MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE RUSSIAN FEDERATION
Federal State Autonomous Educational Institution of Higher Education
“South Ural State University (National Research University)”
School of Electrical Engineering and Computer Science
Department of Computers

THESIS IS CHECKED

Reviewer,
PhD, Associate Professor
___________ N.V. Plotnikova
“___”____________ 2021

ACCEPTED FOR THE DEFENSE

Head of the department,
PhD, Associate Professor
___________ G.I. Radchenko
“___”____________ 2021

DEVELOPMENT OF A SMART SYSTEM FOR PRESERVATION OF GOVERNMENT RECORDS IN DIGITAL FORM

GRADUATE QUALIFICATION WORK
SUSU – 09.04.01.2021.308-628.GQW

Supervisor,
PhD, Associate Professor
___________D.V. Topolskiy

Author,
student of the group KE-228
___________Al-Behadili Hameed Hassan

Normative control
______________ S.V. Syaskov
“___”____________ 2021 г.

Chelyabinsk–2021

Federal State Autonomous Educational Institution of High Education “South Ural State University (National Research University)”
School of Electrical Engineering and Computer Science
Department of Internet of Things

APPROVED
Head of the department
Dr. Sci., Prof.

_____ - ______ - 2021

TASK
of the master graduate qualification work
for the student of the group KE-228
Al-Behadili Hameed Hassan
in master direction 09.04.01
“INTERNET OF THINGS (Master Program “INTERNET OF THINGS”)

1.The topic (approved by the order of the rector from . . 2021, No.)

“DEVELOPMENT OF A SMART SYSTEM FOR PRESERVATION OF GOVERNMENT RECORDS IN DIGITAL FORM”

2. The deadline for the completion of the work: .06.2021.

3. Source of Data for the Work

3.1 Becker, H.S. and Geer, B. (n.d.). Participant Observation and Interviewing: a comparison. Human organization.

3.2 Bikson, T. K. and Law, S. A. (2016) Electronic information media and records management methods: A survey of practices in UN organizations, The Information

3.3 Chachage B. and Ngulube P. (2020). Management of business records in Tanzania: an exploratory case study of selected companies.

4. List of Development Issues

- Producing predictive ability different from the standard clinical methods currently in use.
- Limited amount of data with a wide range of architectural possibilities.
- Fracture risk prediction.

- Increasing performance on the target task by exploiting the trained knowledge of the source task.
- Complementing the ground truth dataset with existing property parcel information that could be indistinguishable based on the recorder alone.

Table of Contents

INTRODUCTION

In this thesis, robust smart system and recognition is performed which is an important part of vision-based form recognition for government records preservation. It is worth to say that the records are vital forms for human being to communicate and carried out many important tasks on a daily basis [1]. Record and other government parts tracking from digital records has revolutionized not only the scientific world but for entertainment like gaming industry and animation creation. It also took its place in traditional hard work on factory automation, virtual reality, rehabilitation purposes & disability support, performance measurement and many more [2]. Now, there are approaches where user can have unique hardware attached to its government which can have good analysis of the joints and geometry of some government parts. But it is neither natural nor comfortable. In addition, it is relatively expensive to wear specific hardware designed for certain tasks. Therefore, vision-based applications are preferred over those approaches for simplicity, cost effectiveness and natural interface experience [3].

In this research work, the main motive is to develop a smart system for the government of Iraq for the purpose of preserving the existing paper-based records to the digital records. Experts have understood the risks associated with sustaining a digital world for several years as mentioned in [4]. The goal of this study is to find ways that digital preservation activity can be communicated to researchers to improve the permanency of their research data as well as government records. The research focused on activities in governmental environment that result in the creation of sustainable information packages (SIP) for three dedicated research digital libraries i.e. electricity, finance, science and technology for digitalization as the approach being followed in [5]. Paper-based information packages are the input to the Open Archive Information System (OAIS) model. The OAIS is an International

Standards Organization (ISO) standard that constitutes the framework for the new digital environment will normalize and transform the paper-based data to the digital records.

Today, mediated communications and knowledge have converged onto digital formats that have been adopted by almost every society on the planet as mentioned in [6]. Although the implications of the digital transformation have been studied from many perspectives, its impact may not be fully understood for generations to come. We proposed a new digital model over four years of governmental records from 2015 to 2019, the accelerating rate of digital data growth has increased both the number of extraordinary opportunities for discovery and the risk of unintentionally removing priceless information sources from the archive of collective human knowledge with a keystroke as explained the process in [7]. While the technology to store data (hardware and software) has improved at an increasing pace, personal and organizational behaviors have been much slower to adapt to the new environment [8]. Data storage has become easier as the price of storage devices has plummeted while data preservation has become harder because of the task's magnitude and unseen, abstract nature.

People needs to use mouse, keypads and remotes to control and manipulate machines. This interface between devices and humans can certainly be improved in coming years. Several researches have been going on to modify existing tools and components which com- bine human and machine interface [9]. It can be thought of by giving commands to a computer or robot with speech and/or some forms. This is the way we humans conduct communication among ourselves and we can definitely extend this way of communications to machines as well [10]. This task is the small attempt in that direction. We would like it to understand human actions and commands with forms.

1. ANALYSIS OF THIS SUBJECT AREA

There are many related issues of record form recognition. The most widely recognized ones are:

- **Record Localization:** Means to decide the picture position of each single location of record.
- **Record Identification:** Plans to recognize the nearness and area of highlights, for example, lines, paragraphs, figures, sections and forms.
- **Record Recognizable Proof:** Thinks about an info picture against a database and reports a match, assuming any.
- **Record Verification Issue:** Checks the case of the normalization of record in an information picture.
- **Record Gauge Issue:** Gauge the area and conceivable direction of a record in a picture grouping.
- **Record Form Articulation Acknowledgment:** Concerns recognizing the full of states (To check if the lines on records changes over time in different records) of different organizations.

Obviously, smart system is the initial phase in any mechanized framework which takes care of the above issues.

1.1 REVIEW OF ANALOGUES

Smart system and recognition for movement of records using different machine learning techniques as it is most important task [11]. It can also be considered as in terms of human mimicking in this huge task for making the robot suitable for industry work. The idea is that, the robot should observe human record movements and be able to understand at some point what is going on around its surrounding [12]. That's why accurate and robust real time record movement detection is priority task for this project which can be the first step forward to proceed. Machine learning will eventually work side by side with humans in industry area where it has to understand record form and be able to repeat at some point of time [13]. It must have a very good insight about where humans' records are with respect to its world coordinates and also it must be able to detect an object for being operated upon. In this task, we are limiting our approach to human records tracking without any records in record [14]. It would be simpler approach where record movements are being observed by machine learning and with form or trajectory detection algorithm developed in the later part of this thesis would help to recognize record to understand that movement.

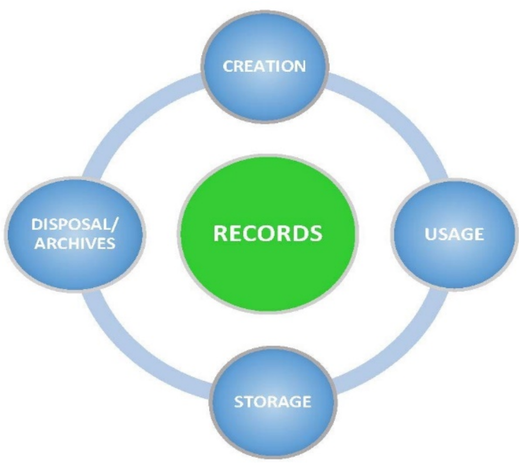

Figure 1.1. An example of four forms. Each form represents a label of every record that belongs to the smart system [15].

Record motion tracking algorithm locates records on frame by frame basis and continuously do so as they move in 3D space [16]. Most tracking algorithms consider object's shape size and color in to account. But once tracking is missed or lost, algorithms fails to find records and again needs to reinitialize [17]. Skin mask is an important feature for smart system which will be the indicator of records and face in the scene. Without clear skin mask, depth is not helpful to track records in this case. Although false skin segmentation vs tracking efficiency needs to be balanced perfectly; by putting strict thresholds would limit false detection but at the same time will have less information to process [18]. Hence efficiency of tracking algorithm will suffer. On the other record, if we keep thresholds too big then the noise and false skin indicator will increase error in detection.

Figure 1.2: All basic record forms take from source [15].

Record motion tracking algorithm locates records on frame by frame basis and continuously do so as they move in record space. Most tracking algorithms consider object's shape size and color in to account [19]. But once tracking is missed or lost, algorithms fails to find records and again needs to reinitialize. Skin mask is an important feature for smart system which will be the indicator of records and face in the scene [20]. Without clear skin mask, depth is not helpful to track records in this case. Although false skin segmentation vs tracking efficiency needs to be balanced perfectly; by putting strict thresholds would limit false detection but at the same time will have less information to process. Hence efficiency of tracking algorithm will suffer. On the other record, if we keep thresholds too big then the noise and false skin indicator will increase error in detection [21].

There are several attempts made for real time robust record tracking. This problem is quite old and many approaches has been developed to detect and understand record movement with many different devices [22]. Earlier people used to wear gloves and some special hard- ware to detect record postilion and used to interact with the world. The record camera was one of the simple solutions for complex hardware problem where people can extract skin information from color records with different chrominance models [23].

State of the art thesis include many record-based identification and tracking approaches. Representation with modeling, specifying key features for detection, contours tracking, or by using classification techniques such as neural network and binary tree classification [24]. Although machine learning algorithms and classifiers needs robust training with lots of data only then they are able to segment record from records. This approach needs lots of processing and hence computationally expensive [25]. The more traditional method where we use record form depth information and try to find record from those records combined; is an ideal approach for this task.

1.2 ANALYSIS OF BASIC TECHNOLOGICAL SOLUTION

Our system is designed to maximize user experience. Moreover, our system differs from existing systems in the following ways.

Just records. Many existing systems require users to wear gloves or markers to be capture by the camera. We feel this limits the usability of the application and aim to keep the prediction of the system entirely marker free.

Real-time. Our system should run smoothly on average modern computers with a dedicated graphics card. The system should also recognize record form at a high frame rate, nearing real-time speeds. Our desired frame rate that we want achieved a frame rate for record storage. In our design and implementation, one driving goal is to squeeze every millisecond as possible.

No calibration. Once trained, our system should require no further calibration before working in a new environment.

Robust and accurate. Our system should have an accurate estimation of where the user’s records are and what forms they use with low false positive. Moreover, the system should be insensitive to various background, user’s location, camera position and other noise.

Arbitrary forms. Our system should be able to easily incorporate new types of forms. With sufficient training, our system should support many complex forms such as those seen in the sign language.

1.3 CONCLUSIONS

Data storage and data preservation are two very different concepts. Storage is dependent on a “place” such as a disk drive while preservation is an “permanent storage achieves” dependent on storing the data over long period of time with secure mechanism or model as described in [26]. Preservation relies heavily on technology that allows users to store more data in more ways on increasingly smaller devices. As devices and storage media evolve, data are left on outdated media or file formats – bits locked in a bottle afloat at sea. This “digital amnesia” is certainly avoidable, but only if the data are properly attended. Threats to data preservation, such as media obsolescence, can be resolved through automation, but ultimately depends on human intervention to assess, describe, and prioritize digital artifacts [27]. Preservation implies “digital libraries” to maintain its readability, content, and meaning. Digitally stored information is an intangible asset whose cost to accumulate is inexpensive and becoming more so as technology progresses.

A system for real-time record collection. We design and implement a complete real-time record form recognition system based on machine learning. The system broadcasts the location and form of the record through a web socket server and can be used by other applications.

An inexpensive way to generate accurate labeled data for record form recognition. We use digital records to easily generate labeled data using the aligned depth records. In [28], the authors use sophisticated computer graphics to generate training samples. We found this expensive and through the use of color gloves, developers can generate their customized form without difficulty. Another advantage of our approach is that the system uses actual raw depth records as training samples. These naturally capture realistic noise such as shadows and hardware noise.

Using computer generated graphics as done in [29] it is very difficult to simulate these noisy effects. Note that the end users do not need to wear color gloves; they are only used in training.

A computational insight about data model for record preservation. To the best of our knowledge, there seems to be no literature in comparing digital from a computational perspective. We provide an in-depth complexity analysis of the two methods rather than merely reporting experimental accuracy as done in most machine learning literature.

Extensive experimental evaluations of the system. We conduct extensive experiments evaluation of the effectiveness of the random forest classifier by systematically exploring a large space of parameters. Interesting results lead to a deeper understanding of random forest

2. DEFINITION OF REQUIREMENTS

In [30] when the author formalized and implemented the model for preservation of digital records, the model would establish institutional trust through its credibility (the perception that professionals are in charge), reliability (outcomes of professionals being in charge), intimacy (acceptance without value judgment) and self-orientation (not self-serving). Building trust models and developing responsibilities across government level within one country will be a significant challenge. Expanding these models beyond national boundaries creates many additional problems. Governments from technologically advanced countries have also been very interested in preserving their own digital contributions and heritage to society as a whole. In particular, England, France, New Zealand, and Australia recognized the issues of preserving knowledge in its digital form for future generations as described in [31]. These early-adopter countries are developing guidelines to help their nations deal with the issue. Each recognizes that the solution cannot rely on technology alone. The OAIS is one of several international (ISO) standards that attempt to resolve problems in an increasingly connected world. Focusing on the average best validation error, two statistics seem to correlate with its absolute improvement. On one record, the improvement seems to shrink with the number of classes and thus the number of original samples in the training set. As the number of selected classes is increased from seven to ten, the difference becomes smaller and even negative. However, even though relatively small, the classification accuracy is higher when the data augmentation is used for twenty classes as well as the whole dataset.

On the other record, the data augmentation becomes more effective for subsets with larger maximum sequence length. While there was a similar amount of original training samples in the three five-class subsets, the maximum sequence length strongly varied. The largest absolute improvement was achieved for the classes where the maximum sequence length was the highest [32]. While the maximum length was equal in the experiments with seven, eight, nine and ten classes, where a shrinking improvement was observable with increasing training set size, the maximum sequence lengths are significantly larger when the experiments are repeated on the classes. In these cases, the data augmentation has positive effects again. As the maximum number of frames gets clipped to smaller numbers for the created sequences, the synthetic samples become less helpful for improving the classification accuracy on the whole dataset. The classification scores are all smaller than the original baseline when the synthetic sequences have a maximum length of only 100, 200 or 500 frames.

These high-level observations do not allow definite conclusions and it seems likely that the effectiveness of the data augmentation depends on many other factors, foremost the shape, type and variations of the selected classes [33]. That is why, it is difficult to determine a definite limit for which the augmentation method starts to fail. However, it is intuitive that the data augmentation is more advantageous on smaller sets which can be more easily enriched with new information. While the visual quality of the synthetic samples is negatively affected by the smaller training set, the generated sequences appear to carry enough information to have a beneficial impact. Moreover, a classifier's proneness to overfitting rises with decreasing amount of training samples, increasing the need for any regularizing input.

Digital technology has been available for over half a century. The widespread use of digital technology in business is about twenty-five years old – since the advent of the World Wide Web. It is only in the past decade that institutions have begun to formalize how our knowledge for future generations will be preserved as performed in [34]. Each of the aforementioned efforts is part of an iterative process. Researchers develop digital preservation practices for their research with constraints imposed by the medium, assistance from exemplars and peers, and with direction from granting agencies. They can apply these practices to their personal lives and provide feedback to the process based on these experiences, making small changes to the system to fit their needs. Thus, the digital environment is a creation of human activity that is both technical and social. Solutions require improved understanding of the communication that crosses organizations and that recognize individual needs and existing record preservation systems have been depicted in [35]. The efforts at this early stage of the paradigm shift will influence future generations' perceptions of who we were, what was done, and why.

2.1 FUNCTIONAL REQUIREMENTS

Digitally sustainable practices for data are both individual and organizational responsibilities in which an information package is the result of a complex system of behaviors, decisions, policies, and practices. According to [36], the success of any strategy within the system depends on how behaviors are shared. Public opinion is less bound to location in a virtual world. Finding group similarities beyond political boundaries and across cultures is possible and incentivized by the market. In accordance with diffusion theory, decisions may be optional, collective, or authority driven [37]. The National Science Foundation is imposing an authority innovation decision on researchers who may act within their discipline to create a collective decision driven by collegial peer pressure while working at a university where academic freedom mandates that the requirements be optional. There are several roles and responsibilities with respect to digital preservation in an organization. Senior administrators need to understand the strategic, long-term importance of the institution's digital assets. Their choices create the options available throughout the organization [38]. Technologists need to maintain the systems. The archivists are the caretakers and the exemplars that practice and create practices for others to emulate. They must not only know the technology, but law and policy for all the entities that the archives are intended for (journals, government studies, privately funded organizations). Their behavior should be strongly influenced by their field, discipline, associates, and accepted social behavior [39]. The authors are ultimately responsible for ensuring archives are managed both locally and on a remote system (at least initially). There should at least be an explanation of how these aspects fit together to ensure that information stored today will be available four years in government of Iraq.

In [40] author calculates depth histogram and thresholds depth record based on that. They assume that the records are in front of the camera all the time and the rest of human government is considered as part of background. They calculate distance of records and other government parts from accumulated histogram by analyzing rapid increase in the histogram slops as described in [41]. They assume that the records are in front of camera and if slope is greater than certain value then this region is selected as candidate ROI. They use color information to segment other records at the same depth level, thus they turn to skin color detection. This method is highly restrictive for record movement, since it requires records to be in front of camera all the times.

It deals solely with depth records and finds record and its movement as described in [42]. They propose classifier that combines a boosting and cascade structure. The features for training are depth differences at the stage of detection and learning. They also implement depth region grow and Depth Adaptive Mean Shift. Although they claim successful results the tracking algorithm can be quite slow because of region growing method.

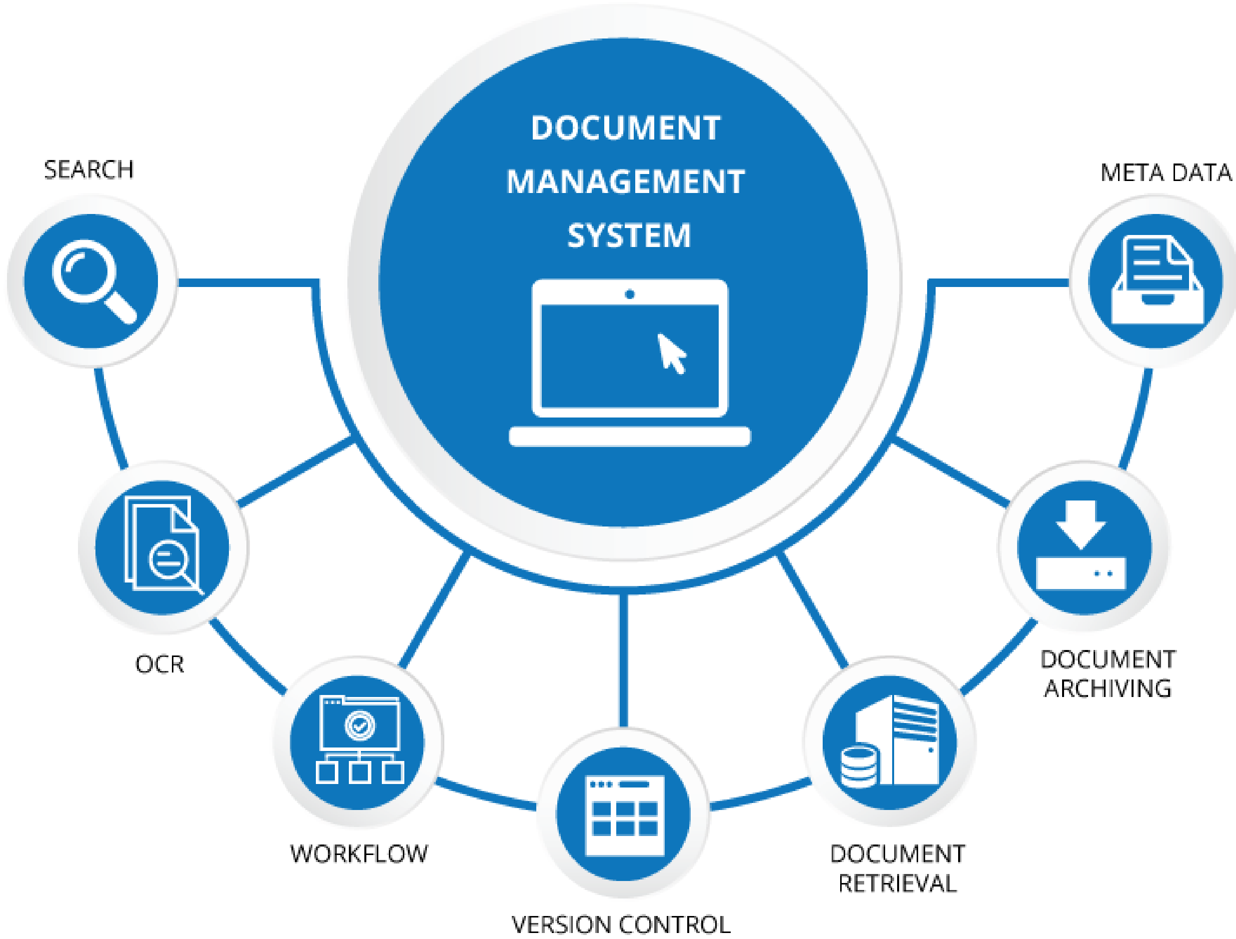

Figure 2.1: Record form recognition using the meta data from DMS smart system. Source [43]

The case examines how rules, resources, and training can be communicated to research faculty to raise overall awareness and improve digital preservation in the conduct of research. A goal of this study was to find an approach to influence better digital preservation at the institution through communication with the knowledge that faculty are strongly influenced within their discipline [44]. Since faculties are evaluated, supported, and tenured at the department level, messages tailored at the department level should be more successful than if from the college or university level. In order to understand how units, interact with each other with respect to data management, the study asked how researchers in the three departments manage their own digital data. Individual researchers are the unit of analysis [45]. The approach is to use the OAIS and the Activity theory to represent a construct of the interrelated processes that describe the digital preservation process. Digital record preservation and management is at the core concept of digitalization [46]. In this work, digital preservation is being performed for the three major digital libraries of government of Iraq whose scope transcends the technology that we use to navigate it. Our digital preservation is based on decisions by the creators of the artifacts for three different digital libraries, the system which manage it, the techniques which maintain the systems and that creates policy and funding for it. Researchers are creators of the information and, as stated earlier, have the greatest responsibility to establish the value of digital artifacts. Since the digital age will rely on the way we preserve our data, preservation practices must evolve to satisfy societal needs. Given the transient nature of the digital world, our decisions of what data to save has fundamentally changed. DOI is a mature theory, elements which quantified using digital system analysis for record digitalization.

The research has been designed solely for the government of Iraq digitalization plan and activity clusters with respect to digital preservation activity. Government Activity Theory provides a framework that includes rules, resources, tools and culture for the preservation of records. Both can explain the introduction and diffusion of digital preservation practices throughout the government by digitalization [47]. The general comparison between the major digital libraries of government with all digital libraries for the efforts of records preservation digitally.

2.2 NON-FUNCTIONAL REQUIREMENTS

There is a need to understand researcher workflows with respect to digital self-archiving attitudes and behaviors in order to improve digital sustainability. Access to local expertise and resources can improve their experiential knowledge. Support and guidance from campus leadership will improve the dialogue on campus that should provide increased collaborative solutions. Researcher behavior is strongly influenced by their domain. Peers in the field are very important to research practices. Acts, such as applying for grants to finance ministry or science ministry, influence overall researcher knowledge, perception, and attitude. However, a researcher's job appointment is within a department in a college or university to which they are accountable. Researchers, technologists and administration should have a dialogue on campus to resolve and learn best practices as mentioned in [48]. This is achieved on a campus through motivation and leadership. In the Iraq, both the public and private sector have begun to develop the framework, policies and practices to address the issue. We used OAIS to preserve archived data, but it also provides a roadmap for the new digital ecosystem. It offers a framework that can be used to identify roles and responsibilities within a bounded system. The digital libraries of Iraq are developing specific guidance for preserving digital artifacts as explained in [49]. Both the finance and science and technology have created mandates that require the consideration of an all-inclusive data management system by employers, administrators, technologists, and organizations. Finance ministry and science ministry requirements are important events in the trend to encapsulate and maintain humanity's knowledge digitally.

The study reveals the short-sighted attitudes of smartest system for the preservation of records. Among the findings of the study are that researchers are dissatisfied with their own level of expertise, but that few are thinking about long-term preservation. They cite that the demand of publishing undermines efforts to change their behavior unless it helps them to complete their research. The government itself has the responsibility for providing the policies, the resources, and fostering attitudes to instill sustainable digital behaviors [50]. Administrators need to have the providence to create systems that support their research faculty. Researchers who apply for these grants must consider the future access of their data, in so doing should consult expertise within their organization and their domain. Local events such as data preservation workshops and classes provide another communication channel to increase dialog and reinforce messaging.

3. DESIGN

It is notable that the digital records have successively used deeper networks but advanced deep learning techniques will provide great efficiency due to its architectural flexibility with convolutional neural network algorithm. Having more data models possess greater representational power making it possible to approximate more complex functions. Also, fast development in the field continuously introduces new techniques and architectures expanding the toolbox for design. However, science has always embraced simplicity, and it is reasonable to argue that the network should be kept as simple as possible. A study has been using the CNN technique even questioned the common practice of using pooling and sophisticated activation functions in machine learning. Instead, they suggest using just recurrent layers with properly selected stride and filter size parameters

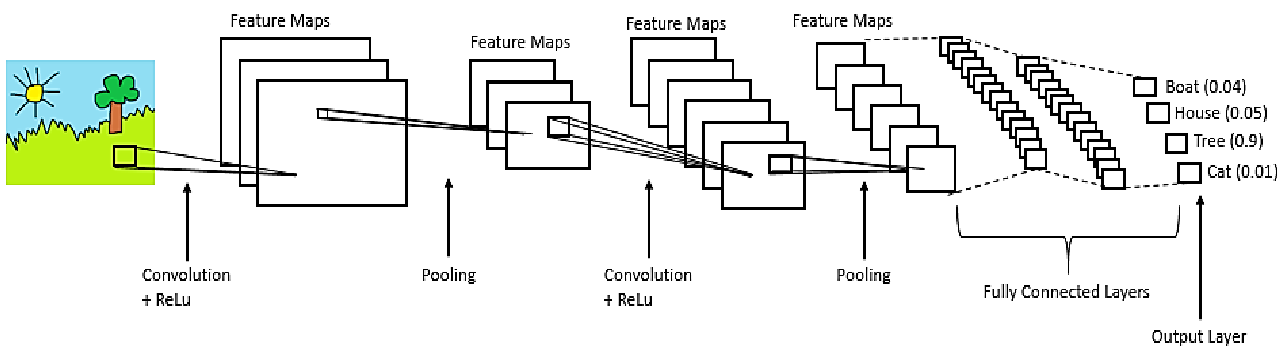

Figure 3.1: Proposed system depicted by flowchart using CNN.

3.1. ARCHITECTURE OF THE PROPOSED SOLUTION

In this architecture, the CNN are responsible for feature extraction, where the features to which they will respond are determined through a learning process of the variable input connections. Simple features, such as specific orientations of lines and edges, will be extracted by lower-level CNN models, while higher-level CNN models extract more global features (e.g. parts of learned patterns). The CNN model receive their inputs through fixed connections from CNN models in the previous layers and are responsible for a robust pattern recognition (i.e. decreasing sensitivity to a deformation or location shift of a pattern). Each CNN model receives inputs from a number of CNN models ('the input window') in one cell plane, i.e. from CNN models that extract the same feature, but at different spatial locations. The models will fire, if it receives an input from at least one of the models. Consequently, the feature will be detected even under a shift in the location of the input features, rendering the system less sensitive to the exact feature locations. The behavior of the models can however also be interpreted from an alternative point of view. The input windows for the different models strongly overlap thus models can be considered as performing a spatial blur on the excitatory signals they receive from the models. This spatial pooling is obtained by averaging these signals from the input window, which are models that perceive the same feature at slightly different locations. Furthermore, the excitatory cell input window is often framed by a small inhibitory region. This enables the models to distinguish a continuous line from the end-point of the line and allows the network to still recognize features after blurring. A consequence of the fixed input windows of the models is a reduction in size of the CNN-planes, thus allowing the network to compress data as it progresses to higher stages.

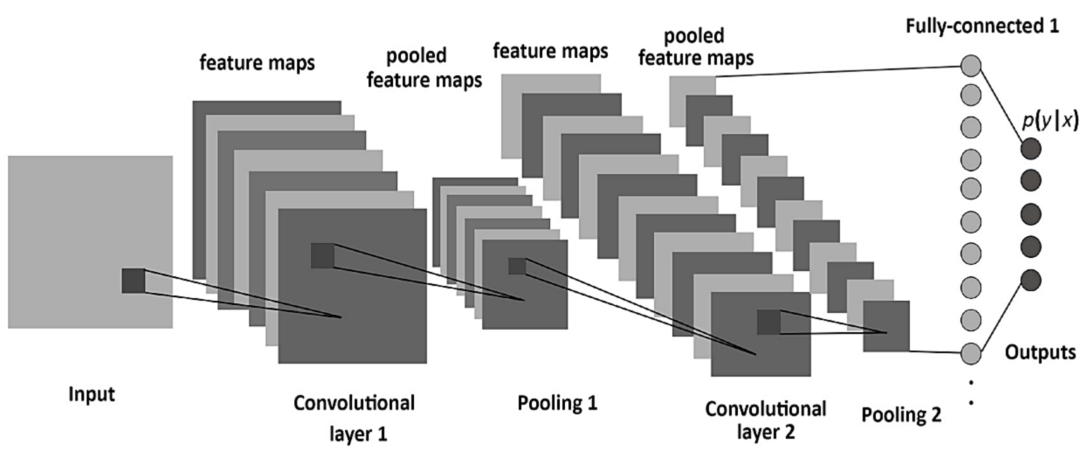

Figure 3.2: Illustration of a deep convolutional neural network.

CNN stands for convolutional neural networks originated from this work whose network has a similar architecture cognition. It translates the behavior of the models in mathematical formulations, respectively as convolutions and subsampling. The network primarily distinguishes itself however from the cognition in its training process. Whereas the neo-cognition relies on a layer-wise, unsupervised training process of the cell layers and supervised training of the output layer, CNN is trained to find a global minimum over all the parameters through back-propagation. The key notions that mimic biology and render CNN's so fusion for record smoothening.

Due to the high density of research in record illumination estimation, multi-exposure records often appears clumped. Thus, the main challenge for this research is to find the patch-wise solution for illumination estimation with human interference with CNN technique by separating different instances of multi-exposure fusion records. Traditional segmentation methods such as the seeded watershed algorithm, which is implemented in fusion records, are slow and sometimes make errors. In addition, the user needs to be experienced in order to use these algorithms as parameter tuning is required. The goal of this research is to overcome these dominant problems in the patch-wise illumination estimation for multi-exposure digital records. We will use a convolutional neural network for multi-exposure digital records. We will train the neural network on hand-annotated records and evaluate its performance with object-based metrics.

3.1.1 Key concepts of CNN's

Convolutional layer as records contain highly correlated 2D information (generally in three RGB-channels), the idea of local connections has been exploited by creating units in a convolutional layer which receive weighted input from local patches in the previous layer, referred to as their receptive fields (RF's). By relying on local RF's, this introduces sparse connections, rather than the whole-scale record statistics passed on to units and causing high-dimensional problems in fully-connected feed-forward networks. The weight vector of a unit is also known as filter bank and through training the weights will be updated to filter for specific features, giving the units the behavior of filter detectors. The weighted inputs and additive bias are then passed through the activation function of said unit, where ReLU's have portrayed much faster learning in deep networks compared to other activation functions.

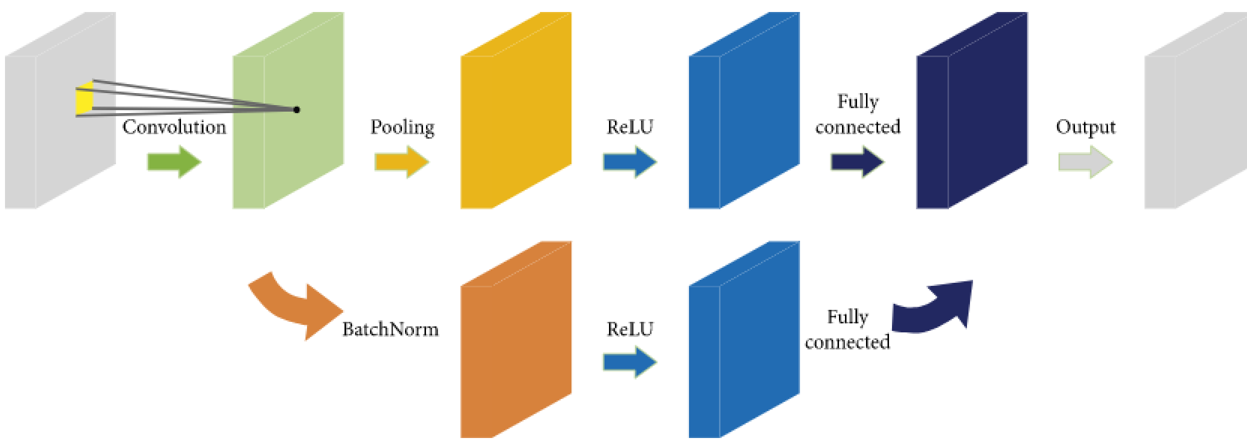

Figure 3.3: Basic building blocks of convolutional neural network.

Furthermore, records tend to portray similar features at different spatial locations. Accordingly, units with receptive fields at different locations will have the same weight vectors to extract similar features, which is known as parameter sharing and further reduces the dimensionality of the problem. On a structural level, the outputs of units with the same weight vector are organized in feature maps. Multiple feature maps (with different weight vectors) in each convolutional layer enable scanning for different features in each local record patch. The optimal number of feature maps needed to capture the record dynamics tends to increase with the complexity of the input records.

3.1.2. Pooling layer

The feature maps in the convolutional layer serve as input for the next stage, being a pooling layer. Given that the relative location of features with respect to each other contains significantly more information than their precise pixel-location, further stressed by the notion that small shifts in the precise feature locations are to be expected in different records, the idea of local pooling of features to reduce spatial resolution is implemented at this stage. This results in a network invariant to small shifts in the records and the stages of convolutional/pooling layers equip the network to handle variable size inputs.

Pooling units will scan the feature maps with (non-)overlapping p1 x p2 receptive fields and out- put a single value. Typical choices for pooling units are subsampling, where the output of a unit in the j^{th} layer is the average of units in the receptive field, supplemented with additive bias b and passed through an activation function and max pooling.

3.1.3. Deep hierarchies

Finally, the use of many layers represents the different stages in the simplified model of the ventral stream. Features at higher-level stages results from combinations of lower-level features, thus enabling deep structures to capture the compositional hierarchy of natural records.

3.1.4. Training and Validation

An implementation is based on hybrid models of CNN algorithm from various approaches. The visualization method proved to be helpful in interpreting the predictions, although it produced different results than expected. The expectation was that the CNN model would learn to identify small signs of skin disease risk. Instead, with custom CNN models they learned wider patterns that possibly predict resistance to fractures. This can of course be seen as the other side of the same thing. The experiments show that deep learning techniques can be used with imaging techniques to produce predictive ability different from the standard clinical methods currently in use. However, due to limited amount of data and wide range of architectural possibilities, their true potential in this context remains to be unfolded. It is also likely that the data analysis alone cannot provide the best fracture risk prediction, regardless of the model. It could contribute to more comprehensive prediction method that utilizes a wider set of input data to cover different aspects of the record edge smoothening. Using hybrid deep learning techniques produced the best record edge smoothening prediction results in the experiments. The results support the advertised benefits of hybrid deep learning, such as faster training speed and suitability for larger datasets.

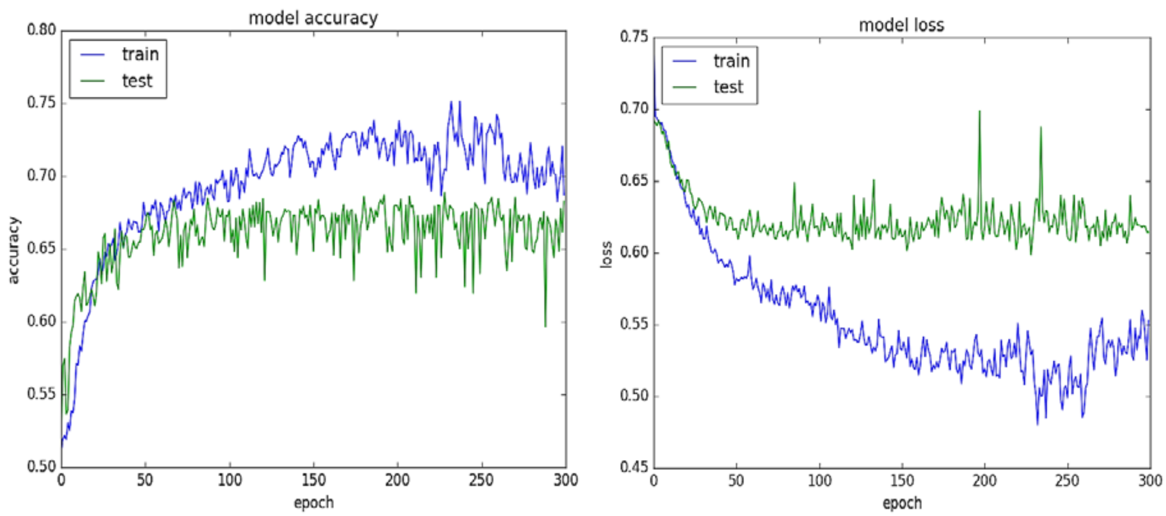

Figure 3.4: The CNN Model Accuracy (accuracy vs epochs) Model Loss (loss vs epochs).

Given the current hype of deep learning, optimizations of CNN architectures and designs is in hot pursuit. This section will therefore be devoted to some of the powerful CNN models that have been developed in the past few years and will be used throughout this work. Specific, revolutionary design-elements that have assisted these improvements are marked throughout this discourse.

3.1.5. Cross Validation

The cross validation is used to validate the data in artificial intelligence based systems, interpretable machine may even teach humans about how to make better decisions. To address the transparency problem, visualization techniques can be used to identify the record regions most important for the model's prediction. It is also possible to highlight the shapes and textures the network sees by visualizing pixel-wise impact on prediction score. Another visualization technique is called guided backpropagation. It produces an illustration of the detected features that contributed to the record edge smoothening prediction. This is done by taking account only the neurons that had a positive effect on the output, while others are set to zero. Combining different visualization techniques can produce effective visual explanations for the CNN model's predictions. To gain better coverage of the problem space, the number of training samples can be artificially expanded. This method called cross validation means modifying samples with random transformations such as rotation and flipping, varying properties like brightness and contrast or by adding random noise.

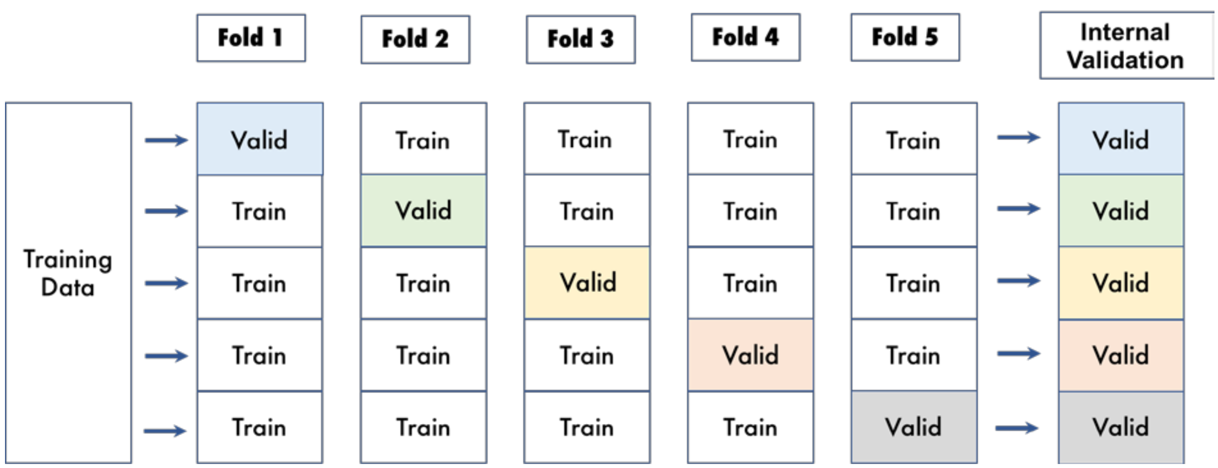

Figure 3.5: An example of dividing dataset for cross validation.

3.1.6. Data Evaluation

As deep learning techniques take numerical data as input, record analysis might seem like a simple case of feeding the network with raw record data. After all, records are nothing but grids of numbers representing pixel intensities. Actually, this approach can work well for simple recognition tasks with small, normalized records. The problem is poor scalability. When the input size and number of layers' increase, the amount of required computation explodes.

Number of Epochs	Loss	Accuracy	Validation Loss	Validation Accuracy
25	0.3301	0.8492	0.7911	0.5360
50	0.2418	0.9047	0.7874	0.5360
75	0.1891	0.9350	0.7896	0.5360
100	0.1702	0.9426	0.7456	0.5379
125	0.1384	0.9606	0.4709	0.7803
150	0.1246	0.9625	0.3569	0.8220
175	0.1020	0.9768	0.3543	0.8295
200	0.0809	0.9829	0.3598	0.8371
250	0.0747	0.9853	0.3662	0.8295
300	0.0678	0.9867	0.3709	0.8314

Table 3.1: The number of epochs per training instances for whole OAIS library.

3.2. ALGORITHM FOR SOLVING THE PROBLEM

Possible sources of variation include the smoothening techniques, filter status, topography, filter properties, smoothening conditions, segmentation methods and other classification influences. Due to such local variations, the task of accurately differentiating between several field record filter can become ambiguous.

This makes it difficult to create robust record processing algorithms that recognize all kinds of field instances without overfitting to the specific scene. In addition to the variation within a single scene, satellite records can show even stronger intra-class variations. This applies to records from different filter regions, from different dates or growing seasons, or under different atmospheric and illumination conditions. Also, in view of increasing public availability of high quality fusion record datasets, the use of additional and more heterogeneous ground truth and fusion record data (e.g. from multiple study areas in parallel) could improve the general model accuracy, robustness and transferability. Additional data augmentation (e.g. random scaling, distortions, color offset or jitter, introduction of record noise) might also help mitigate overfitting and increase the robustness to different fusion settings as well as record lightning conditions.

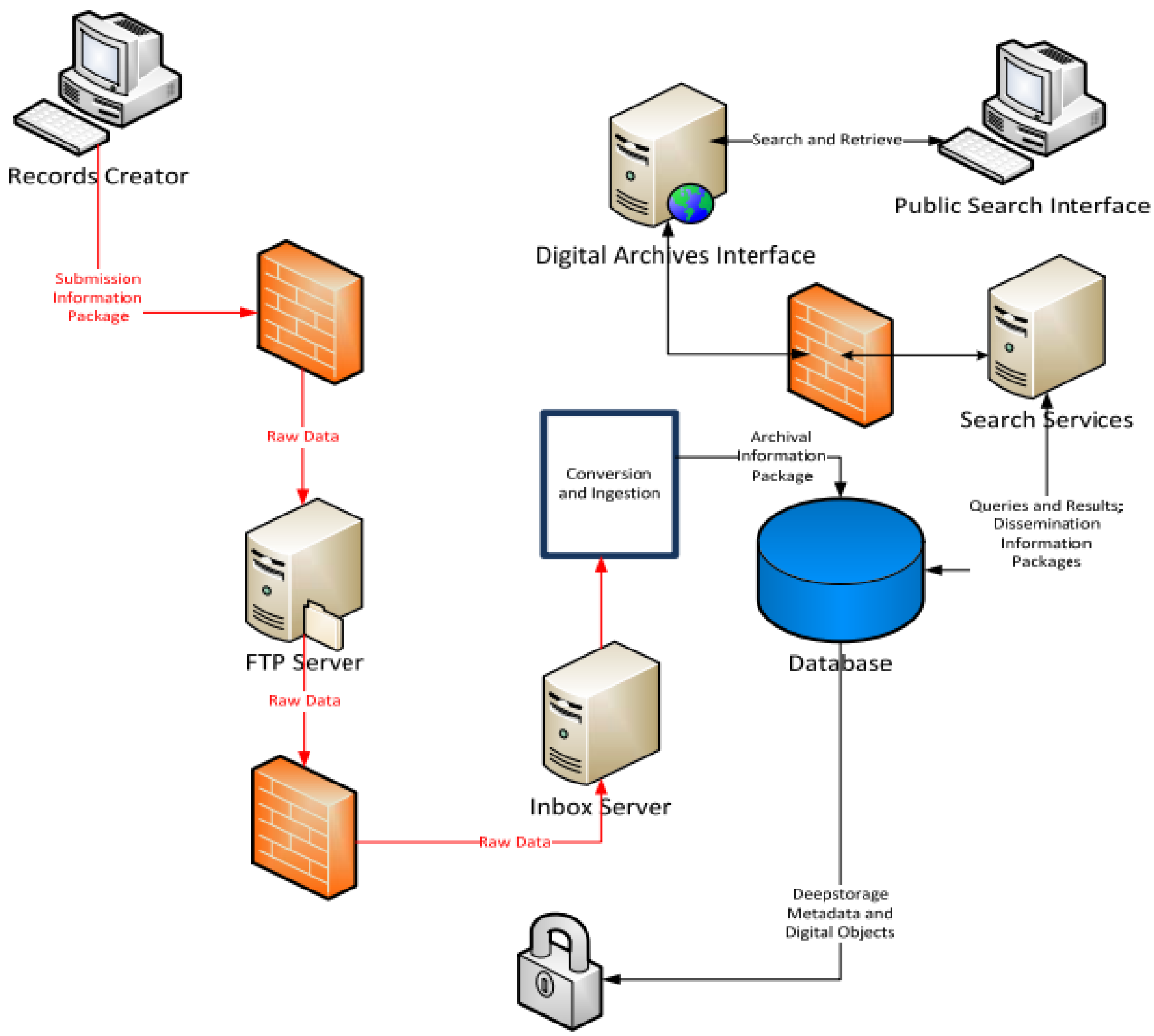

Figure 3.6: Functional overview of preservation system.

In the proceeding discussion, the digital records will serve as a guideline to discuss the most poignant results obtained with CNN's with indication of some remarkable new elements. To give an indication of the performance of these models, the top-5 error rate on the OAIS dataset of the aforementioned nets is presented. It has to be noted that the submitted nets, training and testing schemes offer a myriad of other hyper-parameters, which were tuned to achieve these results (e.g. a 7-network ensemble of CNN-architectures).

The research focuses on compiling the delineate details and classifies digital records with edge smoothing using convolutional neural network algorithm and convolutional neural network while distinguishing between record instances of the same class. Instance segmentation lies at the intersection of edge detection (predicting the bounding box and class of a variable number of records, but not segmenting them) and semantic segmentation (labeling each record pixel with a semantic category label, but not distinguishing between edges of the same category). With semantic segmentation it would be possible to find single instances of CNN algorithm on fields if they were completely surrounded by areas of different record classes.

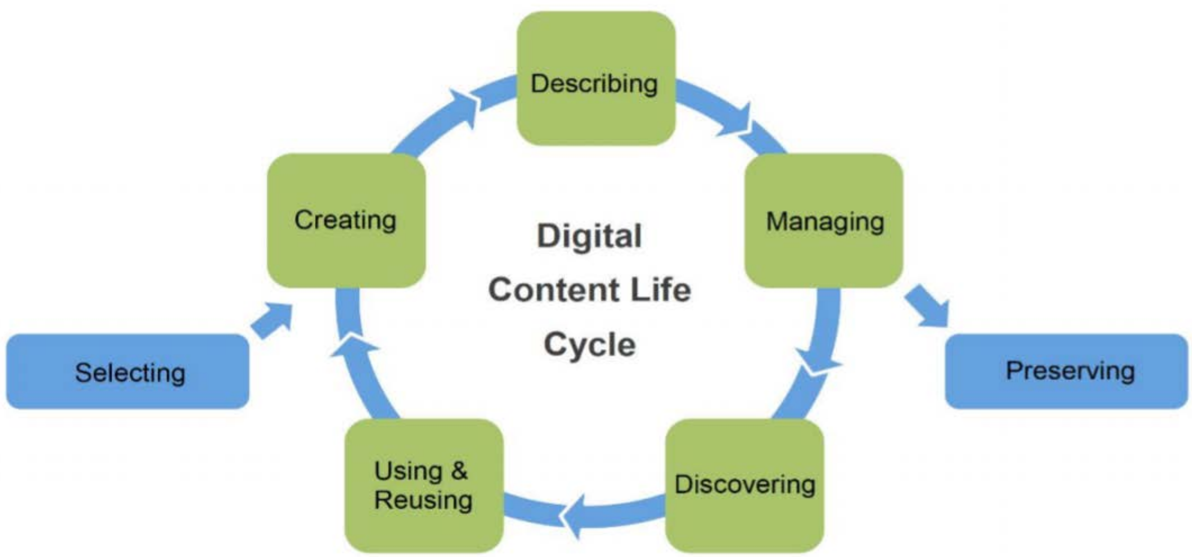

Figure 3.7: Convolutional neural network flowchart that is being followed in this work for OAIS database for digital records.

These datasets usually digital records from manual tracing of field boundaries. However, manual tracing of record edge smoothening is heavily dependent on the used recorder and supplementary data. It also is a highly subjective task, inevitably leading to inaccuracies and ambiguities depending on the priorities of the operator. The ground truth dataset can be complemented by existing property parcel information that could be indistinguishable based on the recorder alone. Furthermore, a single field record edge smoothening can potentially show several subfields due to unreported land use changes within the growing season. Given these constraints, even a very powerful automated record processing algorithm can only become as good as the ground truth data but will never be in perfect alignment with reality. Although the model can generalize within certain limits (local generalization), its transferability is mainly limited by the variability of the training data. However, filter field are often directly adjacent to each other and have touching boundaries. In this case, edge segmentation would yield the record edge smoothening. Same filter instance segmentation is able to distinguish between these connected or overlapping instances of the same class. Deep learning has not reached mainstream adoption in the remote sensing community yet.

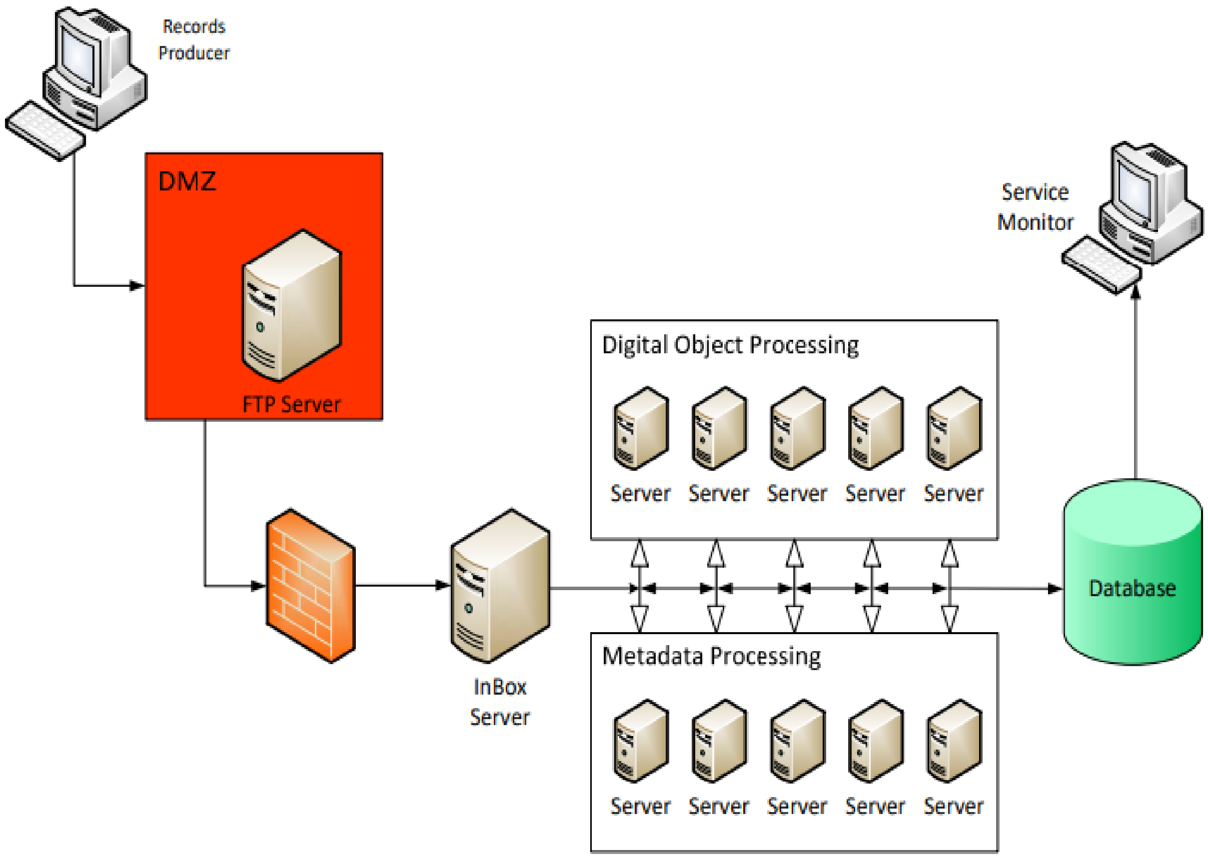

Figure 3.8: The process of digital record object creation along with database.

3.2.1. Preservation of Digital Record

The digital record is basically a buffer of intensities per pixel. For this reason, the record at this stage is very dark and not suitable for a computer vision application. When the record is fusion, each pixel is mapped to a three-channel RGB representation with red, green and blue channels. The CNN algorithm determines the intensities of the other channels, based on information from neighboring pixels. This is a loss stage, as information in terms of bits per pixel is being lost. The digital records have a huge tone, which is due to the Bayer mosaic pattern. In the Bayer mosaic, green is the dominant color as the color arrays in the record sensor are placed in rows of green-blue and red-green. As can be seen, color transformation corrects for that hue tone. The effect of gamma correction is not directly visible when compared to the output of the color transformation stage, but adds to the highlights and black tone appearance of the final result. The effect of the tone mapping stage is clearly visible, as it makes the record pleasant to the human vision, by fixing the brightness of different parts of the record. Furthermore, the training of the aforementioned deep CNNs generally required multiple GPU's and was a time-consuming process (up to 3 weeks). Thus, it is clear that training the same, high-performance CNN architectures on real-life data is in most cases unfeasible. The CNNs trained to recognize the records dataset have nevertheless a strong capability to extract very distinctive features from natural records, rendering them useful for datasets of other records. This is supported by the idea that the lower levels of these architectures extract general features from the records and become increasingly more task-specific deeper in the network. From this the idea of transfer learning sprung forth, which aims to kick start the process of the random weight initialization by using the pre-trained weights of deep CNN's to facilitate the training process.

3.2.2. Key Notions Concerning Digital Records for Storage

The CNN algorithm learning entails training a source network on a source dataset for a source task, transferring the learned knowledge (e.g. features or weights) to a new network for a different learning task and/or on a different target domain. Transfer learning is used with the objective of increasing performance on the target task by exploiting the trained knowledge of the source task. It can either serve as a 'Kick-starter' for the process improving initial performance, increase final performance or speed up the training of the target task. Three main types of transfer learning can be distinguished with respect to similarities in the source/target domains/tasks: (1) inductive transfer learning with the goal of inducing a generalizable, predictive model for the target task from a different source task; (2) transfer learning where a model should be derived in the target domain for the target task lacking labeled data, based on the source task containing a substantial amount of labeled data; and (3) unsupervised transfer learning which targets unsupervised learning in the target domain based on a different source task.

For purposes of record classification, the approach of inductive transfer learning has gained momentum in the past years. Given the large training time and requirements on labeled data, architectures are pre-trained on the records (source domain) for 1000-class classification purposes (source task) and knowledge is transferred to a labeled dataset of choice (target domain) for e.g. classification as target task. Thus, it primarily exploits the learned feature-representations in the source domain and aims to transfer these to the target domain with the aim of improving classification performance whilst making use of deep, data-greedy CNN architectures.

The transferability of different levels of features was touched upon in this work with results confirming that the deepest layers are most task-specific and better results can be obtained by using features from levels upwards in the model. This was confirmed by the work, in which furthermore was suggested that pooling of features as well as combinations of features may improve performance. We added to the evidence by documenting higher performances for CNN's initiate with pre-trained weights and fine-tuned for the task at hand, rather than through random weight initialization and training from scratch. Furthermore, they investigated the transferability of specific feature layers and reported co-adaptation of intermediate feature layers, where splitting between frozen and trainable layers at this level might result in a drop in performance.

3.2.3. List of Development Issues

- The experiments unveil that deep learning techniques can be used with data storage to produce predictive ability different from the standard clinical methods currently in use.
- The problem with limited amount of data and wide range of architectural possibilities, their true potential in this context remains to be unfolded.
- In future the problem might arise as likely that the data analysis alone cannot provide the best fracture risk prediction, regardless of the CNN model.
- The CNN learning technique cannot be used with the objective of increasing performance on the target task by exploiting the trained knowledge of the source task.
- The ground truth dataset can be complemented by existing property parcel information that could be indistinguishable based on the recorder alone.

3.3. DESCRIPTION OF DATA

The dataset has been acquired from an open-source repository known as the OAIS digital records dataset. The physical appearance of records in records can vary depending on many physical and human geographical factors, both from the ground perspective as well as in digital recorder. Possible sources of variation include the cultivated species, plant status, topography, soil properties, weather conditions, cultivation methods and other human influences. Due to such local variations, the task of accurately differentiating between several field records can become ambiguous. This makes it difficult to create robust algorithms that recognize all kinds of field instances without overfitting to the specific scene. In addition to the variation within a single scene, satellite records can show even stronger intra-class variations. This applies to records from different agricultural regions, from different dates or growing seasons, or under different atmospheric and illumination conditions. Also, in view of increasing public availability of high quality geo datasets, the use of additional and more heterogeneous ground truth and satellite record data (e.g. from multiple study areas in parallel) could improve the general model accuracy, robustness and transferability. Additional data augmentation (e.g. random scaling, distortions, color offset or jitter, introduction of record noise) might also help mitigate overfitting and increase the robustness to different geographic settings as well as atmospheric and lightning conditions. The data reference model is given in this link: https://libguides.bodleian.ox.ac.uk/digitalpreservation/oaismodel

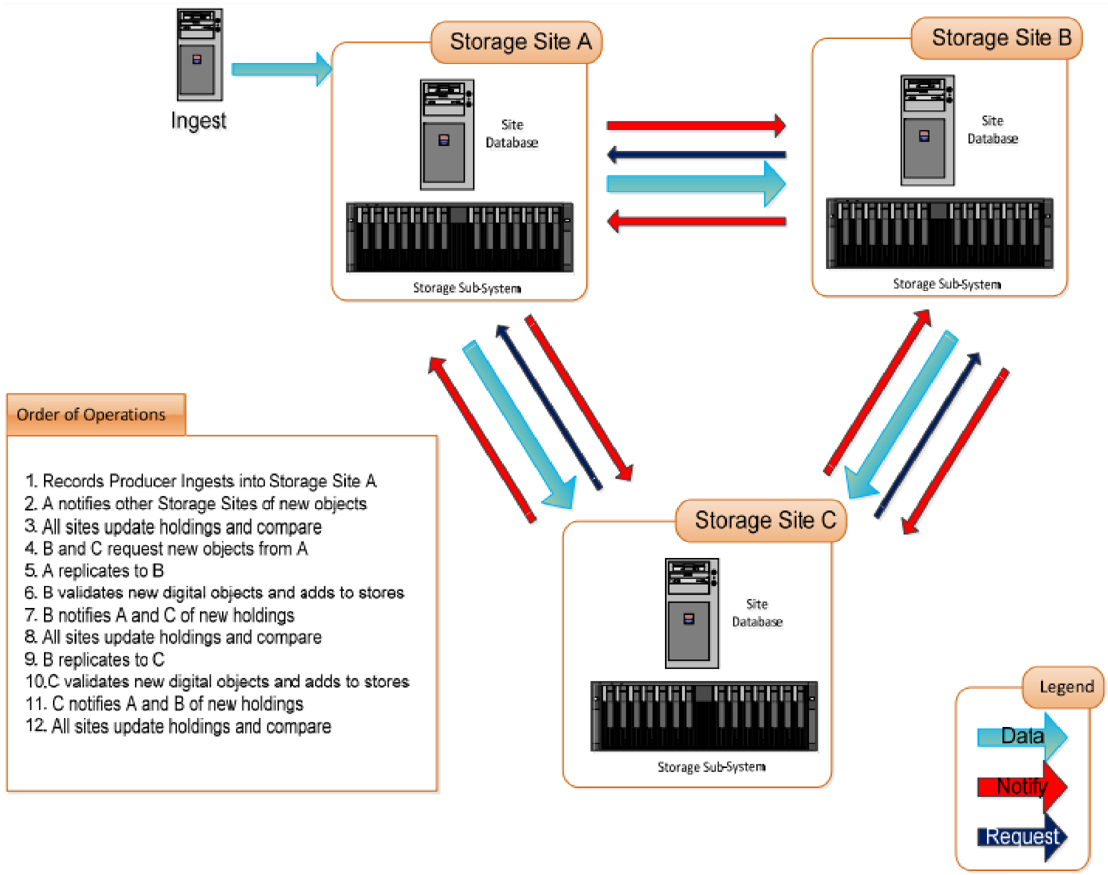

Figure 3.9: The final storage site replication process followed by record preservation.

Figures above shows records under various exposures in all sequence. More details concerning every one of the sequences, like name, number of source records, and spatial resolution have been stated in the caption of figures. Amongst the 8 sequences of test, two of them have been considered in the moving object scenes and the others have been considered in the static scenes. In addition to that, the first 5 sequences. The levels of exposure are set manually according to the camera tools. For uncontrolled outdoors environment, moving records such as, trees moving due to wind, make acquiring well-aligned sequences quite a challenge. However, all the records in each dataset were registered in order to align the frames to be in the same angle and direction, so the content of each record will not vary according to the used measure. Therefore, this will be not variable in our experiments.

4. IMPLEMENTATION

4.1. IMPLEMENTATION OF INTERFACES

In this research work, the main motive is to develop a smart system for the government of Iraq for the purpose of preserving the existing paper-based records to the digital records. Experts have understood the risks associated with sustaining a digital world for several years. The goal of this study is to find ways that digital preservation activity can be communicated to researchers to improve the permanency of their research data as well as government records. The research focused on activities in governmental environment that result in the creation of sustainable information packages (SIP) for three dedicated research digital libraries i.e. electricity, finance, science and technology for digitalization as the CNN approach. Paper-based information packages are the input to the Open Archive Information System (OAIS) model. The OAIS is an International Standards Organization (ISO) standard that constitutes the framework for the new digital environment will normalize and transform the paper-based data to the digital records.

Today, mediated communications and knowledge have converged onto digital formats that have been adopted by almost every society on the planet. Although the implications of the digital transformation have been studied from many perspectives, its impact may not be fully understood for generations to come. We proposed a new digital model over four years of governmental records from 2014 to 2021, the accelerating rate of digital data growth has increased both the number of extraordinary opportunities for discovery and the risk of unintentionally removing priceless information sources from the archive of collective human knowledge with a keystroke as explained the process. While the technology to store data (hardware and software) has improved at an increasing pace, personal and organizational behaviors have been much slower to adapt to the new environment. Data storage has become easier as the price of storage devices has plummeted while data preservation has become harder because of the task's magnitude and unseen, abstract nature. It is notable that the digital records have successively used deeper networks but advanced deep learning techniques will provide great efficiency due to its architectural flexibility with convolutional neural network algorithm. Representation with modeling, specifying key features for detection, contours tracking, or by using classification techniques such as neural network and binary tree classification. Although machine learning algorithms and classifiers needs robust training with lots of data only then they are able to segment record from records.

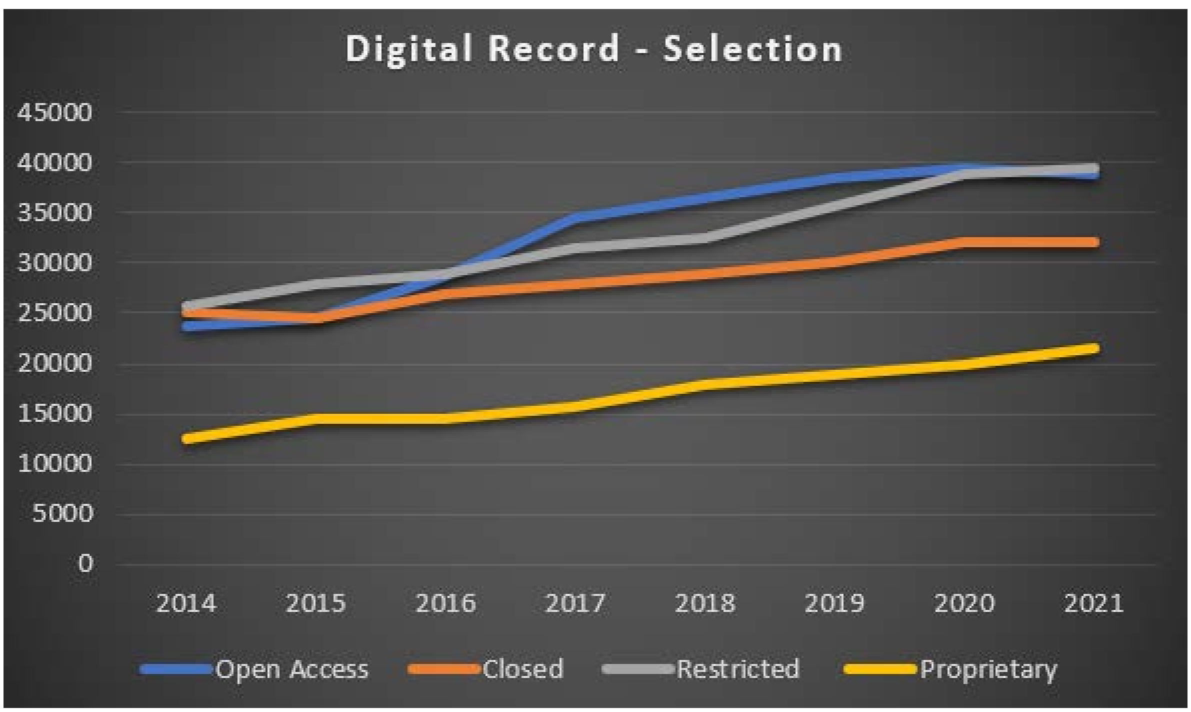

Figure 4.1. The selection of digital records repository in terms of open access, closed access, restricted and proprietary documents from OAIS database from year 2014 until 2021 on x-axis with number of articles on the y-axis.

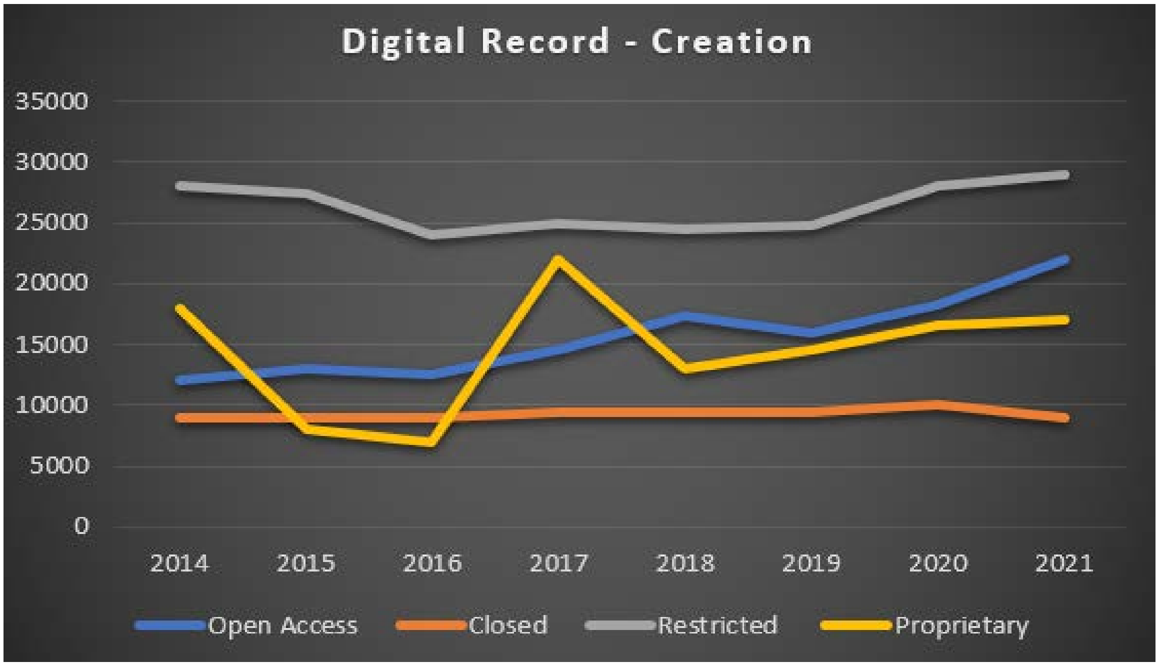

Figure 4.2. The creation of digital records repository in terms of open access, closed access, restricted and proprietary documents from OAIS database from year 2014 until 2021 on x-axis with number of articles on the y-axis.

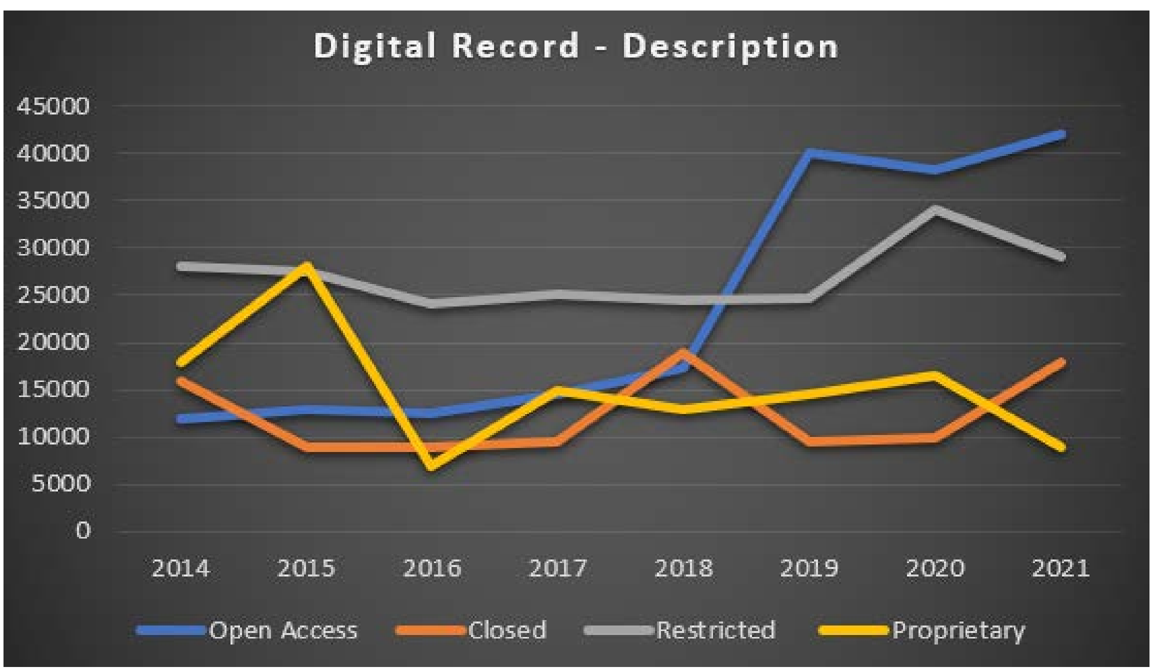

Figure 4.3. The description of digital records repository in terms of open access, closed access, restricted and proprietary documents from OAIS database from year 2014 until 2021 on x-axis with number of articles on the y-axis.

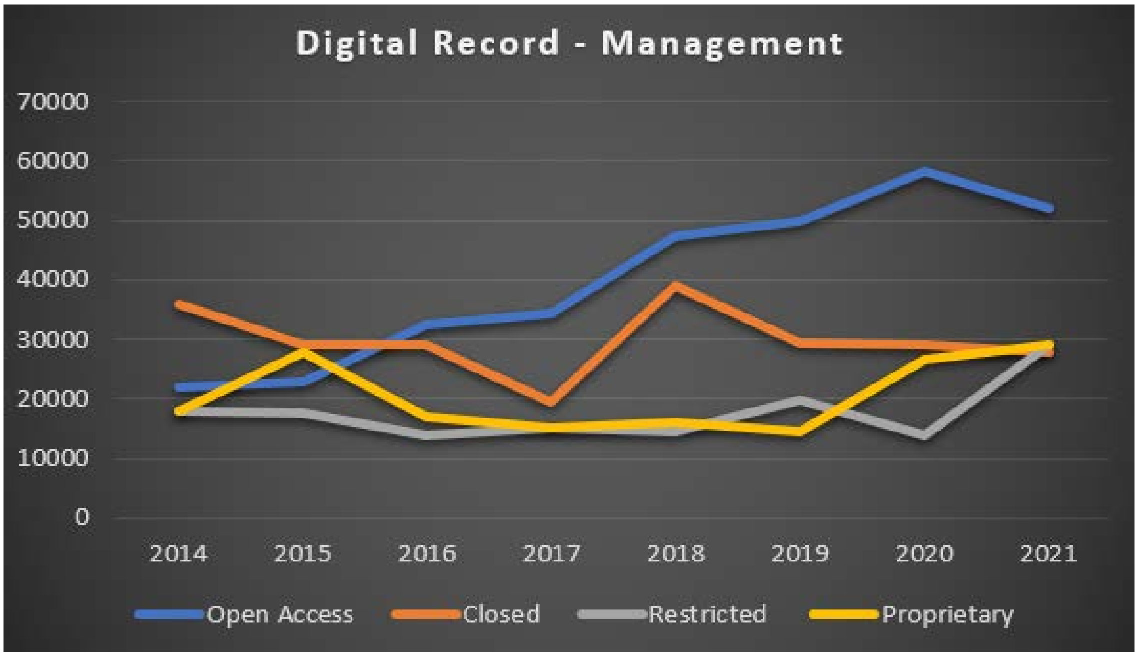

Figure 4.4. The management of digital records repository in terms of open access, closed access, restricted and proprietary documents from OAIS database from year 2014 until 2021 on x-axis with number of articles on the y-axis.

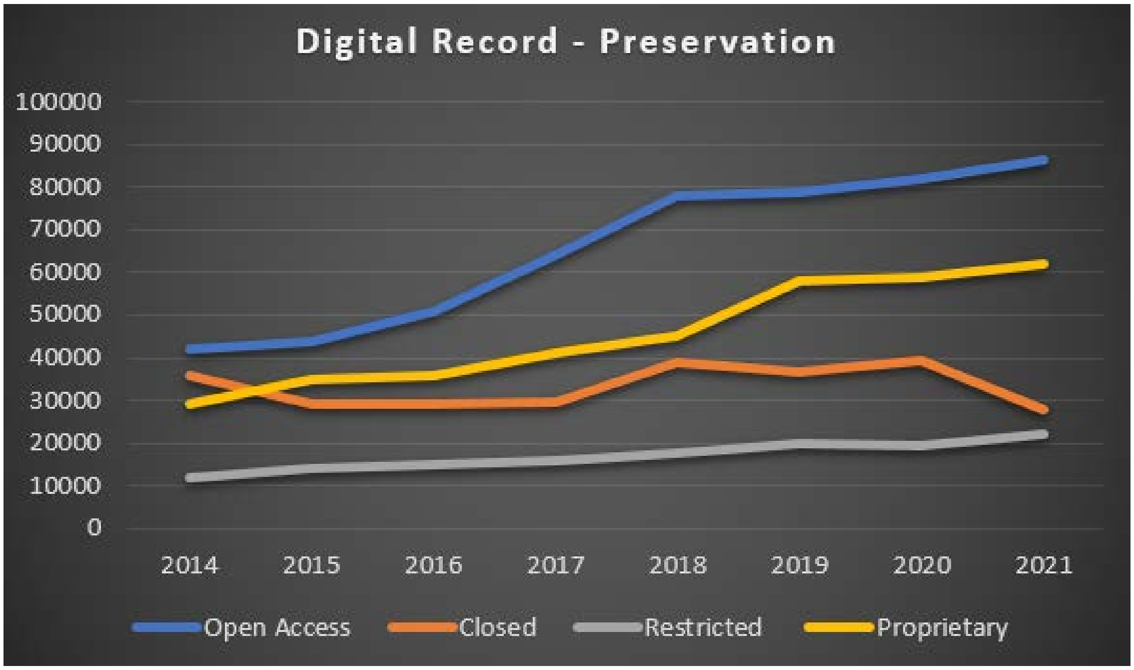

Figure 4.5. The preservation of digital records repository in terms of open access, closed access, restricted and proprietary documents from OAIS database from year 2014 until 2021 on x-axis with number of articles on the y-axis.

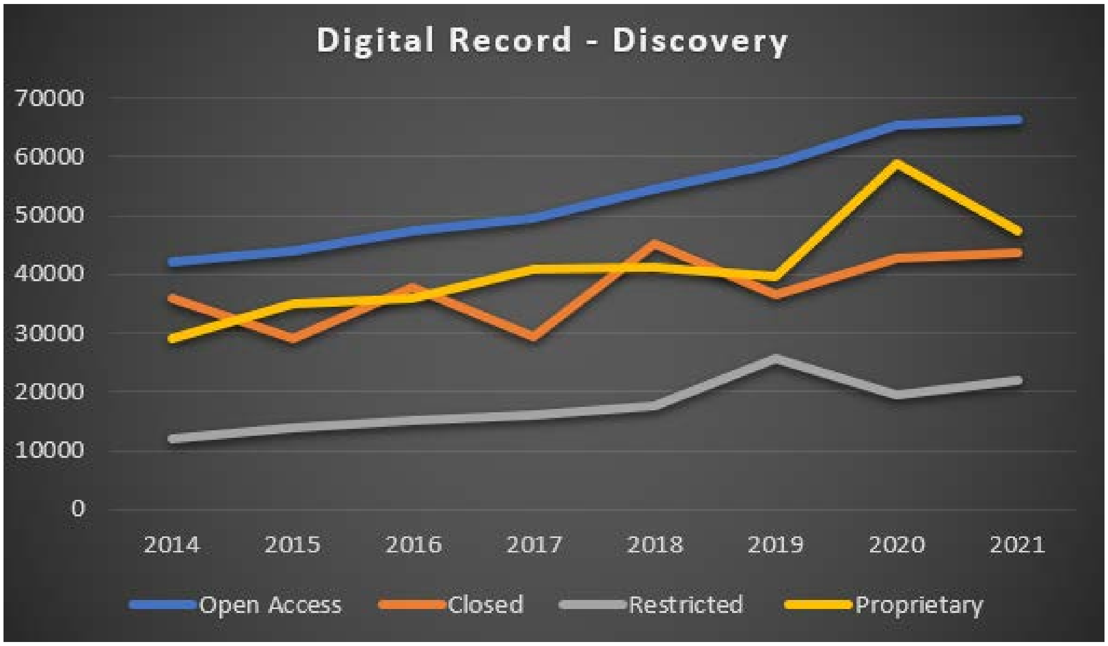

Figure 4.6. The discovery of digital records repository in terms of open access, closed access, restricted and proprietary documents from OAIS database from year 2014 until 2021 on x-axis with number of articles on the y-axis.

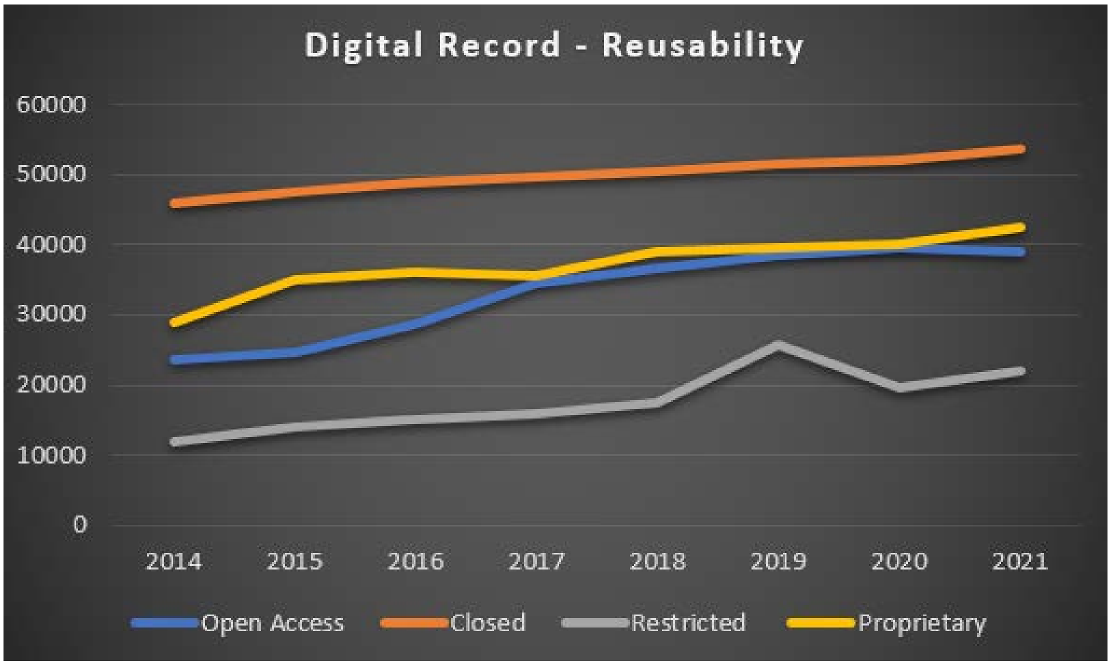

Figure 4.7. The reusability of digital records repository in terms of open access, closed access, restricted and proprietary documents from OAIS database from year 2014 until 2021 on x-axis with number of articles on the y-axis.

A goal of this study was to find an approach to influence better digital preservation at the institution through communication with the knowledge that faculty are strongly influenced within their discipline. Since faculties are evaluated, supported, and tenured at the department level, messages tailored at the department level should be more successful than if from the college or university level. In order to understand how units, interact with each other with respect to data management, the study asked how researchers in the three departments manage their own digital data. Individual researchers are the unit of analysis. The approach is to use the OAIS and the Activity theory to represent a construct of the interrelated processes that describe the digital preservation process. Digital record preservation and management is at the core concept of digitalization. In this work, digital preservation is being performed for the three major digital libraries of government of Iraq whose scope transcends the technology that we use to navigate it. Our digital preservation is based on decisions by the creators of the artifacts for three different digital libraries, the system which manage it, the techniques which maintain the systems and that creates policy and funding for it. Researchers are creators of the information and, as stated earlier, have the greatest responsibility to establish the value of digital artifacts. Since the digital age will rely on the way we preserve our data, preservation practices must evolve to satisfy societal needs. Given the transient nature of the digital world, our decisions of what data to save has fundamentally changed. DOI is a mature theory, elements which quantified using digital system analysis for record digitalization.

The ground truth dataset can be complemented by existing property parcel information that could be indistinguishable based on the recorder alone. Furthermore, a single field record edge smoothening can potentially show several subfields due to unreported land use changes within the growing season. Given these constraints, even a very powerful automated record processing algorithm can only become as good as the ground truth data but will never be in perfect alignment with reality. Although the model can generalize within certain limits (local generalization), its transferability is mainly limited by the variability of the training data. However, filter field are often directly adjacent to each other and have touching boundaries.

5. TESTING

5.1. TESTING METHODOLOGY

Digital record preservation and management is at the core concept of digitalization. In this work, digital preservation is being performed for the three major digital libraries of government of Iraq whose scope transcends the technology that we use to navigate it. Our digital preservation is based on decisions by the creators of the artifacts for three different digital libraries, the system which manage it, the techniques which maintain the systems and that creates policy and funding for it. Researchers are creators of the information and, as stated earlier, have the greatest responsibility to establish the value of digital artifacts. Since the digital age will rely on the way we preserve our data, preservation practices must evolve to satisfy societal needs. Given the transient nature of the digital world, our decisions of what data to save has fundamentally changed. DOI is a mature theory, elements which quantified using digital system analysis for record digitalization. State of the art thesis include many record-based identification and tracking approaches. Representation with modeling, specifying key features for detection, contours tracking, or by using classification techniques such as neural network and binary tree classification. Although machine learning algorithms and classifiers needs robust training with lots of data only then they are able to segment record from records. This approach needs lots of processing and hence computationally expensive. The more traditional method where we use record form depth information and try to find record from those records combined; is an ideal approach for this task. The OAIS is one of several international (ISO) standards that attempt to resolve problems in an increasingly connected world. Focusing on the average best validation error, two statistics seem to correlate with its absolute improvement. On one record, the improvement seems to shrink with the number of classes and thus the number of original samples in the training set. As the number of selected classes is increased from seven to ten, the difference becomes smaller and even negative. However, even though relatively small, the classification accuracy is higher when the data augmentation is used for twenty classes as well as the whole dataset. On the other record, the data augmentation becomes more effective for subsets with larger maximum sequence length. While there was a similar amount of original training samples in the three five-class subsets, the maximum sequence length strongly varied.

5.2. CONDUCTING THE TEST PROCEDURE

The research has been designed solely for the government digitalization plan and activity clusters with respect to digital preservation activity. Government record activity theory provides a framework that includes rules, resources, tools and culture for the preservation of records. Both can explain the introduction and diffusion of digital preservation practices throughout the government by digitalization. The general comparison between the major archives of repositories with all journals for the efforts of records preservation digitally. Digital technology has been available for over half a century. The widespread use of digital technology in business is about twenty-five years old – since the advent of the World Wide Web. It is only in the past decade that institutions have begun to formalize how our knowledge for future generations will be preserved as performed. Each of the aforementioned efforts is part of an iterative process. Researchers develop digital preservation practices for their research with constraints imposed by the medium, assistance from exemplars and peers, and with direction from granting agencies. They can apply these practices to their personal lives and provide feedback to the process based on these experiences, making small changes to the system to fit their needs. Thus, the digital environment is a creation of human activity that is both technical and social. Solutions require improved understanding of the communication that crosses organizations and that recognize individual needs and existing record preservation systems.

ARTICLE	**TECHNIQUE**	**DIGITAL LIBRARIES**
[51]	Trustworthy Digital Object (TDO)	National Digital Information Infrastructure Preservation Plan (NDIIPP)
[52]	Byte Replication and Encapsulation	Information Communication Technology (ICT)
Proposed	Convolutional Neural Network (CNN)	Open Archival Information System (OAIS)

Table 5.1. The compassion for the preservation of records in digital form among existing digital libraries.

The efforts at this early stage of the paradigm shift will influence future generations' perceptions of who we were, what was done, and why. Today, mediated communications and knowledge have converged onto digital formats that have been adopted by almost every society on the planet. Although the implications of the digital transformation have been studied from many perspectives, its impact may not be fully understood for generations to come. We proposed a new digital model over four years of governmental records from 2014 to 2021, the accelerating rate of digital data growth has increased both the number of extraordinary opportunities for discovery and the risk of unintentionally removing priceless information sources from the archive of collective human knowledge with a keystroke as explained the process. While the technology to store data (hardware and software) has improved at an increasing pace, personal and organizational behaviors have been much slower to adapt to the new environment. Data storage has become easier as the price of storage devices has plummeted while data preservation has become harder because of the task's magnitude and unseen, abstract nature.

CONCLUSION

In this work, digital preservation is being performed for the three major digital libraries at governmental level, the scope transcends the technology that we use to navigate it between the gap of preservation from paper-based to digital format. This gap leads to the preservation of the digital data. These gaps can create a nearly impossible task of recovering paper-based data but using this smart system, the records can be recovered from any digital repository where it is stored. This is the time that many government regulations require documents be stored for reporting purposes in digital format as many other prospering countries have upgraded their system using CNN. In results, we achieved good preservation of most of the records from 2014 to 2021, the goal is to preserve which most of the records it belongs to digital format. The distribution of records preservation for major archives in term of paper based and digitally preserved records from year 2014 to 2021. The preservation of digital records was recorded very low in 2014 as much of the record archival were paper-based in all major repositories of the government. However, year-by-year the smart system tends preserve the records in digital format from old paper-based format and till 2021 large amount of records are being converted to the digital format with a whopping accuracy of 98.68% for all the records where the system was trained on 80% of data and tested on rest of the 20% of data. In this work, digital preservation is being performed for the three major digital libraries of government of Iraq whose scope transcends the technology that we use to navigate it. Our digital preservation is based on decisions by the creators of the artifacts for three different digital libraries, the system which manage it, the techniques which maintain the systems and that creates policy and funding for it. Researchers are creators of the information and, as stated earlier, have the greatest responsibility to establish the value of digital artifacts. Since the digital age will rely on the way we preserve our data, preservation practices must evolve to satisfy societal needs.

FUTURE WORK

The OAIS is one of several international (ISO) standards that attempt to resolve problems in an increasingly connected world. Focusing on the average best validation error, two statistics seem to correlate with its absolute improvement.

- In future, the improvement will shrink with the number of classes and thus the number of original samples in the training set. As the number of selected classes is increased from seven to ten, the difference becomes smaller and even negative.
- However, even though relatively small, the classification accuracy is higher when the data augmentation is used for twenty classes as well as the whole dataset.
- In future, the research will aim to perform data augmentation becomes more effective for subsets with larger maximum sequence length. While there was a similar amount of original training samples in the three five-class subsets, the maximum sequence length strongly varied.
- The largest absolute improvement was achieved for the classes where the maximum sequence length was the highest and it would be determined in future.
- In future, the maximum length was equal in the experiments with seven, eight, nine and ten classes, where a shrinking improvement was observable with increasing training set size, the maximum sequence lengths are significantly larger when the experiments are repeated on the classes. In these cases, the data augmentation has positive effects again.

- As the maximum number of frames gets clipped to smaller numbers for the created sequences, the synthetic samples become less helpful for improving the classification accuracy on the whole dataset.

REFERENCES

1. American Psychological Association (APA): efficiency. (n.d.). Dictionary.com Unabridged. [Electronic Resource] URL: http://dictionary.reference.com/browse/efficiency

2. Baxter, P. and Jack, S. (2018). Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers. [Electronic Resource] URL:

http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.666.2447

3. Becker, H.S. and Geer, B. (n.d.). Participant Observation and Interviewing: a comparison. Human organization. [Electronic Resource] URL: https://www.jstor.org/stable/44127708?seq=1

4. Bikson, T. K. and Law, S. A. (2016) Electronic information media and records management methods: A survey of practices in UN organizations, The Information Society: [Electronic Resource] URL:

https://www.tandfonline.com/doi/abs/10.1080/01972243.1993.9960136

5. Chachage B. and Ngulube P. (2020). Management of business records in Tanzania: an exploratory case study of selected companies. [Electronic Resource] URL: https://journals.co.za/doi/abs/10.10520/AJA1560683X_318

6. Wickman, Danielle, “Measuring performance or performance measurement? Measuring the impact and sustainability of the PARBICA Recordkeeping for Good Governance Toolkit” [Electronic Resource] URL: https://search.informit.org/doi/abs/10.3316/ielapa.201002756

7. Demirtel, H. and Gökkurt Bayram, Ö. (2017). Efficiency of electronic records management systems: Turkey and example of Ministry of Development. Procedia - Social and Behavioral Sciences. [Electronic Resource] URL:

https://www.sciencedirect.com/science/article/pii/S1877042814040701

8. Duranti, L. (2015). Concepts and principles for the management of electronic records, or records management theory is archival diplomatics. [Electronic Resource] URL: http://dx.doi.org/10.1108/09565691011039852

9. Fang-Ming Hsu Tser-Yieth Chen Shuwen Wang, (2019). Efficiency and satisfaction of electronic records management systems in e-government in Taiwan. [Electronic Resource] URL: http://dx.doi.org/10.1108/02640470910966907

10. Harvard's Records Management Services. (2015). Controlling Costs and Promoting Efficiency. [Electronic Resource] URL: http://library.harvard.edu/sites/default/files/ControllingCostsAndPromotingEfficie ncy. pdf

11. Henriksen,H.Z. and Andersen, K.V. (2018). Electronic records management systems: implementation in the Pakistani local government. [Electronic Resource] URL: http://dx.doi.org/10.1108/09565690810858505

12. Hounsome, C. (2020). The Records Management Challenges of Amalgamation. [Electronic Resource] URL: http://www.condar.ca/CONDAR%20Articles/article7Amalgamation.pdf

13. Johnston, G.P. and Bowen, D.V. (2015), The benefits of electronic records management systems: a general review of published and some unpublished cases. [Electronic Resource]

URL:https://www.emerald.com/insight/content/doi/10.1108/09565690510632319/full/html?src=recsys&fullSc=1&mbSc=1

14. Keith, G., (2015). Implementing an electronic records management system. [Electronic Resource] URL: http://dx.doi.org/10.1108/09565690510614229

15. Ki-Young, J., Phillips, D. T. (2015). Operational efficiency and effectiveness measurement. International Journal of Operations & Production Management, [Electronic Resource] URL: http://dx.doi.org/10.1108/EUM0000000006223

16. Mäkinen, S. and Henttonen,P. (2011), Motivations for records management in mobile work", [Electronic Resource] URL: //dx.doi.org/10.1108/09565691111186867

17. Meijer, A. (2016). Accountability in an information age: Opportunities and risks for records management. [Electronic Resource] URL: //link.springer.com/article/10.1007/BF02438902

18. Mokhtar, U.A. and Yusof, Z.M. (2019), Electronic records management in the Malaysian public sector: the existence of policy, [Electronic Resource] URL: http://dx.doi.org/10.1108/09565690910999201

19. Myers, M. D., (2017). Qualitative Research in Information Systems. [Electronic Resource] URL: https://link.springer.com/chapter/10.1007/978-0-387-35309-8_1

20. Myers, M.D. and Newman, M. (2007). The qualitative interview in IS research: Examining the craft. Information and Organization. [Electronic Resource] URL:

https://www.sciencedirect.com/science/article/abs/pii/S1471772706000352

21. National Archives and Records Administration. (n.d.). Electronic records management: resource guide. [Electronic Resource] URL:https://www.archives.gov/files/era/pdf/rmsc0305.pdf

22. Patience Coetzer, X. (2020). The status of records management at the University of Zululand. [Electronic Resource] URL:http://uzspace.unizulu.ac.za/handle/10530/1219

23. Porter-Roth,B. (2016). Applying Electronic Records Management in the Document Management Environment: an Integrated Approach. [Electronic Resource] URL:

https://piworld.tradepub.com/form/applying-electronic-records-management-in-the-document-management-environment-an-integrated-approach/w_aaaa1654/prgm.cgi?sr=hicat&_t=hicat:1087

24. Quan-Haase, A. (2013). Theoretical Perspectives on Technology. Don Mills, Ontario: Oxford University Press. Richmond, H. (n.d.). Why Records Management? The Top Ten Business Reasons. [Electronic Resource] URL: http://condar.ca/CONDAR%20Articles/article1- %20ten%20reasons.pdf

25. Reed, B. (2017),"Electronic records management in Australia", [Electronic Resource] URL:http://dx.doi.org/10.1108/eb027111

26. Shaw,A. and Shaw, D.T. (2016). Electronic Records Management Criteria and Information Security. [Electronic Resource] URL:https://ro.ecu.edu.au/cgi/viewcontent.cgi?referer=https://scholar.google.com/&httpsredir=1&article=1018&context=isw

27. Sprehe, T. J. (2015). The positive benefits of electronic records management in the context of enterprise content management. [Electronic Resource] URL:https://www.sciencedirect.com/science/article/abs/pii/S0740624X05000195

28. Tambudzai Nengomasha , C. (2009). A study of electronic records management in the Namibian public service in the context of e-government. [Electronic Resource] URL:

https://repository.unam.edu.na/handle/11070/447

29. Turner, D. W. (2010). Qualitative Interview Design: A Practical Guide for Novice Investigators. [Electronic Resource] URL:

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.674.8927&rep=rep1&type=pdf

30. Umi Asma' Mokhtar and Zawiyah Mohammad Yusof, (2009). Electronic records management in the Malaysian public sector: the existence of policy. [Electronic Resource] URL: http://dx.doi.org/10.1108/09565690910999201

31. Washington State Archives (n.d.). Records retention cost-benefit analysis. [Electronic Resource] URL: https://www.uwmedicine.org/sites/stevie/files/2018-12/Records_Management_WA-General-Records-Retention-Schedule-v6.0.pdf

32. Yampolskiy, R. V. (2011). Efficiency Theory: A Unifying Theory for Information, Computation and Intelligence. [Electronic Resource] URL:

https://www.tandfonline.com/doi/abs/10.1080/09720529.2013.821361

33. Financial Reporting Council, The UK Corporate Governance Code. London: The Financial Reporting Council Limited, [Electronic Resource] URL:

https://www.frc.org.uk/OurWork/Publications/Corporate-Governance/UK-Corporate-Governance-Code-September2012.pdf; accessed on 15th May 2021.

34. Gollins, Tim, “Parsimonious preservation: Preventing pointless processes!”, [Electronic Resource] URL:http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1086.679&rep=rep1&type=pdf

35. Gough, Roger (ed.), “An End to Sofa Government: Better working of Prime Minister and Cabinet”, [Electronic Resource] URL:

http://image.guardian.co.uk/sysfiles/Politics/documents/2007/03/27/DemocracyTaskForce.pdf

36. Hill, Brian, “Forrester Research / ARMA International Records Management Survey”, [Electronic Resource] URL:https://www.arma.org/page/research

37. International Records Management Trust (IRMT), “Consultancy Projects”. [Electronic Resource] URL:

http://www.irmt.org/documents/educ_training/educ_resource/IRMT_ed_case_study_v1.pdf

Jisc infoNet, “Guidance notes for the Records Management Maturity Model.” [Electronic Resource] URL: https://www.jisc.ac.uk/guides/records-management/maturity-model

38. Lemieux, Victoria, “The records-risk nexus: exploring the relationship between records and risk”, [Electronic Resource] URL:

https://www.researchgate.net/publication/235296430_The_recordsrisk_nexus_Exploring_the_relationship_between_records_and_risk

39. Lemieux, Victoria, “One step forward, two steps backward? Does e-government make governments in developing countries more transparent and accountable?” [Electronic Resource] URL: https://openknowledge.worldbank.org/handle/10986/23647

40. McKemmish, Sue, “Traces: Document, record, archive, archives”, [Electronic Resource] URL: https://www.researchgate.net/publication/292880129_Traces_Document_record_archive_archives

41. Lipchak, Andrew, “Information management to support evidence-based governance in the electronic age “, [Electronic Resource] URL:

http://www.irmt.org/documents/research_reports/project_reports/Final%20DGF%20Report%20Revised.pdf

42. McLeod, Julie et al, “A Critical Assessment of Records Management Capacity and Compliance Toolkits”, [Electronic Resource] URL:

https://www.emerald.com/insight/content/doi/10.1108/09565690710833116/full/html

43. McLeod, Julie, Sue Childs and Rachel Hardiman, “AC+ erm – Accelerating positive change in electronic records management”, [Electronic Resource] URL:

https://core.ac.uk/download/pdf/5900118.pdf

44. McLeod, Julie, Sue Childs and Susan Heaford, “A Critical Assessment of Records Management Capacity and Compliance Toolkits” [Electronic Resource] URL:

https://www.academia.edu/47049871/A_critical_assessment_of_records_management_capacity_and_compliance_toolkits

45. National Archives, “About the National Archives – National Archives History”. [Electronic Resource] URL: http://theses.gla.ac.uk/7506/1/2015PhiriPhD.pdf

46. National Records of Scotland, “Implementation of the Public Records (Scotland) Act” [Electronic Resource] URL:

https://www.nrscotland.gov.uk/record-keeping/public-records-scotland-act-2011

47. Perkins, Eric, “Records Management Training: Introduction to Records Management”. [Electronic Resource] URL: http://www.vita.virginia.gov/uploadedfiles/vita_main_public/unmanaged/vita_rm_trainin g_2010v2.pdf; accessed on 9th May 2021.

48. Share, Felex. “Education Records Management in Shambles.” [Electronic Resource] URL:

https://www.herald.co.zw/education-record-management-in-shambles/

49. University College London, “University College London Records Management Policy”. [Electronic Resource] URL:

https://www.ucl.ac.uk/library/about-us/records-office/university-college-london-records-management-policy#:~:text=The%20principal%20aims%20of%20records,its%20timely%20and%20secure%20destruction

50. Gladney, Henry. (2006). Principles for Digital Preservation. Commun. ACM. [Electronic Resource] URL:https://dl.acm.org/doi/abs/10.1145/1113034.1113038

51. Matlala, Mpubane. (2020). Long-term preservation of digital records at the University of KwaZulu-Natal archives. [Electronic Resource] URL:

https://www.ajol.info/index.php/jsasa/article/view/189859