. The bank would also have a copy of this passbook, known as a bank passbook.

3. Deposit/Debit Slip – A bank deposit slip accompanies a cash or cheque payment and provides the bank teller with all relevant metadata about the transaction so that the right amount of money reaches the correct account. This information typically includes the account holders name, their account number and what type of deposit is being made [9].

4. Withdrawal Slip – A withdrawal slip is a written document containing an instruction for the bank to pay a stated amount to a stated individual. It is quite similar to a deposit slip in that it is effectively a record of a transaction. A withdrawal slip typically contains the account number from which to withdraw funds from, branch information, details about the payee, the amount to withdraw, and other metadata such as the date and time of the transaction [10].

5. Cheque book – A cheque book was introduced to allow a bank's customers to make payments to a third party without having to first withdraw the funds themselves. It made the process of transferring money between banking organisations more seamless. If a person received a cheque they could take it to their own bank, the bank would then stamp it and send it to the central bank who would release the funds. The central bank would then also add their stamp to the cheque and request to be reimbursed by the institution who issued the cheque.

6. Bank Statements - Paper statements produced periodically on a monthly, quarterly or annual

basis allowing account holders to see all transactions processed on their account including interest added to their balance [11].

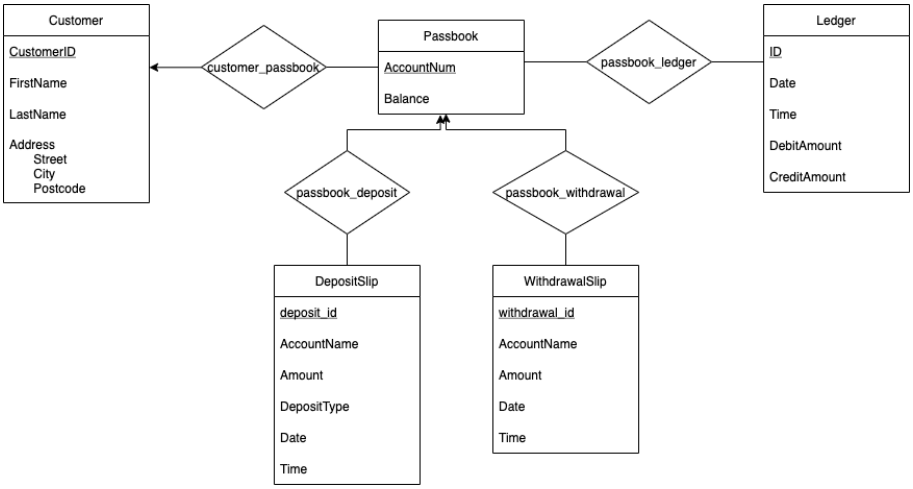


Figure 1: *- ER Diagram showing how the paper-based document used in the pre-digital banking system can be modelled for a DBMS*

Every time a function was executed these documents would be updated in turn to ensure processes were followed correctly. The double entry accounting practices involved and the strict procedures which every bank put in place made the industry highly secure in the absence of modern day technologies. From looking at the documents listed above we quickly notice how they cross reference each other and can immediately identify relationships. Figure 1 shows an ER diagram that would be designed today to model this data-set. The diagram below highlights the striking similarities in how data was stored and organised pre-digitally.

Since the advent of digital data management systems, the banking industry has begun to adapt and now has large data storage records for all its customers and operations. The benefits of this evolution have been huge and can be seen in several ways:

1. Removing Data Redundancy: The introduction of modern relational databases led to the inception of a technique known as database normalisation. The purpose of database normalisation is to eliminate redundant (repetitive) data and ensure that information is stored logically. This is achieved by dividing large tables into smaller ones and linking them using relationships. The benefits are that it removes insertion, update and deletion anomalies as information is not stored in multiple places [12].

2. Security Problems: The transition to digital DBMS meant that customers information was more secure as it wasn't subject to traditional threats such as fire and theft. However, new security concerns began to arise, these include SQL injection, weak password protection, improper user security, Distributed Denial of Service (DDoS) attacks [13].

3. Reducing Operational Latency: Naturally the time it took to perform UPDATE, INSERT and DELETE operations was also significantly reduced. In the modern-day frameworks such as Spring Data exist to abstract these functions further so that backend services with REST APIs can be developed with ease [14].

However it could be argued that technological advancements have led to some drawbacks. These include but are not limited to:

1. Data integrity concerns: Data integrity is the accuracy and consistency of data stored in a database. Pre-digital systems did not have to give this as much consideration as information was updated by hand on paper. In digital DBMS, data integrity is imposed when the database is designed, and is validated through the use of ongoing error-checking routines. In modern banking organisations there is a role for security professionals who use concepts such as: data encryption, data backup, access controls, input validation, data validation to ensure data integrity across the system [15].

2. Concurrency issues: As IT infrastructure evolved it became clear that making simultaneous requests on a dataset was a realistic concern. Concurrent access is not an issue if users are just reading data, however, when we have a mixture of REAS and WRITE operations it becomes a challenge. DBMS concurrency control is a group of strategies which can address these conflicts. Strategies include: timestamp protocols, lock-based protocols and validation-based protocols [16].

3. Loss of face to face customer service: The introduction of digital DBMS has led to constant evolution within the banking sector, nowadays online banking is used predominantly to fulfil customer needs. Some people feel the loss of face to face customer service has been detrimental and do not like using online interfaces to transfer large sums of money.

iii. Healthcare - 210016568

Before the invention of the computer, almost every industry, including healthcare, stored data by making full use of papers. Digital data management has gradually replaced paper-based management systems with computers, which can be classified into three stages: manual management, files management, and database system management. In such a traditional industry, healthcare data has transitioned from purely paper-based tracking to digital information in the past 40 years. Nevertheless, many types of medical data are not yet digitised or integrated into health data management systems.

Previously, statistical data in a hospital was mainly to record each patient's condition in detail. Each patient's case was completed by hand in the early stages, which helped the hospital better analyse the condition and follow-up treatment. However, as time has progressed and hospital functions have increased, they are increasingly generating more and more data. Evidencing that paper-based data management is unsustainable.

As paper records in hospitals disappeared, digital data management has become an integral part of the healthcare system. A single outpatient visit or hospitalisation can generate thousands of data elements-from opening clinical records to diagnosis and treatment and patient bills. In 2012, the global digital healthcare data was estimated to be about 50 PB, and in 2020, only eight years later, it was estimated to be as high as 2,500 PB[17].

Currently, many fields in the health information system represented by Hospital Information System (HIS) have been widely used in the hospital care environment[18]. As the Figure 2 showing, HIS is a factor of health informatics, which mainly emphasises the management needs of hospitals. It aims to create a paperless environment that can cover all hospital operations, such as clinical, administrative, and financial systems[19]. Health Data Management (HDM) plays an essential role in HIS, which is the "housekeeper" in the electronic medical record/Electronic health records (EMR/EHR). In this process, data is efficiently and logically organised, checked for errors, processed, retrieved, and kept safe in a safe environment without redundancy.

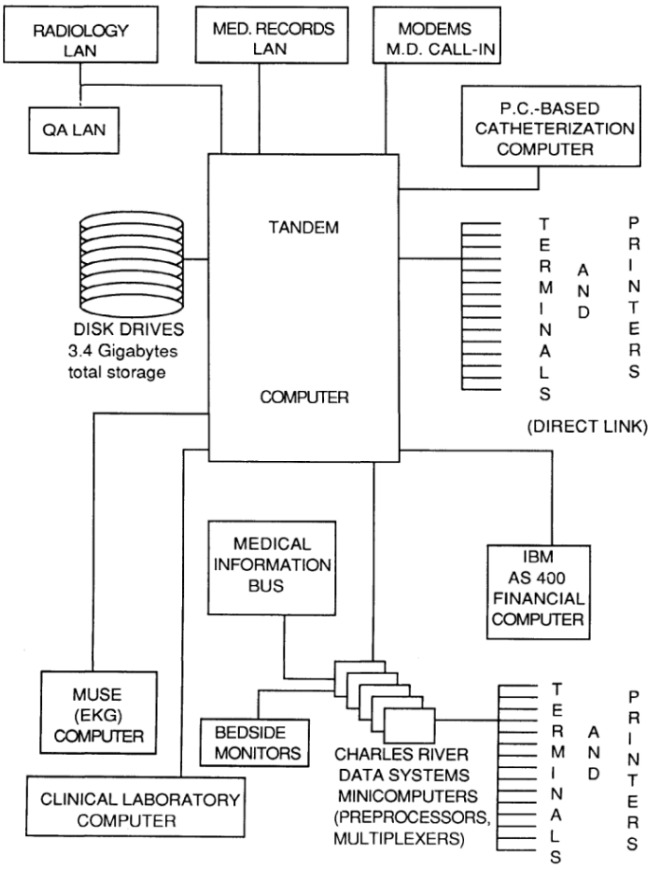


Figure 2: *Hospital Information System Diagram[20]*

Health data includes but is not limited to: patient demographics, medical notes, laboratory test results procedures, and surgical imaging, such as X-rays, computed tomography (CT) and MRI prescription referrals, and other communication provider information[21]. The source of health data traditionally comes from the patient through the provider (like medical staff), and the provider enters the data into the medical record system. EMR usually uses an electronic version of the paper-keeping record with EHR used for more complete records implemented in hospitals and larger medical institutions. Patients can access their health data through the Personal Health Record (PHR), the custodian. However, not all data in EMR and EHR can be shared directly with patients, such as biopsy of suspected cancer or other life-changing and worrying test results or predictions.

Healthcare providers can provide better care through better access to patient health data. Patients can provide data in addition to face-to-face consultations and view their health records to follow their care plan to stay engaged. However, not all data in EMR and EHR can be shared directly with patients, such as biopsy of suspected cancer or other life-changing and worrying test results or predictions. The medical data management system allows clinicians and their employees to access essential details through a platform that ensures vital details are not overlooked. These systems enable doctors to ensure that they receive timely and appropriate reimbursements and help them track patient outcomes and discover inefficiencies that may affect their health in practice.

If clinical and administrative information is contained in a separate database that is not integrated, providers and staff will not have access to all the data needed to make informed decisions about the future of patient care, finance, and practice.

Essentially, these types of data management methods are in line with historical trends, allowing to record the necessary data and extract information but may vary from efficiency. They all encounter data storage and data security problems and need to deal with them in various ways.

IV. Key issues of Pre-Digital Data Management Systems

i. Security - 200029960

Security is a vital concept within a data management system that is formed through the encompassment of tools, processes, and methodologies. Within modern systems, the database and the management system have programmes designed to protect them and every application that accesses it to reduce the threat of cyber-attacks, misuse, and damage. However, security within pre-digital data management systems is at significant risk due to the lack of control and management the system allows.

Within both pre-digital and modern systems, human error is a real threat to security. It is the cause of approximately half of data breaches reported in 2019[22], as shown in Figure 3:

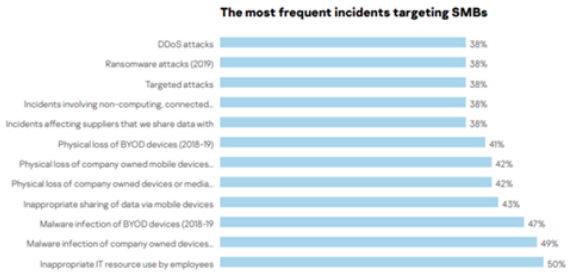


Figure 3: *Kaspersky found that in 2019 SMBs were most frequently affected by incidents due to inappropriate IT resource use by employees[22]*

As illustrated by Kaspersky's research of the most frequent incidents targeting small and midsized businesses (SMB), further factors relating to human error significantly contributed to the number of security incidents in 2019. These include loss of company owner devices or media (42%), loss of company-owned mobile devices, and loss of bring your own devices (BYOD) (42% and 41%, respectively). The same issues are relevant and more prevalent in pre-digital data management systems as paper documents can easily be damaged, mishandled, or lost, which poses a high risk to any organisation operating with paper-based data management systems. The advantage of modern database management systems is the ability to utilise password management to documents and further enhance security through multifactor authentication, which reduces the number of opportunities that human error can arise. Controlling privileges to areas within the system reduces the amount of information exposed, even if a user commits an error that leads to a data breach[23].

There are examples of top-secret multi-million-pound military paper documentation being stolen which confirms that "you cannot realistically encrypt the written word"[24]. This incident could have been prevented if the information had been digitalised through encryption and access controls being in place. During early 2021, the UK Government released figures for the percentage

of businesses that have identified or experienced breaches in a 12-month period as shown in Figure 4:

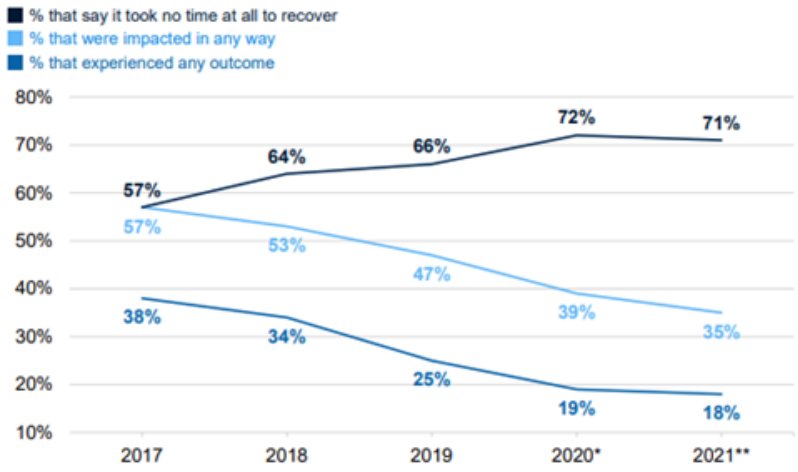


Figure 4: *Over 600 businesses per year that identified a breach or attack in the previous 12 months[25]*

Initially, this looks as though breaches and attacks are reducing through time from 2019 to 2021. However, this is likely due to the introduction of the General Data Protection Regulation (GDPR) in 2018, which required all businesses in the UK to implement essential cyber security measures. In addition, this could likely reflect the rising trend of cloud usage and backups, which digital data management systems allow. There is clearly a substantial security issue in pre-digital systems as coupling the likelihood of misplaced, lost, or even damaged records with the lack of ability to backup and recover a database after a failure poses a risk. The severity of the risk is minimised within modern data management systems as databases can be restored from an up-to-date, consistent state if a frequent backup process is implemented. Having a digital backup provides assurance to the organisation and customers and many modern management systems opt for data backup in the cloud as it has the benefit of scalability, reduced costs to alternatives, and no additional workload for employees[26]. It is however best practice to have a backup copy on premise for ease of access with the cloud backup in more serious incidents.

ii. Accuracy - 210016568

Accuracy is one of the most basic requirements in data management[27]. It means the closeness of repeated observations of a certain feature of the same object or repeated estimates of a parameter to each other. It is critical for a data manager to record the data accurately and thus predict the potential trend accurately.

In pre-digital systems, human errors are the leading cause of data accuracy[28]. One of them is some small mistakes like typos without disturbing the meaning. For instance, upper-case I, lower-case l and number 1 are very easy to confuse. Yet, some unconstructed data may be easy to cause misunderstandings due to the ambiguous description[27]. For example, a set of data without text description is not easy to understand the meaning of his expression. The cause of this situation is likely to be because the recorder did not label the name, purpose, and source of each item of data caused.

Similarly, as for modern systems, it also plays an integral part in systems. Even minor typos will affect the operation of the data management systems[29]. Considering the first example I

mentioned, it is hard for a computer to distinguish upper-case I, lower-case l and number 1; if an error occurred during data collection, the computer would treat "paI", "paI", and "pa1" as three different words. This leads to further incomplete and inaccurate data when querying the database content.

Nevertheless, one of the benefits is that with the development of advanced technology, automatic correction services are provided, which prevents basic typos[30]. These techniques (in Figure 5) can correct the content in the database based on context, common input errors, thus reducing errors due to the data collection process. Unlike the pre-digital systems, the correction of data errors relies almost exclusively on human intuition and awareness.

Figure 5: *Correction process for interconnected typos[30]*

Moreover, the correlation among different systems could detect inconsistent data[31]. In order to apply data in various systems and assist in understanding the data clearly, the unified data representation is constructed with less misled. Nowadays, common approaches include setting constants when performing programming to ensure data consistency in subsequent systems. More checkboxes are increasingly used than input boxes when collecting data to ensure data consistency between the user interface and the data collection system and introduce error checking mechanisms when submitting data.

With the proper data storage, the previous data are organized formally, and it is easy to analyze and visual through some basic statistical methods. The users have a higher probability of deducing the likely data trends and thus helping institutes' operation, so-called accuracy.

iii. Long-Term Preservation - 210018181

Data preservation has always been a fundamental issue. As data collection becomes more and more electronic, modern society still requires paper documents. Electronic data is more likely to be tampered with and erased. Therefore, some highly confidential data is still stored in paper files. Nowadays, we can still find documents hundreds or even more than a thousand years ago, thanks to people's research on preserving paper documents for a long time.

Affected by paper carriers, paper archive materials have the disadvantages of being damp, flammable, and ageing. Therefore, to keep the paper for a more extended period, natural materials with a strong alkali, such as lime and asbestos, are usually used to make paper. Because there are many acid gases in the air, these gases will cause the paper to age and break, making the paperless susceptible to ageing and damage. This method is changed from the material itself.

Another critical method is to create a suitable environment for paper documents. If the humidity of the environment is too high, the moisture content of the paper will increase. Once the moisture content of the paper increases, it will absorb the acid gas in the air because the acidic substance has a significant effect on the cellulose in the paper, so when the humidity of the paper is very high, the paper is easier to oxidize. When the relative humidity exceeds 85%, the physical strength

of the paper will lose more than 30%, the elasticity will be wholly lost, and the original state will not be restored after drying[32].

In addition, if the relative humidity is lower than 25%, the paper will become hard and brittle due to loss of moisture due to drying. The elasticity and folding resistance will be weakened, and it will be easily damaged. Therefore, the humidity in the archive's warehouse is essential, and it must be kept as stable as possible, preferably at 45%-60%[33]. When the temperature rises, the material molecules gain heat energy, the molecular movement speeds up, the kinetic energy increases, the number of activated molecules increases, and the chemical reaction also speed up. Cellulose-related reactions that are hardly carried out at ordinary times will also occur, and the speed of paper ageing will increase. When the temperature is too high, the paper will lose moisture and become hard and brittle. If the temperature continues to rise, the cellulose will be dehydrated, carbonized, and burned[34]. Therefore, the temperature of the archive's warehouse must be controlled, not too high or too low, preferably between 14-24°C[35]. So, a good environment can affect the storage time of paper.

Compared with current digital databases, digital databases have relatively low environmental requirements. If the database is stored on a disk, then there are specific requirements for the environment. However, the current Internet technology supports storing the database on the server, which means that the digital database does not have any requirements for the environment, as long as the server always exists. This data is permanently stored in the database, and the only thing that needs to be paid attention to is the data security issue[36]. Moreover, the digital database needs to take up very little space, but paper files need to take up much space to store it. Another advantage of the digital database is that its storage is very convenient as it can create a backup anytime and anywhere. On the other hand, when backing up paper files, more space is required.

iv. Classifying Data - 210018181

When we manage paper documents, we all encounter a common problem: how to classify data. The general method is to classify documents according to the time and content of the documents. Once the number of paper documents is large, this classification work will become very labour-intensive and time-consuming, and the paper document management system's data-sharing mechanism is very inefficient. When we want to inquire about some relatively confidential documents, we often need to apply for permission from the relevant responsible department. These application steps are also very cumbersome.

People usually need to go to different levels of managers to approve level by level, which causes the data sharing of paper documents to become inefficient. Moreover, the same paper document may have multiple versions, which will increase the difficulty of document management. However, the database management system does not have such shortcomings. It can automatically classify and quickly realize data sharing. For example, if people connect to the same database, they can share data and easily share data. The user sets the query authority, which can increase the security of the data and avoid the disclosure of some highly confidential files.

These are difficult to achieve with paper management systems, but because digital data can be tampered with, and minimal traces of the tampering are available to identify the authenticity of the data [37]. So, a popular trend now is to combine paper files with digital files and let them work together. The advantage of doing this is that we can digitize paper documents, which means that we can input the paper documents into the computer and then classify and manage the data. At the same time, we also need to keep the paper documents because paper quality documents can verify the authenticity of digital data. Let paper files and databases make up for each other's

shortcomings while improving data management efficiency; it also ensures the authenticity of the data.

V. Conclusion - 200029960

Overall, it is clear that technological advancements have considerably reduced the burden of security, accuracy, long-term preservation, and the classification of documents that pre-digital data management systems pose. However, it is essential to note that the issues have not diminished within modern database management systems but have allowed a more controlled approach to storing, accessing, and retrieving data. The industries explored brought to light the differences and similarities between modern database management systems and the pre-digital systems implemented within the library, banking, and healthcare sectors before digitalisation changed how they operated worldwide.

References

[1] American Library Association. Library association (1876-77). library journal. [EB/OL], 1876. https://babel.hathitrust.org/cgi/pt?id=hvd.32044011710068&view=1up&seq=5&skin=2021.

[2] Mohinder Partap Satija. *The theory and practice of the Dewey decimal classification system*. Elsevier, 2013.

[3] Matseliso Moshoeshoe-Chadzingwa. Library and information science text for africa, lo aina: book review. *South African Journal of Libraries and Information Science*, 71(1):102–103, 2005.

[4] Rahmatollah Fattahi. A comparison between the online catalog and the card catalog: some considerations for redesigning bibliographic standards. *OCLC Systems & Services: International digital library perspectives*, 1995.

[5] L Morupisi and NP Mooko. Using the online public access catalogue at the university of botswana. *Information development*, 22(3):197–204, 2006.

[6] Apurba Jyoti Majumder. Web dewey: The dewey decimal classification in the web. 2007.

[7] Education and Training Unit. Filing, 2021. https://www.etu.org.za/toolbox/docs/admin/filing.html.

[8] Go Cardless. What Is a Passbook and How is it Used? | GoCardless, 2021. https://gocardless.com/guides/posts/what-is-a-passbook-and-how-is-it-used/.

[9] Go Cardless. What Is a Bank Deposit Slip? | GoCardless, 2021. https://gocardless.com/en-au/guides/posts/bank-deposit-slip/.

[10] HDFC Bank. Withdrawal Slip: Know What is Cash withdrawal Slip and How to Fill it | HDFC Bank, 2021. https://www.hdfcbank.com/personal/resources/learning-centre/save/know-what-is-cash-withdrawal-slip-and-how-to-fill-it?__cf_chl_jschl_tk__=BPH2MB_VA3kYhJqjYurdXSMfJf_ql25_nLJpJhEPxp8-1637333033-0-gaNycGzNDZE.

[11] Adam Barone. Bank Statement Definition, 2021. https://www.investopedia.com/terms/b/bank-statement.asp.

[12] Richard Peterson. What is Normalization in DBMS (SQL)? 1NF, 2NF, 3NF, BCNF Database with Example, 2021. https://www.guru99.com/database-normalization.html.

[13] Andreas Hope. Security considerations in database operations, 2018. https://blog.dbwatch.com/security-concerns-in-database-operations.

[14] Spring. Spring Data, 2021. https://spring.io/projects/spring-data.

[15] Michael Teeling. What is Data Integrity? | Veracode, 2012. https://www.veracode.com/blog/2012/05/what-is-data-integrity.

[16] Richard Peterson. Dbms concurrency control: Timestamp & lock-based protocols, 2021. https://www.guru99.com/dbms-concurrency-control.html.

[17] 12. Deloitte. Embrace healthcare revolution. [EB/OL], 2020. https://www2.deloitte.com/content/dam/Deloitte/cn/Documents/life-sciences-health-care/deloitte-cn-lshc-embrace-healthcare-revolution-en-180914.pdf.

[18] R Haux, S Koch, NH Lovell, M Marschollek, N Nakashima, and K-H Wolf. Health-enabling and ambient assistive technologies: Past, present, future. *Yearbook of medical informatics*, 25(S 01):S76–S91, 2016.

[19] Peng-Ting Chen, Chia-Li Lin, and Wan-Ning Wu. Big data management in healthcare: Adoption challenges and implications. *International Journal of Information Management*, 53:102078, 2020.

[20] Gilad J Kuperman, Reed M Gardner, and T Allan Pryor. *HELP: a dynamic hospital information system*. Springer Science & Business Media, 2013.

[21] Karamjit Kaur and Rinkle Rani. Managing data in healthcare information systems: many models, one solution. *Computer*, 48(3):52–59, 2015.

[22] Kaspersky. It security economics in 2019. [EB/OL], 2019. https://go.kaspersky.com/rs/802-IJN-240/images/GL_Kaspersky_Report-IT-Security-Economics_report_2019.pdf.

[23] Nikita Tresa Cyriac and Lipsa Sadath. Is cyber security enough-a study on big data security breaches in financial institutions. In *2019 4th International Conference on Information Systems and Computer Networks (ISCON)*, pages 380–385. IEEE, 2019.

[24] Infosecurity. Top-secret anglo-french drone plans stolen. [EB/OL], 2012. https://www.infosecurity-magazine.com/news/top-secret-anglo-french-drone-plans-stolen/.

[25] Emma Johns. Cyber security breaches survey 2020. *London: Department for Digital, Culture, Media & Sport*, 2020.

[26] He Zhonglin and He Yuhua. A study on cloud backup technology and its development. In *International Conference on Information and Management Engineering*, pages 1–7. Springer, 2011.

[27] Marcel G Schaap and Feike J Leij. Database-related accuracy and uncertainty of pedotransfer functions. *Soil Science*, 163(10):765–779, 1998.

[28] Timothy D Ludwig and David T Goomas. Performance, accuracy, data delivery, and feedback methods in order selection: A comparison of voice, handheld, and paper technologies. *Journal of Organizational Behavior Management*, 27(1):69–107, 2007.

[29] Yohay Carmel, Dennis J Dean, and Curtis H Flather. Combining location and classification error sources for estimating multi-temporal database accuracy. *Photogrammetric Engineering and Remote Sensing. 67 (7): 865-872.*, 67(7):865–872, 2001.

[30] Behzad Soleimani Neysiani and Seyed Morteza Babamir. New labeled dataset of interconnected lexical typos for automatic correction in the bug reports. *SN Applied Sciences*, 1(11):1–6, 2019.

[31] Sameer S Pradhan, Eduard Hovy, Mitch Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. Ontonotes: A unified relational semantic representation. In *International Conference on Semantic Computing (ICSC 2007)*, pages 517–526. IEEE, 2007.

[32] Frederick Thomas Carson. *Effect of humidity on physical properties of paper*, volume 445. US Government Printing Office, 1944.

[33] John Bogaard and Paul M Whitmore. Explorations of the role of humidity fluctuations in the deterioration of paper. *Studies in Conservation*, 47(sup3):11–15, 2002.

[34] EL Graminski, Edwin J Parks, and EE Toth. The effects of temperature and moisture on the accelerated aging of paper. 1978.

[35] Adriaan BJ Du Plooy. The influence of moisture content and temperature on the aging rate of paper. *Appita Journal*, 34(4):289–292, 1981.

[36] Silvana Castano, Mariagrazia Fugini, Giancarlo Martella, and Pierangela Samarati. *Database Securi*. Addison—Wesley, 1995.

[37] Jing Han, Ee Haihong, Guan Le, and Jian Du. Survey on nosql database. In *2011 6th international conference on pervasive computing and applications*, pages 363–366. IEEE, 2011.