MAJOR CHALLENGES AND OBSTACLES FACING DIGITAL REPOSITORIES IN THE DIGITAL ERA

Shahad Ahmed Abdulgaffar

Information Science Department, King Abdulaziz University
Jeddah, ***SAUDI ARABIA***

ABSTRACT

In light of the technological and digital revolution and heavy dependence on digital resources, digital data saving has become a vital means that helps to keep and save intellectual productions. In fact, the digital repositories are one of the most important devices and records that contribute to the perpetuation of intellectual and scientific life. This paper aimed at shedding light on the challenges facing digital repositories, especially in light of the immense and continuous increase of data. It tries to identify those challenges and determine the solutions that can contribute to face them.

The lack of independent and detailed studies that tackled the challenges and focused on practical solutions was the reason behind conducting this paper. It targets collecting the related intellectual production and identifying the shortcomings and what is needed to be studied in the future, especially in light of big data and its impact on the continuity and development of digital repositories.

The survey methodology was used to gather studies and related literature and thus generalize the benefit and try to propose solutions. 15 Arab and foreign studies were covered which is estimated to contribute to achieving the goal of identifying the challenges facing digital repositories.

After reviewing Arab and foreign related literature and studies, the main results that came out of this paper, especially in light of the huge data, focused on the following challenges of digital repositories: intellectual property rights, the lack of qualified staff for content management, and huge budgets. A number of studies have suggested an approach to utilize modern technology and cloud computing to solve storage and saving problems. Others, however, considered cloud computing itself as one of the challenges of digital repositories. Some studies have emphasized the need to define a unified global policy for building digital repositories in light of Data ever-growing enormous size and the information and technical revolution, since the ultimate challenge was not due to technical and infrastructure problems.

This paper recommended the necessity of conducting specialized studies in finding applicable solutions to address the challenges and the optimal use of fuzzy

logic and artificial intelligence techniques, the reduction of the dependence on the human component, and the reduction of the budgets of the salaries and training of employees to make them restricted only to building a strong infrastructure.

Keywords: *Digital Repositories, Big Data, Cloud Computing, Intellectual Property, Virtual Archiving, IRB Scientific Research Ethics Board, Digital Repository Challenges*

INTRODUCTION

In light of the technological and digital revolution and heavy dependence on digital resources, digital saving is the means that helps to preserve intellectual production so as many of the hard printed materials copies would be saved in digital form after their digitization. In fact, these materials live in libraries, museums or archives in environments exposed to climate fluctuations, rot, pests and light. To be properly preserved, digital and digitized materials require special building and management. In this context, the idea of the digital repository emerged first by the creation of "Arxiv" site, which is the specialized repository in the field of physics allowing exchanging views on research papers and scientific articles before publication in 1991 A.D. Three years later, "Cop prints" appeared in the year 1997 A.D for cognitive sciences, languages and philosophy, with the aim of collecting intellectual production before publication as a solution to remove the obstacle of material subscriptions that hinder sharing intellectual production [1]. It is a database available on the web that attracts multiple types of intellectual production of science researchers in various forms of digital materials about a specified subject or for an institution. It is saved, organized and broadcast without physical restrictions, and with a minimum of legal restrictions for researchers [2].

Moreover, since the current era has become an overwhelming source of data, with an increasing speed in data size, tens of millions of people have been in direct contact with billions of devices. Trillions of transactions operate in unimaginable amounts of data, which leads to generating an equivalent amount of Data by people, and most of what we do is greatly influenced by our ability to access massive amounts of data, whether it is online or through our computers or mobile phones, all of which are called big data [3].

Accordingly, this paper aims at revealing the implications of big data on digital repositories, and the challenges they face in light of the information and communication technology revolution. Henceforth, the study problem tries to answer the following question: "What are the most prominent challenges of digital repositories, especially in light of Mega data?" Its goal is to identify the challenges that digital repositories faced in light of this big data. Indeed, this topic is of

particular importance in clarifying the challenges and in trying to avoid and restrict their number since the expected growth of data from all kinds of sources raises concern that the current digital storage systems and publishing will be unable to analyze and deal with them and provide access to them [3]. In addition, and as the previous reference explained, in 2003 people all over the world produced only 5 billion Gigabytes of data, but in the period from 2003 to 2012, 7.2 trillion Gigabytes were produced, and according to the study, it currently produces constantly and increasingly about five billion Gigabytes every two days. This confirms the importance of conducting studies on clarifying the challenges and coming up with solutions and recommendations. The topic requires really a very large amount of studies and explanatory research especially in relation to the small number of studies in general and the scarcity of Arab studies in particular.

This motivated us to prepare this paper trying to search for references and to collect information in order to generalize the benefit.

This paper used the method of surveying different theoretical resources represented by books, references, Arab and foreign scientific periodicals in addition to studies, research, Arab and foreign scientific theses as well as official seminars and reports related to the subject of the study. This would create a scientific background for the study so as it would become a source for collecting data and information in the period between 2010 and 2020 A.D.

After surveying Arab and foreign related literature and studies, the main results that this paper reached, especially in light of the huge data, focused on the following challenges of digital repositories: intellectual property rights, the lack of qualified staff for content management and huge budgets. A number of studies have suggested an approach to utilize modern technology and cloud computing to solve storage and saving problems. Others, however, considered cloud computing itself as one of the challenges of digital repositories. Some studies have emphasized the need to define a unified global policy for building digital repositories in light of Data ever-growing enormous size and the information and technical revolution, since the ultimate challenge was not due to technical and infrastructure problems. The main recommendations were the necessity of conducting specialized studies to find applicable solutions to address the challenges and the optimal use of fuzzy logic and artificial intelligence techniques; the reduction of the dependence on the human component; and the reduction of the budgets of the salaries and training of employees to make them directed only to building a strong infrastructure.

This paper consists of five parts: - reviewing a number of literature and previous studies related to the topic. - clarifying the problem. - reviewing the results of the related survey study. - introducing the results, conclusion and recommendations. - proposing topics for future studies.

LITERATURE REVIEW

A number of Arab and foreign studies related to the current study will be reviewed. They addressed the digital repositories and their challenges, especially in light of big data. In fact, it is noticed that there are few studies that focused on the challenges faced by the digital repositories in light of big data in general with a scarcity of Arab studies in the period between 2010 -2020 A.D. They are sorted from the most recent to the oldest.

A study [18] entitled "Managing Multimedia Big Data: Security and Privacy Perspective, 2020" which is one of the most recent studies as it was released in March of the year 2020. that clarified the most important big data challenges in light of the diversity of the data (Multimedia) namely external threats or malicious attacks and privacy. It focused on clarifying challenges, especially in the security aspects, and recommended more focus on security aspects in light of big data.

A study [4] entitled "Big Data Redux: New Issues and Challenge Moving Forward, 2019" followed the survey method of more than 520 studies using Google Scholar, and the analysis process lead to the fact that the biggest challenge in light of big data is data ownership. It clarified in a special way the problem with the service provider's validity in transferring data from one repository to another without the user's knowledge or approval. The study revealed that until 2013 there was no ideal solution for managing big data and then many developments were achieved in terms of storing big data and retrieving it, but there are still challenges even with technical developments in big data management. The study showed that one of these most important challenges is the validity of data, privacy and intellectual property. The study concluded that more attention is required to address analytical challenges and find better ways to solve complex problems that need new approaches to process large and more complex data.

The study [5] entitled "Application of Big Data Technology to Library data, 2018" followed the survey method and made it clear that the increasing nature of data requires a study on how to combine big data and technologies. The study was prepared for reviewing the research conducted in the use of huge data technology. It mentioned the works that were studied and their problems and challenges in the light of the big data. Among those works were the digital data repositories in which data is stored, archived and indexed. It came out by identifying a number of challenges and recommendations about the need to conduct more and more research on facilitating exploration and organizing information in new and accessible ways to big data.

In a study [6] entitled "Qualitative Data Sharing: Data Repositories and Academic Libraries as Key Partners in Addressing Challenges, 2018" there

was a focus on how to share digital repositories and academic libraries with researchers despite the challenges of intellectual property, ensuring legal and ethical participation, and sharing big data.

It concluded that some of the challenges associated with sharing qualitative moral and legal data and coming up with a number of solutions that could be addressed, including providing guidance for working with the IRB and explicit regulations in data sharing to meet the challenges.

In a study [7] entitled "DABAR - the National Infrastructure for Digital Repositories, 2017" the focus was on the national infrastructure of digital repositories for a number of research and higher education institutions in Croatia to build a digital repository DABAR (Digital Academic Archives and Repositories) to construct safe, reliable and interoperable repositories in a simple way. As it allows the management of its digital assets without the need to deal with technical issues related to the implementation and maintenance of digital repositories, the study clarified the technical challenges facing the construction of digital repositories. It demonstrated that DABAR has been recognized as a reliable, safe and user friendly national infrastructure for the creation and maintenance of interoperable digital sustainable repositories.

The study [8] entitled "Open Source Software for Digital Preservation Repositories: A Survey, 2017", surveyed eleven open source projects, including digital repositories, and discussed solutions that form the basis for overcoming challenges in data saving. It reviewed the latest findings of digital repository software in data preservation and saving, and recommended raising awareness about digital saving strategies in view of the importance of information and increasing the need for access to it, especially in light of the development of software solutions from highly specialized repositories into general purpose repositories. Consequently, storing different types of data in large quantities, as solutions for digital preservation of digital repositories were recommended through the implementation of new saving strategies.

The study [9] entitled "Institutional Digital Repositories at the Algerian University: Preparing a Mechanism for Building and Implementing the Digital Repository of the University of Constantine 2 Abdel Hamid Mehri, 2017", explained and diagnosed the current situation of information institutions and their beneficiaries and their related partners. It specifically aimed at examining and diagnosing institutional digital repositories in Algerian universities and included in the open DOAR repository directory, the study highlighted the challenges facing digital repositories.

It came out with results, most notably the lack of policies for digital repositories such as the policy of saving, content management and copyright and

metadata management, and recommended the need to develop policies for saving and content management.

The study [10] entitled "A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools, 2016" is a practical survey that aims at exploring the potential impact of big data challenges, especially on databases such as digital repositories. It came up with a system to deal with big data, and recommended researchers to develop solutions to identify challenges and take advantage of the various techniques used in statistical analysis, machine learning, data mining, cloud computing, and data stream processing.

The study [11] entitled "The Challenges of Big Data, 2016" aimed at clarifying the challenges in the presence of big data, especially in digital repositories. It came out with the need to increase the storage space and use of cloud computing technologies regardless of whether the data is inside the repository or is retrieved from an external repository with a large bandwidth of the network and thus address the storage problem and facilitate the transfer of data.

The book [12] entitled "Digital Repositories in Saudi Universities, 2016" explained the challenges facing digital repositories and mentioned intellectual property and cost, and then came out with projects to develop digital repositories in light of big data through the use and provision of policies and guidelines to enrich the repositories.

The study [13] entitled "Managing Big Data Issues within a Research Data Repository: Dealing with the 21st Century Data Explosion, 2015" discussed managing big data in data repositories. It clarified the challenges facing repositories in light of big data represented by scientific, cultural, ethical and technical challenges. The study came out with solutions for managing big data in digital repositories such as open archiving and virtual archiving. However, it clarified that full saving repositories are still facing the challenges of low human and technical resources in light of the constant demand for information.

The study [14] entitled "Major Challenges and Solutions for Utilizing Big Data in the Maritime Industry, 2015" clarified key aspects of the challenges identified by reviewing the developed institutions and discussing and providing possible solutions to them. It concluded that the essence of the problem is the lack of clear rules in terms of rights and responsibilities regarding the processing of big data. It came out by dividing the data in the digital repository into two sections: one for private usage and one for general usage. It recommended developing guidelines for enhancing data exchange and cooperation between companies, and eliminating concerns about data ownership as one of the most important challenges.

The study [15] entitled "Big Data: Challenges and Opportunities for Digital Libraries, 2015" is an inductive study on the concept of big data and its uses and challenges. It discussed the problem of capturing big data in digital repositories and the challenges in storing big data from institutions to the level of individuals. It clarified that the most important challenges are related to privacy. The study concluded that technology shifted its challenges from big data to the challenges of working in the cloud and artificial intelligence.

The study [16] entitled "Big Data Challenge: a Data Management Perspective, 2013" reviewed the challenges of big data from multiple axes. It presented a brief overview of research and problems related to big data. The study came out with solutions and answers that help in making use of big data and in reducing challenges and barriers. The main challenge was how to operate machine learning algorithms, traditional statistics when querying, and indexing data. Deep analysis of big data is also an important challenge. It recommends conducting a number of studies on big data in order to meet the challenges.

The study [17] entitled "Retooling Libraries for the Data Challenge, 2010" shows that digital libraries and digital repositories, especially institutional platforms regardless to how much they are suitable for research purposes, are at risk of embarrassing errors in retrieval when dealing with big data. It clarified a number of challenges to face shrinking budgets and staff. It recommended paying attention to researchers in the context of big data and emphasizing their role in the process of searching for information and trying to find a solution to meet the challenges.

PROBLEM FORMULATION

The current era has witnessed an information and technology revolution with an increasing speed in the level of big data. Tens of millions of people are in direct contact with billions of devices, and trillions of transactions are operating in unimaginable amounts of data, which generate equivalent amounts of data by People.

Most of what we do is greatly affected either by our ability to get access to massive amounts of data online or through our computers or mobile phones, all of which are called data [3]. And since the digital repositories [12] are a database that both contains the research works deposited by researchers and works to make them available and saved for a long-term according to their type, or whether they are specialized in a specific subject or institutional or national repositories, it was necessary not only to identify the challenges that these repositories faced in Big data but also to conduct studies trying to tackle these challenges and find solutions.

Hence, the problem of the paper that this survey study seeks is to reveal the implications of big data on the digital repositories, and the challenges they face in light of the information and communications technology revolution. The problem of the study is trying to answer the following question: "**What are the most important challenges of digital repositories in the light of big data?"** To do so, the following sub questions were answered:

- What is the main challenge that all studies unanimously agreed upon?
- What are the most prominent solutions?
- What are the deficiencies that were observed out of this survey study according to the point of view of the researcher?

THE RESEARCH METHODOLOGY

A descriptive analytical survey approach was used. It describes and explains what exists through the case study method to analyze the causes and effects of its suitability to the nature of the study. This will be analyzed using theoretical references represented by Arab and foreign books, references and scientific periodicals in addition to Arab and foreign studies, research and scientific theses as well as seminars and official reports related to the subject of the study. This would not only create a scientific background for the study but also use it as a primary source for collecting data and information. This task will be done over the period from 2010-2020 A.D.

After reviewing the theoretical references, it is noticed that:

In the study no [4], a comprehensive concept of big data was discussed, which is large amounts of data that exceed the ability of technology to store, manage and efficiently process using traditional methods and tools of databases and repositories. It has major dimensions, which are size, speed and diversity. It clarified the most important challenges as follows:

- Storage and transportation.
- Content management.
- Intellectual property.
- Security.
- Privacy.
- Operating costs and expenses.
- Transfer data from one repository to another without the user's knowledge or consent.
- Data correctness.

The study also clarified the emergence of a new challenge, which is cloud computing, and noted the similarity of studies in depicting the challenges of digital repositories. It concluded that more attention is needed to address analytical challenges and find better ways to solve complex problems that require new approaches to process large and more complex data. As big data continues to grow, problems and challenges will continue to grow significantly.

The study was characterized by shedding light on one of the solutions to face and reduce challenges, namely building high-performance software and systems that can produce effective results, but did not come out with any concept of a proposed system.

The study [6] was similar to studies no [4, 12]. It demonstrated that the most important challenges are intellectual property, ensuring legal and ethical participation, and sharing big data. It was characterized by a solution proposal to address some of the challenges associated with sharing qualitative moral and legal data and came up with a number of solutions like providing guidance to work with the IRB and including explicit regulations in data sharing to meet the challenge. This would be achieved through changing any formal rules, educating actors from different scientific fields, and coordinating their efforts in specific projects. It made a proposal to prepare a series of workshops that bring together IRB employees from research universities, magazine editors, public and private financiers, and representatives of social science societies to discuss how they can share ethical human data throughout the research life cycle. In this way, it achieves the educational role and builds bridges that enable data repositories and other academic libraries to achieve their missions.

The study [5] dealt with the impact of big data on information databases including central data repositories that store, save and index data.

It came out by identifying the following challenges:

- **Lack of data analysts:** Data analysts not only need statistics and computer science skills, but also field knowledge skills and the ability to collaborate. Therefore, one of the most important challenges is the ability to manage big data information. The training through short courses may not be sufficient.
- **Ability to utilize big data:** Big data faced a challenge with big companies, and more than half of the organizations were unable to deal with big data due to the lack of staff.
- **Big data searching** is much slower than searching for other disciplines.
- **Budget issues** caused by the fact that investment in high-performance information technology is required. However, much of the research data that was produced ten years ago still needs digitization and is not a simple task since it requires a lot of time and employees.

- **Technical challenges** linked to technologies such as data capture, storage, processing, and presentation. A large set of data often contains some false data. Therefore, removing this data requires some effort. Because of the heterogeneous search data types and formats, their integration becomes very difficult. It requires having algorithms or processing operations performed on them. It needs a budget for building tools and providing other support as well.
- **Privacy:** Big data explores data and discovers knowledge. Consequently, new risks to system breaches may arise due to the possibility of accessing a large amount of data.

It recommended that more and more research is needed on facilitating prospecting and organizing information in new and accessible ways in light of big data. It is consistent with a number of studies on the most challenging points facing digital repositories in light of big data such as studies no [9, 12, 15].

The article [7] tried to clarify a model for designing national infrastructure for digital safe, reliable and interoperable repositories in a simple way. 109 digital repositories were created in DABAR by research and higher education institutions in Croatia.

DABAR has been recognized as a reliable, safe and user-friendly national infrastructure for the creation and maintenance of interoperable and sustainable digital repositories. It also helps reduce development and maintenance costs. It provides interoperability for all of its repositories with national and global electronic infrastructures. This model is applicable outside the science and higher education system. For example, it can be used in the primary and secondary education system as a repository of educational content or as a commercial platform for hosting digital repositories.

The study showed that the challenge facing the DABAR team and society is to further enhance open access and support institutions in setting prerequisites for publication in open access. This study solved the problem of cost and maintenance since it still considers availability and property rights a challenge.

In the study [9], the challenges of developing digital repositories are explained by the lack of a building pattern for digital repositories and content management that can be globally applied. In fact, all experiences are different and inherent in the enterprise environment and the literature is full of processing patterns and case studies on the development of local institutional digital repositories. Given that there are differences between the repositories with regard to programs, policies, types of institutions and user needs, the repository environment is very complex and not comparable.

At this time, best practice guides can guide the repository preparation process, and each organization can choose its own policy or adopt and re-use the experiences of other repositories. The study recommended the necessity of developing policies related to saving, content management, quality control and metadata.

The most prominent challenge that the researcher came out with is the lack of a content management policy in the repositories in light of big data. The need to develop policies is stressed in relation to availability and saving through establishing an advisory group that helps in taking decisions, designing a unified policy, and planning policy through:

- Studying the economic and organizational feasibility of the repository and determine the objectives of the repository.
- Providing opportunities to exchange information and knowledge on the local, national and international levels.
- Providing better access points for information sources.
- Defining the contents, acquisition policy, deposit, filing, regulation, quality, metadata and risk management.

The study agreed with more than one study about the fact that intellectual property is one of the most important challenges for digital repositories in light of big data such as the study no [12].

The study [18] explained that the most important issues, challenges, and threats are related to security, unauthorized access, and attacks. It recommended the need to follow the policy of encryption, especially with the diversity of data and access to multimedia.

The work done on [11] dealt with a distinct solution to overcome the challenge of big data and the difficulty of managing it with the technology of cloud computing which was mentioned as one of the challenges in the study no [4] and as a solution in the study no [10]. However, it also showed that there is a challenge in the cloud computing which is its management by humans.

The paper [10] explained the challenges that affect digital repositories under the big data namely storage and data analysis and flexibility in expansion. It recommended developing solutions to identify challenges and benefit from the various techniques used in statistical analysis, machine learning, data mining, cloud computing, and data stream processing.

The work [8] came out with a solution to face the challenges of storing big data in digital repositories, which is the use of the OAIS form where information is obtained from research articles and then stored in the determined place. For example, obtaining information about drugs and checking which are available in repositories as a database are kept in the designated category according to the

type of disease. In addition, it recommended focusing on the digital saving strategies that each system provides to ensure long-term access, to ensure the integrity and validity of the data stored through authentication features and to set access control mechanisms and the ability to track user actions via data. it was characterized by clarifying the proposed model.

The book [12] stressed the challenges facing the digital repositories in light of the increasing level of data, with what is intended from each of the challenges according to the following points:

- **Cost and the need for a budget for employees:** the term cost was addressed as a challenge, but the intention is the budget of the repository staff, salaries and needs.
- **Content producing difficulty:** That is, the difficulty of encouraging researchers to deposit their work voluntarily, and therefore the repository needs to bear the costs in order to encourage researchers to deposit.
- **The need for budget and ongoing support:** In light of the big level of data the digital repository needs continuous support by the repository management.
- **Intellectual property rights:** With the huge data, the digital repository may violate intellectual property.
- **Need for an incentive for researchers:** Researchers need incentives to disclose their bibliographic data for their scientific production.

The researcher emphasized the need to protect intellectual property through cooperation with the project to save and make available research (SHERPA). The latter aims at helping to resolve intellectual property issues, quality and metadata. In fact, copyright is a central issue for researchers accepting to publish in digital repositories. Therefore, providing guidelines and a clear deposit policy is one of the elements that underline repository developers' interest in intellectual property issues.

It came out with a management mechanism:

- Obtain permission from the original publisher of the work (copyright removal).
- Clarify the deposit policy.
- Clarify general deposit rights.

This study was characterized by identifying the most important challenges and finding solutions, especially in the intellectual property barrier.

The article [15] tried to clarify the challenges facing digital libraries and databases represented by digital repositories according to the following:

- The need to train staff.
- Need for a budget.
- Privacy.

Moreover, it focused on capturing big data in digital repositories with a huge storage level of data and hundreds of millions of mathematical models in addition to individual businesses where algorithms and fuzzy logic can be used to facilitate retrieval. However, the study did not address in details how to use these algorithms, or gave a detailed explanation of how to deal with big data, nor did it provide any proposed visualization or detailing of any of the challenges mentioned.

In [14] the author identified the major aspects of the challenges through reviewing the developed institutions and discussing and providing possible solutions to them. It concluded that the essence of the problem is the lack of clear rules in terms of rights and responsibilities regarding processing big data as well as:

- Privacy problems
- Data monopoly
- Handling and analyzing data
- Data monitoring

In this, it was similar to studies no [4,5,6,7,12,15]. It came out by dividing the data in digital repositories into two parts: one for private usage, and one for general usage. It recommended developing guidelines to enhance data exchange and cooperation between companies and remove concerns about data ownership as one of the most important challenges. However, it did not clarify those principles.

The study [13] was one of the best studies in terms of coverage, as it dealt with *Challenges:*

- Access to all other data sets within the repository.
- Cost.
- Human intervention.
- Storage and data management tools.
- Ethical and technical challenges.

Solutions:

- Open archiving: the process of depositing an author or scientific research in digital form in one of the available repositories.
- Virtual archiving: It is the speed of publishing and making available scientific research, which is achieved in light of its immediate and virtual

archiving, and this is in contrast to the process of publishing in academic scientific journals that usually take 12 months or more.

- Carefully create a sorting system to define the path to which specific data should be directed as part of the saving and distribution process.

The Study no [16] is the only study among the reviewed studies that addressed the challenges of dealing with big data and repositories with advanced technology other than cloud computing that was addressed in a number of other studies.

The Study [17] dealt with the same challenges treated in the studies no [4, 15], which are budgets and employees. The study did not provide any perception except that it recommended paying attention to research operations in light of big data and confirming the role in the process of searching for information and trying to find a solution to meet the challenges.

ACHIEVED NUMERICAL RESULTS

After answering the main question and sub-questions and after using the survey method, the results of this paper were as following:

- 15 studies were consulted during the period from 2010 to 2020, as shown in the following graph:

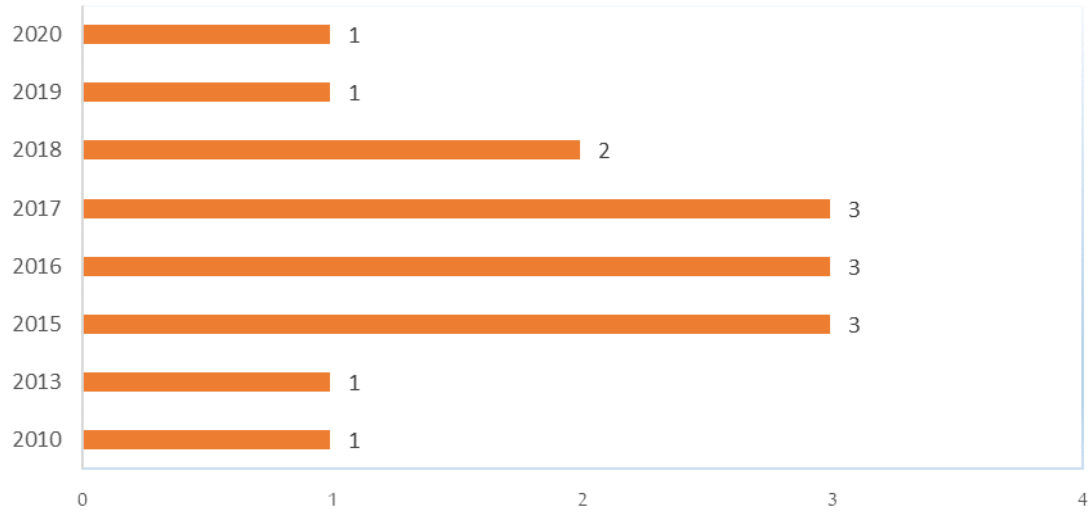

Figure 1. *Reviewed references by year*

- The studies agreed on the following challenges that were classified according to their frequency in the studies:

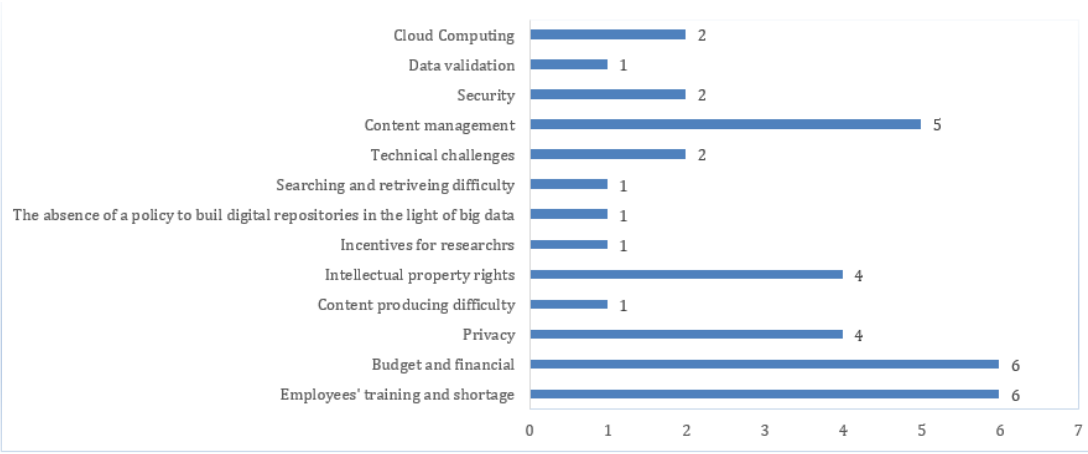

Figure 2. *The most prominent challenges*

- It is noticed from Figure 2 that most studies have agreed that the shortage of trained and qualified staff to solve the problem of managing digital repository content in light of big data is still the main challenge. Then, came the need for budgets and material support, intellectual property rights and privacy as obstacles and challenges for digital repositories.
- We notice that the technical challenges are not significantly mentioned which may be due to the technical revolution in this era. ▪ Some studies came out with solutions as shown in the following figure:

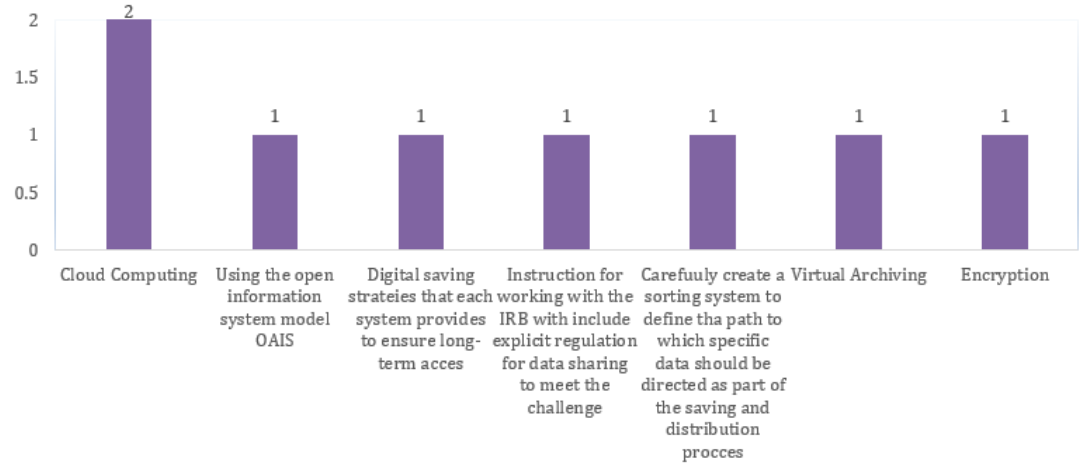

Figure 3. *The most prominent solutions*

- According to Figure 3, two studies agreed on the need to apply cloud computing as one of the storage and archiving solutions in digital repositories in light of big data.

- As far as the researcher knows there is almost a total lack of studies that dealt with the challenges of digital repositories in light of big data, and it may be mentioned as one of the examples and models in some studies.
- As far as the researcher knows there is a Lack of studies that came out with an actual applicable system, and most of them presented recommendations and proposals for implementing mechanisms that contribute to solving the challenges of digital repositories in light of big data.
- Among the possible solutions that we can come up with:
  - To solve the problem of intellectual property, which was mentioned in one of the studies, is having agreements with the Council of Ethics for Scientific Research to solve the problem of intellectual property.
  - One of the possible solutions is to set global standardization policies for building digital repositories in light of huge data.
  - One of the solutions, mentioned in some studies, is to hold training workshops to train the largest possible number of employees and thus solve the problem of the shortage of staff.
  - Follow the virtual archiving method.
  - The size, type and speed are only highlighted in a simple paragraph related to Big Data as far as the researcher knows.
- Two studies came out with two systems, namely:
  - The Open Archives Information System (OAIS) where information is obtained from research articles and then stored in the determined place. For example, obtaining information about drugs and checking their availability in repositories as a database that is saved in the determined category according to the type of disease.

Figure 4. *Open Archives Information System (OAIS) [8]*

  - Typical workflow system for big data project.

Figure 5. *Typical workflow of big data [10]*

CONCLUDED COMMENTS AND RECOMMENDATIONS

This paper aimed at identifying the challenges facing digital repositories in light of big data not only to contribute to finding appropriate solutions by programmers and specialists but also to benefit from the characteristics and advantages of digital repositories in saving intellectual production and facilitate access to it.

Thus, we came out with the following conclusions:

- The great similarity among researchers in the reviewed studies about the most important challenges.
- Intellectual property rights are a major challenge with the technical and information revolution.
- The challenges that were referred to in the studies differed according to years with the exception of the challenge of lack of qualified employees and the budget which have continued since the oldest study in 2010 A.D.
- The term cloud computing appeared in studies from 2016 onwards.

However, there are some remarks that should be mentioned:

- The paucity of Arab studies that addressed the challenges of digital repositories. They were restricted just to the presentation of successful models or statistical surveys.
- The studies as far as the researcher knows there is lack the practical part that can be taken and implemented.
- There are no studies that have clarified the challenges of digital repositories in light of big data in one form, but it is discussed as an example or an axis as far as the researcher knows.

Recommendations

The following recommendations were made:

- The need to complete specialized studies in addressing the challenges of digital repositories in light of big data, according to actual systems that can be applied.
- Take advantage of technology such as artificial intelligence, and fuzzy logic to solve the problems of the lack of qualified employees and thus reduce the need for employees and budgets for training and salaries.
- Benefiting from cloud computing as a solution to saving and storage problems.

REFERENCES

1) Ahmed, Walaa. Digital Repositories, available at: http://w-ahmed17. blogspot.com/2015/11/blog-post_29.html, accessed: 24-3-2020.
2) Omar, Iman (2011). The Origins and Evolution of Open Digital Repositories, available at: http://www.journal.cybrarians.info/index.php?option=com_content&view=article&id=607, accessed: 7-3-2020.
3) Al-Tayyib, Zainab, Al-Riba'e, Al-Suleiman (2018). New information specialist roles for handling big data, available at: https://www.qscience.com/docserver/fulltext/jist/2018/2/jist.2018.16.pdf?expires=1585050506&id=id&accname=guest&checksum=42454D7AF0D7020BA9719BCCD50C644E, Accessed: 24-3-2020.
4) Espinosa, J. Alberto; Kaisler, Stephen; Armor, Frank; Money, Wiliam H. (2019). Big Data Redux: New Issues and Challenge Moving Forward, Proceedings of the 52nd Hawaii International Conference on System Sciences, Available at: URI: https://hdl.handle.net/10125/59546, Received: 23-3-2020.
5) Kaladhar, A.; Naick, B.R. Doraswamy; Rao, Somsaekhara (2018). Application of Big Data Technology to Library data: A review, International Journal of Library and Information Studies, Vol. 8 (2) Apr-Jun, 2018 ISSN: 2231-4911.
6) Mannheimer, Sara; Pienta, Amy; Kirilova, Dessislava; Elman, Colin; Wutich, Amber (2018). Qualitative Data Sharing: Data Repositories and Academic Libraries as Key Partners in Addressing Challenges, American Behavioral Scientist 63 (5), 643-664
7) Celjak, Drazenko; Bekic, Zoran; Cundekovic, Marko; Jertec, Ljitjana, Jertec; Millinovic, Jasipa, Zubic, Alen (2017). DABAR - the national

infrastructure for digital repositories, University of Zagreb, University Computing Center, Available at: URI: https://www.semanticscholar.org/paper/DABAR-the-national-infrastructure-for-digital-Celjak-Bekic/197730c946f04f23bd38747426668f6b11b587f2, Received: 28-3-2020.

8) Rosa, Calos; Craveiro, Olga; Dormingues (2017). Open Source Software for Digital Preservation Repositories: A Survey, International Journal of Computer Science & Engineering Survey (IJCSES) Vol. 8, No. 3, June 2017.
9) Ahsan, Babouri; Nabil, Aknoush (2017). Institutional digital repositories at the Algerian University: Preparing a mechanism to build and implement the digital repository of the University of Constantine 2 Abdel Hamid Mahri, Iraqi Journal of Information, vol (18), p (1,2), pp. 132-168.
10) Acharjya, D.; P., Kause (2016). A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 2, 2016.
11) Mardis, Elaine (2016). The Challenges of Big Data, The Company of Biologists Ltd | Disease Models & Mechanisms (2016) 9, 483-485 doi: 10.1242 / dmm.025585
12) Nagy, Salah (2016). Digital repositories of universities in Arab countries, Arab Center for Studies in Library and Information Sciences, Cairo, pp. 33-42, available at: http://acrslis.weebly.com/uploads/1/6/0/7/16070576/digital_repositories. pdf, accessed: 23-3-2020
13) McNally, James W. (2015). "Managing Big Data Issues Within a Research Data Repository: Dealing with the 21st Century Data Explosion, Inter-university Consortium for Political and Social Research (ICPSR), Annual International Conference on ICT: Big Data, Cloud and Security Conference, Available at: https://www.researchgate.net/publication/300053038_Managing_Big_Data_Issues_Within_a_Research_Data_Repository_Dealing_With_the_21st_Century_Data_Explosion/citations, Received: 23-3-2020
14) Koga, Sadaharu (2015). Major challenges and solutions for utilizing big data in the maritime industry, World Maritime University, Master of Science.
15) Hacken, Richard (2015). Big Data: Challenges and Opportunities for Digital Libraries, Brigham Young University, nternational Conference on Computing in Engineering and the Sciences, Available at: https: //scholarsarchive.byu.edu/facpub/1580, Received: 23-3-2020

16) CHEN, Jinchuan; CHEN, Yueguo; DU, Xiaoyong, LI, Cuiping, LU, Jiaheng; ZHAO, Suyun; ZHOU, Xuan (2013). Big data challenge: a data management perspective, Comput. Sci., 2013, 7 (2): 157-164, DOI 10.1007 / s11704-013-3903-7.
17) Salo, Dorothea (2010). Retooling Libraries for the Data Challenge, Ariadne Issue 64, Available at:
http: //www.ariadne.ac.uk/issue64/salo/, Received: 23-3-2020
18) Kumar, Rajeev; Agrawal, Alka; Khan, Raees (2020). Managing Multimedia Big Data: Security and Privacy Perspective, Available at: https://www.researchgate.net/publication/339584595, Received: 31-3-2020.