A Conceptual Framework for Persistent Preservation of Digital Information

Md. Ahsan ul Hasan*

Abstract

In every aspect of our life whether in scientific, business, cultural, public and other domains the digital information is growing at an exponential rate, which also raises the concern towards preservation of the valuable information for future use. This valuable information in digital form is increasingly at risk for the vulnerable characteristics of the digital objects and rapid advancement of the technology and media. There is a growing concern by the international researchers and designated communities towards persistent preservation. A book left on a shelf for hundred years might be fine, but digital data must be read and checked constantly to ensure their integrity. This paper is an attempt to construct a dynamic and stable framework to solve the problem of persistent preservation of digital information.

Key Words: Digital Preservation, Digital Entity, Digital Archive, Metadata.

1.0 Introduction

Preserving digital documents, whether they are personal, public or even commercial, lie at the heart of the information society. Even though authorities paying lot of interest, research into their long-term viability and the meaningful accessibility of their contents remains unresolved. The enormous knowledge generated by the international communities in the form of digital 'bits' raises the concern of long-term preservation of today's digital information. This knowledge, in the institutional concept sometimes referred as 'intellectual assets' or information generated by various activities in digital form, or converted to digital form, demand measures to be taken to keep their integrity for many years from now. The specified period is generally driven by various compliance and regulation that varies in different parts of the world.

Three significant obstacles persists in the world of digital preservation-

A. Constant growth of the volume of digital information that demands long term preservation;
B. Rapid advancement of formats in which the bits of information are encoded, the technologies and approaches in use, and the medium in which those bits are embedded;
C. New compliance legislations to meet that are emerging worldwide.

This paper follows the obstacles; traditional procedures; general criteria to follow to define an ideal and persistent preservation environment; that meets most of the demands of a preferred solution of the long-term preservation of digital information.

*Lecturer, Faculty of Business, ASA University Bangladesh

1.1 Background

The concept of preservation has experienced a radical change since last decade by the digital environments. Generally the concept of preservation is treated as keeping things unchanged for a long period. The rapid advancement of technology and storage media shift the view of keeping things unchanged to a 'pre-planned, systematic and consistent change'.

1.2 Objectives

The term 'preservation' recognises a particular state of the archival concept which deals with the human recognisable interpretation of the entity and concerns over the life-span of the digital entities that is longer than the underlying preservation media that holds the bits. The emphasis is on the digital form of information and retention of this information for a certain period of time to comply with the State and Organisational legislations. So the Research Objectives are to-

i. Develop an understanding of interactive, robust and experiential solution to help realising the activities of capturing, formulating, maintaining, refining, accessing and disseminating digital objects;
ii. Analysing generic processes and approaches of preservation solution to explore the actual problem domain of long term preservation paradox;
iii. Outlining of a Proposed Solution which incorporates all the generic archival operations and employ approaches and mechanisms from emerging technologies to help support distributed, redundant and complete solution of the problem

2.0 Generic Operations on Digital Entities for Preservation

Persistent preservation of the digital entities can be quite complex because a digital entity can be a single document, or a compound document with multiple components. A digital entity can also be the archival collection that is assembled by grouping multiple documents along with their descriptive, origin, and authenticity metadata. The generic operations, which are developed by the Persistent Archive Research Group of the Global Grid Forum[i] to preserve digital information in longer period of time. And they are:

Appraisal: The operation recognises disposition of the digital entities and recognises various terms and conditions to determine the selection of the entities which needs long term preservation even after their regular life-span.

Accession: The process establishes a formal acceptance to the dealt digital entity and acknowledges the acquisition; which means a formal registration of digital entities into the preservation schedule.

Arrangement: It defines the digital entities into archival system context. The processes used to recognise a digital entity after acquisition by defining its context attributes against the repository structure. Conceptually this can be viewed as logical representation of the digital entities into a repository collection hierarchy.

Description: The process describes the digital entity's form by which it is recorded, which can map the structural, functional and content issues in an entity. Description recognises the format of the entities in which they have been encoded.

Preservation: The process identifies the measures needed to retain recorded digital entities for a long period; which includes replication for disaster recovery and also pretend consistent authentication, integrity compliant and active mechanism to recognise the readability of the digital entities.

Access: The operation performs the retrieving of the desired digital entities by establishing a searching mechanism using their respective descriptive Meta information to spot them into the specific repository location. The process also recognises the transition of the digital entities from the repository to the access point.

2.1 Major Impediments of Persistent Preservation

Despite increasing concern about digital preservation, there are numerous technical, organizational, legal and economic barriers to a comprehensive infrastructure for protecting and preserving digital assets. The most familiar problems in digital preservation are media failure or deterioration and rapid changes in computer hardware and software that make older systems obsolete on a regular basis. The general obstacles that impede the long-term preservation of digital objects are:

Technological Obsolescence: One of the major issues regarding consistent preservation of digital objects is the fragility of the physical media upon which the objects are stored. Whatever appropriate storage environment is set, still the digital media would have a specific life-span. Moreover, new technology inventions and deployment of new storage devices and software changes the old environment in terms of products and methodology employed earlier to record, preserve, retrieval and the dissemination of the digital objects.

According to the researchers the technology life-cycle revolves into merely 18 months of time frame. That implies approximately every one and half year new bunch of technology replaces the old ones. And for the storage medium it takes 5 to 10 years to shift from the old to an advance one. Taking into consideration of these rapid technological changes, most advanced medium of storage cannot ensure a comparatively long life-span. That is the reason all the efforts have been taken, to outline the strategy of storing digital objects within a storage media, always keep an eye on the partial and most often a short term solution to the question of preserving digital objects.

Data Migration: Migration is a kind of advancement of the concept of refreshing, which is a periodic activity of transferring digital objects from one version of hardware-software to a new one or one generation of system to an advance one. The actual goal here is to transform objects from one state to other in a way which would not compromise with the content integrity and authenticity; though it is not possible to make an exact replica of the previous copy to the new version.

Migration is not a relatively new approach though, it is a time-consuming, rigorous process and the existence of vulnerability to data loss or corruption shows the fragility of the concept.

Migration always requires a systematic regeneration or a total new approach to deal with each different kinds of format.

State and Organisational Legislations: The vulnerability of the digital object preservation for longer periods makes it difficult to the preservation authority to cope with various rules and regulations specified by the state and organisations. The situation gets worse when taking into

account the changing nature of these legislations to confront the upgradation of technology and medium for storage.

Now-a-days most of the organisational intellectual properties are kept in digital form. The migration operation which involves regeneration of the encoding format of the digital entity could create a complex situation among the involved parties. Moreover, when the preservation methodology cannot promise the authenticity and the integrity of the digital entity, the organisational intellectual assets and other various evidential documents become useless.

2.2 Ideal Preservation Solution

Building a sound preservation solution for a specific organisation promises a lot of issues to be dealt with to capture, retain, manage, audit, search, retrieve and destroy the records. The preferred solution should be extensive and robust enough to comply with the organisation's every regulatory and legal need which will in terms significantly reduce the cost of employee training.
The major emphasis should be on the following issues for making a complete Archiving framework –

1. The archiving system should have ability to retain all the files, records of transactions, emails and other documentations to support organisational regulatory decisions, corporate governance and legal litigations.
2. To ensure complete compliance against the above, the system should be equipped with proper functionality which will ensure secure and total capture and disposition of records, advance and complete searching technique to satisfy timely presentation of documents for legal purpose.
3. Management of records throughout their lifecycle with comprehensive audit ability checklist to ensure eventually the enforcement of corporate record classification against organisational policy.

The ideal solution should be a uniform, extensible, flexible and sustainable solution which approach would be the same approach to be undertaken each time, the records of a universal range to be upgraded in future. In other words, it should be capable of having a 'consistent' procedure to convert, translate and organise the records from the previous state to the presently acceptable form. The word 'consistent' implies-

- Does not need continuous evolution of methodologies
- Does not require repeated changes of approaches to deal with the new record formats or software-hardware paradox.
- Does not involve continuous regeneration of recordkeeping theories and practises.
- Does not imply rigorous translation or transformation of each individual record.

2.3 Metadata Concept

Metadata is usually identified and classified as 'data about other data'. In its simplest instance metadata can be described as the list of all the electronic journals of a particular university library, where a single entry will include the name of the author (s), title, topic, year of publication, etc. Linguistic differences could exist in different parts of the world but without constituting a firm metadata structure, an archiving solution on whatever field cannot exist at the first place.

Several international metadata standards exists[ii] namely Dublin Core, ONIX, MARC 21, DOI, CERIF, GILS, IEE/LOM and they contain different terminology to define data. In December 2002 the individual responsible bodies of these standards had declared a consensus named [iii]CORE-RESOLUTION stating that they have reached a common platform where, in their language, the meaning of units of the individual standards can be 'comparable' and 'mappable'. This is a very useful situation for the world for archiving solution to build a globally shared preservation metadata structure.
However, the situation becomes complex when defining the specificity in the metadata structure while maintaining robustness in the resulting structure to define broad applicability.

2.4 Approaches of Designing a Persistent Preservation Solution

The central issue is revolving around the 'format specification' under which the digital material is encoded and the 'unpredictability' of the future generation computational features (and formats).
There are three approaches (Moore. R., W., 2003) which has being focused by the researchers to overcome the 'unpredictability' of the digital data and they are-

i. Data Migration: This concept ignites an archival activity to transform the original encoding format to a new one that is recognised by the advanced (available) computational applications.
ii. Emulation: The exact differences between the Emulation Technique and the Data Migration is, instead of migrating the format specification of the digital entity this approach tries to 'migrate' the underlying software application and the original operating system environment (on which the application is supposed to operate on the digital entities, to present them in a human recognisable form) to the new generation of computing system (which also includes the hardware specifications into an 'encapsulated' framework).
iii. Ontology Migration: involves the conceptualisation and characterisation of the data model in order to map the semantically, procedurally and structurally relevant objects into appropriate Ontological framework and employ migration operations into these frameworks to retain the authenticity and integrity of the digital objects.

3.0 An overview on the existing research initiatives

3.1 InterPARES Project

The International Research on Permanent Authentic Records in Electronic Systems[iv] (InterPARES) project began in 1999 with a vision to developing the conceptual and methodological framework necessary towards the preservation of digital information maintained in an integrated and authentic fashion for indefinite time. However the complexity raises handling objects having multiple and variable entries into the database, which requires representing themselves based on the content generated from the database by various operations.

In the year 2002 the IntePARES project team with additional researchers of various disciplines began the second phase of the project by labelling it as InterPARES-2 which has completed in 2006. The InterPARES-2 is basically dealing with the complex objects which could be interactive, dynamic or experiential in nature.

3.2 Data Grid Technique

The revolutionary data grid technology uses the concept of data management abstraction to handle digital information spread over multiple administrative domain.
Theoretically, management of digital information in data grid concept incorporates four basic categories or naming conventions, they are-

a. To implement the virtualization of data storage, the naming of resource is necessary to have control on accessing data from different administrative domains.
b. To impose security measures like implementing single-sign, identification of the actual person the user names should be evaluated across the administrative domains.
c. To implement data virtualisation, identification of file names across the different administrative domains is necessary.
d. To implement federation of data grids, processing of context attributes is required to understand and manage various states of the digital entities caused by the various activities and time.

3.3 Emulation Approach

Emulation concept comes basically from the idea of keeping the obsolete hardware and software in a secured museum environment to use them reformatting the digital content into forms that are recognisable by the advanced generation of software. Jeff Rothenberg[v] (January, 1999), a senior researcher of the RAND Corporation suggested the technique of emulation which is capable of creating the environment under which the extinct software can run on the present or future computer to make the preserved content readable.

The emulation technique cannot guarantee to keep all the attributes of the digital entity intact by the time it would be translated under emulated environment, however, as long as the authenticity and integrity- the most important features are not been compromised in any way, the technique can fulfil the requirements of long term preservation.

The most vulnerable side of emulation concept is- it is based on assumptions and the future computer system will generate the category of operating system calls, which the emulator attached to the digital entity is capable of mapping with the original system calls recognised and specified in the emulation specifications.

4. Designing a Persistent Preservation Framework

A typical characteristic in the design of the digital preservation solution is to deal with the fulfilment of different requirements assigned by the different user communities; resulting into a steady growth of digital objects and their associated resources. The possible solution which will incorporate all the generic archival operations to help support distributed, redundant and complete preservation consists of four sections and they are

i. Capturing the data (Data Capture)
ii. Managing the data (Information Management)
iii. Archiving the data (Archival Storage)
iv. Accessing the data (Access Control)

4.1 Capturing the Data

The processes of capturing the data are described below-

Receive Digital Object process simply receives the submission of articles and their associated identification information (e.g. Depositor or Author's name, registration number, etc.). In some cases, this transfer of objects can include a formal declaration of a change of legal obligation/authorisation over the object from the author or the depositor. The identification information associated with the incoming object would assist the administration to formulate the decisions (e.g. the relevance, importance or how long to archive).

Generating Copyright Compliance process avoids legal complications over the submitted objects by forming the agreement with the author as the object submitted is by all means their own piece of work or the depositor is submitting the object on behalf of the author carries the rights to do so. This declaration information is stored in a special log book called declaration log, which would be kept safe to avoid the legal obligation.

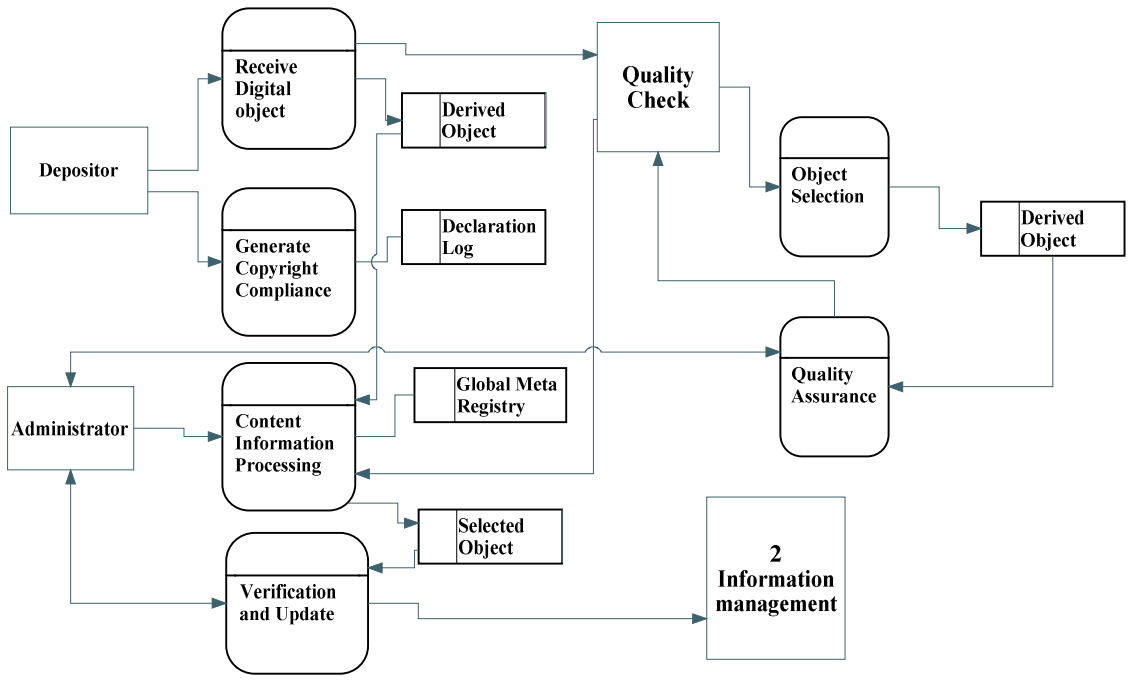

Fig. Capturing the Data

Content Information Processing makes a formal acceptance by recording the acquisition of the object by registering an entry of derived digital entity into the 'metadata registry'. This operation adds a semantic tag and a timestamp to the each. The registry is basically kept independently from the underlying storage systems which hold the actual digital content.

This process generates functionalities to extract the metadata from the actual content and follow a certain documentation standard to form 'structured information' to make it ready for preservation. Structured information specifies the metadata and the contextual information about the content and makes it ready for preservation in its original format.

Verification and Update: Before updating the 'content' this functionality verifies the 'context' of the digital entity. The verification mechanism works on the 'provenance' information that describes the creation of the entity, relationship with the other entities (semantic and temporal tags) into the digital collection. This function is also responsible for recognising the category of Ontological Structure the dealt entity should belong, this decision depends on analysing the four types of relationships namely, structural, logical, spatial and temporal relationships with the other existing entities in the collection. Based on the decision the function puts the value into the 'Ontology-tag' of the entity into the Global Metadata Registry.

Quality Check

Object Selection determines the nature of the digital entities; especially make selections over the digital entities applying various terms and conditions to finalise the decision of preserving the dealt entity. These terms and conditions are established by the management body of the organisation.

Quality Assurance function issues a request to the Administrator to send the Original form of the digital entities as they arrived to the Admin domain. Admin function responds while the administrator goes through the whole document and ensures-

a) The digital entity is relevant to the field specified,
b) Understandable by the designated subscriber community without the help of the originator,
c) Overall documentation satisfies standards and
d) Change of custody over the digital document is declared.

The process sends the resulting document from the administrative entity to the capture module. For this it issues a request for transfer to the 'Content Information Processing' function of the capture module and transmits while gets the confirmation. The standard transfer protocol is used in this process of transferring the digital entities between functional modules.

4.2 Information Management

The processes under this module are-

Receive Object Updates incorporates the processes of sending confirmation to the Capture module and keep administrator informed in a periodic basis regarding the status of the repository updates. The capture module provides the metadata information about the new digital entities which this modules spread over two places- administrator verification and the underlying database. This module attaches the repository status tag to the metadata structure.

Administering Information performs two basic functionalities, they are- organising the entries at the global metadata registry in a hierarchical fashion and categorising the format specification upon which the digital entity is encoded. For the capability of identifying the individual derived entity, a persistent logical identifier is established at this stage which is unique for every digital entity and independent of the remote storage system file names. This operation adds an extensive

and unique collection of attributes to the digital entities and this attribute set establishes a pointer to the global metadata registry to represent each individual entity distinctly.

Categorising the format specification supports the infrastructure independent representation of the digital entity. In this process the digital entities are converted from its proprietary encoding format to an established encoding standard.

The other responsibilities involve holding the integrity and verifying the authenticity of the structured information entities. The Meta information identifies the actual set of content of each digital entity and the repository status information generates from the received article information updates module recognises the operations held over the new entities. Administrating information operates on the strategy set by the administrator functional entity. Thus this module handles the responsibilities of creating and structuring digital entities by the concept of 'information repository abstraction'.

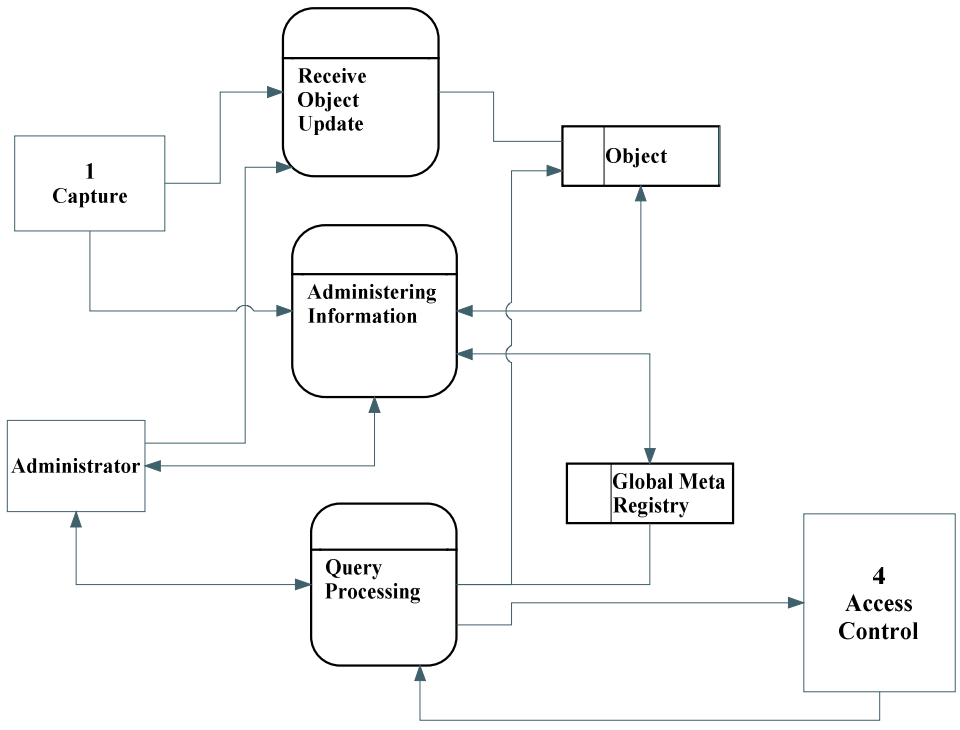

Fig. Information Management

Query Processing handles the query request from the access module and executes the request that generates the result set to transmit back to the access module.

4.3 Archival Storage

Function incorporates the following processes-

Receive Object function receives the request from the Capture module for the incoming digital object and upon completion of the transfer of the structured content of the object it gets back to the Capture module by issuing a confirmation of receipt. The request for the incoming article content can be comprised of the indication about appropriate storage media types, volumes, etc.

Storage Structure Management deals with the appropriate steps to store the object's content to the most applicable and appropriate media. This functional entity is governed by the storage policy and structure set by the administrative functional entity which is eventually governed by the management decisions. In some cases the (Hedstrom 2002)directions proposed by the Capture module are also to be taken into account towards the storage of the digital objects. Eventually, all the derived digital objects are characterised into their appropriate ontological structure.

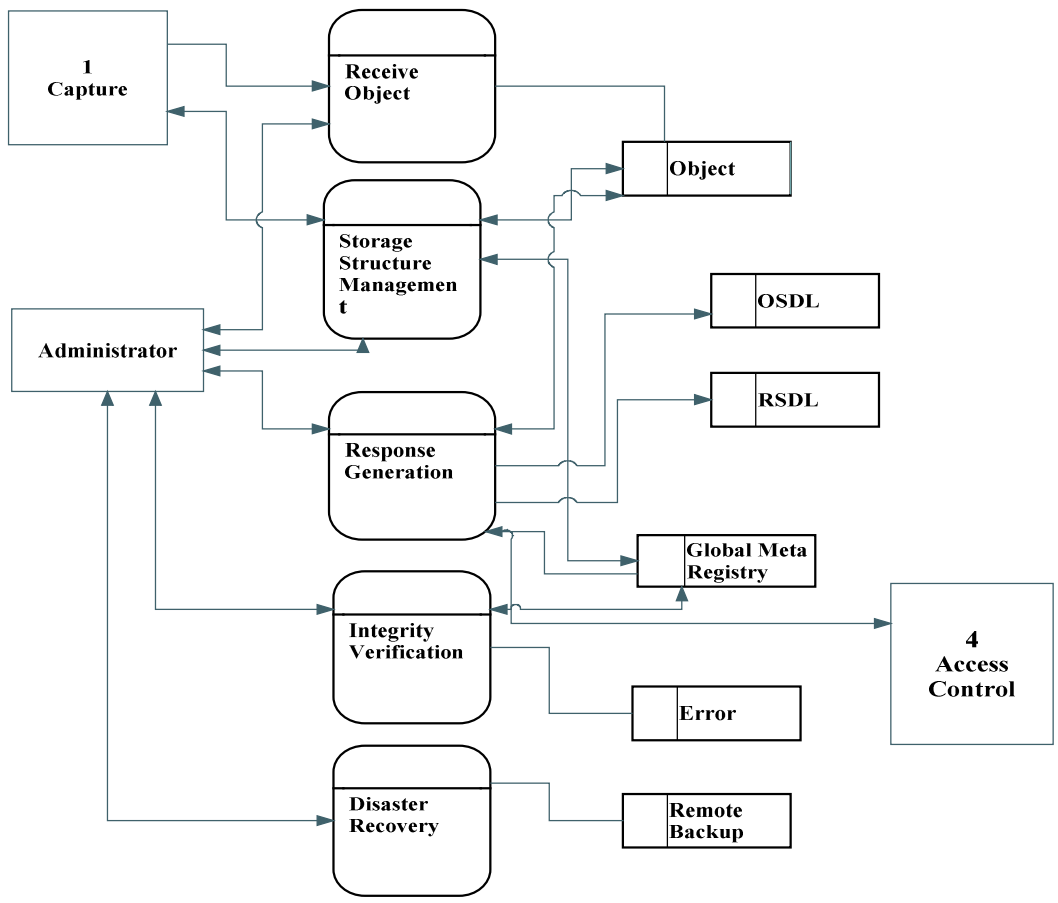

Fig: Archival Storage

This function also incorporates the responsibility of providing any extra measures to the content of the digital objects like any special confidentiality measures that can impose an extra authentication and/or authorisation inconsequence to the request for a view from the access module. These added responsibilities can ignite actions towards the consideration of the storage type, place (for example storing at a remote site away from typical storage space), imposing handling constraints or an extra backup, etc.

The 'Operational Steps Description Log' (OSDL) holds the specific characteristics of the structure of the digital entities. This structure defines the order in which the relationships need to be applied to interpret the digital entity. On the other hand, the set of operations performed by the different categories of repositories are specified in a structured collection named 'Repository Steps Description Log' (RSDL), which holds all the necessary documentation of the System Software (OS) Specification and underlying Repository (hardware) Specification corresponding to each entry into the OSDL.

This function takes the added responsibilities of generating reports for the administration regarding the operations performed over the content of the objects in both periodic and special consideration basis and also enhancing the responsibility by producing the periodic documentation.

Response Generation incorporates the responsibility of generating response to the requests made by the Access module. This function basically gets the requested content of the article straight from the specified repository block and sends it to the access point. If it experiences any constraint imposed by the storage strategy structure then the Storage Structure Management function gets activated in an automated fashion and issues a decision request to the Administrative functional entity. The function will be in a wait state until it gets the response from the Storage Structure Management function.

Integrity Verification function ensures the integrity of the structure of the digital objects while transferring the content between the functional modules from the repository. The basic focus of verifying the integrity is to keep the administrative attributes of the digital entities consistent even after a migration operation operated on certain category of digital entities. This function identifies all the 'potential breach of integrity' over the digital content from analysing the notifications received from the functional modules and reports back to the Administration.

Disaster Recovery function is outlined by the Administrative entity in its disaster recovery policy. The functionality supports mechanisms to create replica of the contents stored in the repository and maintains them in a remote, separate storage medium.

4.4 Access Control

Highlights the following processes-

Coordinate Access function handles the single point of interface between the access point and the digital collections of the repository. This interface could be established through internet or intranet services or through a WAN or LAN, etc.

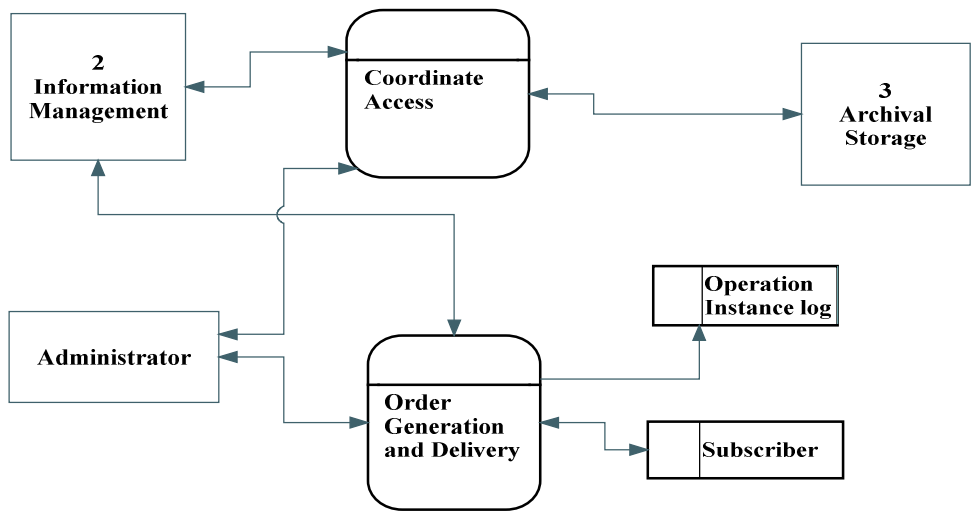

Fig: Access Control

The request for queries are transferred straight to the Information Management module where Query Processing function performs the necessary actions and upon completion of the getting back the result set from that it displays the response to the access point. The request for the digital content may involve either or both modules namely Archival Storage Function and Information Management to formulate the request. The order of the digital content can be either in an ad hoc basis or a special consideration ignition based action for the Administration functional entity which will be handled by the 'condition based activation' for the order/request in a automated fashion which would be resulting into a periodic generation of the requested material.

Order Generation establishes the access to the desired digital entity as long as its pointer is existing in the global metadata registry. The **Order Delivery** schedule generates the mechanism for accessing digital objects in a distributed environment across multiple administrative domains. This facility includes processes for moving digital content and their attendant metadata in high volume and at the same time configuring the authentication of the access points across the administrative domains.

5.0 Evaluation of the Proposed Framework with Generic Archival Operations

The generic operations on digital entities listed earlier in this paper (appraisal, accession, arrangement, description, preservation and access) could be mapped into the proposed framework.

APPRAISAL: To identify digital entities, the proposed framework selects entities by evaluating them under various conditions for long-term preservation established in the Quality Check Module.
ACCESSION: To provide the formal agreement of acquisition of the digital objects, the Object Registration of the Capture Module undertake the formal acceptance by making a unique entry to the infrastructure independent Metadata Registry.

ARRANGEMENT: The proposed framework organizes and manages the context information associated with each digital object. Context defines the relationship between different entities into the collection. The Verification and Update activity handles the organisation of context information. In case of discrepancy the mechanism notifies the admin. The organisation of context information also helps determine the level of granularity which is applicable to the digital collection.

DESCRIPTION: This function incorporates the appropriate recording of the digital entities, their associated functions and other attributes. The framework supports this consideration by describing individual entity distinctly into its Global Registry. This is a baseline requirement for the persistent archival solution to assign its collections a persistent logical identifier.

PRESERVATION: This is the aim of the whole effort to be carried out by archivists. Certain considerations exist while planning the continuous preservation of the digital objects. The proposed framework holds two kinds of logs to support repository and administrative domain independent preservation and available access of the desired documents, namely Operation Step Description Log (OSDL) and Repository Step Description Log (RSDL).

OSDL emulates the Steps Descriptions to execute the representation application of the specific categories of digital entities. And RSDL holds the emulated version of the system software and the platform specification that corresponds to the each entry established into OSDL. These Logs are directly handled by the 'Storage Structure Management' under the Archival storage Module.

The advantage of maintaining this type of architecture is- each time a need arise for migration, instead of migrating the encoding format of the individual digital entities it can be migrated. It also supports the 'redundancy consideration' to preserve and access of the digital entities because of its infrastructure independent nature. Disaster recovery functionality provided in the framework which ensures the replication of the content and metadata for each entity and this could be kept in a remote site to avoid local disaster over the archival systems.

ACCESS: This function uses the 'descriptive metadata' to identify desired digital entity and retrieve them from specified location. Access incorporates the capability to spot the document of interest and establishes interfaces to retrieve the desired material.

6.0 Understanding the Framework against the Core Capabilities of Data Grid Technology

The framework proposed in this paper has constructed some key features of persistent archival concept. By analysing the core capabilities of the several data grid mechanisms, the elements needed to maintain a persistent archive following Grid features can be identified.

Logical Name Space: the proposed framework supports the 'name transparency' of the Persistent Archival solution which is constructed in the Global Metadata Registry of this model. This registry defines each digital entity by establishing a name which is unique across the repositories and administrative domains under this registry. A mapping is constructed between logical name and the physical existence of the digital entity through establishing a pointer at the time the entity is going to be selected for preservation.

Information Repository Abstraction: The proposed framework maintains an infrastructure independent organisation of digital entities into a logical collection hierarchy. This hierarchical collection namely Global Metadata Registry organises semantic correspondence to each existing digital collection in its global name space. The Registry constructs administrative, descriptive, context and provenance conceptualisation into ontological structure. Infrastructure independency implies flexibility of restructuring the database scheme.

This concept is used by the Administering Information Content sub-process under the Information Management functionality. The capability of identifying the individual derived entity, a persistent logical identifier is established at this stage which is unique for every digital entity and independent of the remote storage system file names. This operation adds an extensive and unique collection of attributes to the digital entities and this attribute set establishes a pointer to the global metadata registry to represent each individual entity distinctly. Categorising the format specification supports the infrastructure independent representation of the digital entity.

Storage Repository Abstraction: The proposed framework has two types of Logs, *Operational Steps Log* and *Repository Steps Log*. Operational Steps Log incorporates all the steps required to manipulate different categories of digital entities while the Repository Steps Log holds all the characteristics of different categories of repositories to execute the digital elements. These Logs can be characterised as 'Knowledge Repository'. In this way both the digital entities and their appropriate repositories can be replicated. This Log also helps to define the protocol needed to operate on digital entities which belongs to the specific categories of repositories.

Distributed Resilient Architecture: The proposed framework supports Client-Server Architecture. The 'implementation' of this framework should recognise local servers established with each local repository. Each request to the system generates 'interaction-state', which eventually retrieve all the related elements and their state information from the Global Metadata Registry thereby interacts to the desired Local Server. When the request failed to retrieve desired elements the automated search mechanism looks for a replica within the Global registry and in case found a match the request is redirected to that local server under which a replicated version

is found into the respective repository. Each interaction-state operations are recorded into the Operational Steps Log within the respective digital collections' state attributes.

7.0 Limitation of the Proposed Framework

While preserving an institution's all intellectual assets digitally and indefinitely; it is hard to predict the future computing system that possibly recognises the form used to encode information digitally in today's world. A number of potential solutions exist, however, these solutions fall short of delivering an optimum solution. Even in this paper the framework proposed is not free from flaws. While following this framework, the underlying representation application, system and repository specification accumulates a high volume of extra material to be preserved with the actual digital entities into an ontological structure which is not ensuring an optimum solution of the problem domain. As a result, this paper tries to describe a solution which is close to the optimum solution.

8.0 Conclusion

Recent progress in digital preservation is a consequence of the growing awareness of longevity as a critical issue for sustainable and useable digital libraries, increased investments in research and development, and efforts to focus on discrete and potentially solvable aspects of the problem. The proposed solution of this paper for preserving digital information illustrated all the issues relevant to a digital entity that starts with its origination, exists with various operations and transformations throughout its life-span and ends with disposition, and of course the roles and responsibilities of designated individuals associated with it. At the end, this paper is an initiative to establish a Proposed Framework based Solution concerning Persistent Preservation of Digital Information over a longer period of time; by accumulating different theories and approaches and meeting its critical success factors the Framework outlines a dependable solution for Long-term Preservation, without compromising Authenticity and Integrity of Digital Information.

References

Calanag, M,L, Sugimoto, S & Tabata, K 2001, *A Metadata Approach to Digital Preservation*, viewed 13 January 2012, <"http://www.dcpapers.dublincore.org/ojs/pubs/article/download/653/649"

Chen, S,S 2007, 'Digital preservation: Organisational Commitment, Archival Stability, and Technological Continuity', *Journal of Organisational Computing And Electronic Commerce 17(3)*, pp. 206-209.

Dawson, C,W 2005, *Projects in Computing and Information Systems, A Student Guide*, Addison Wesley.

Duranti, L 2002, *International Research on permanent Authentic Records in Electronic Systems*, viewed 5 February 2012, < http://www.interpares.org/ip1/ip1_index.cfm >.

Hamer, AC 2006, *The ICSA Guide to Document Retention*, ICSA Publishing Limited, London, UK.

Hedstrom, M 2002, *Digital Preservation: Problems and Prospects*, viewed 15 February 2012, < http://www.dl.slis.tsukuba.ac.jp/DLjournal/No_20/1-hedstrom/1-hedstrom.html >.

Hedstrom, M & Montgomery, S 1999, *Digital Preservation: Needs and Requirements* , viewed 7 June 2012, < http://www.oclc.org/research/activities/past/rlg/digpresneeds/digpres.pdf >.

IBM Haifa Research Lab 2007, *The Need for Preservation Aware Storage*, viewed 25 January 2012, < https://www.research.ibm.com/haifa/projects/storage/datastores/papers/preservation_data_store_osr07_dec_30.pdf >.

InterPARES1 Project 1999, *International Research on Permanent Authentic Records in Electronic Systems*, viewed 5 February 2012, < http://www.interpares.org >.

InterPARES-2 Project 2006, *International Research on Permanent Authentic Records in Electronic Systems*, viewed 5 February 2012, < http://www.interpares.org/ip2/ip2_index.cfm >.

Lynch, C 1999, *Canonicalization: a fundamental tool to facilitate preservation and management of digital*, viewed 13 January 2012, < http://www.dlib.org/dlib/september99/09lynch.html >.

Moore, RW 2003, *Persistent Archive Concepts*, viewed 18 January 2012, < http://www.ggf.org/documents/GFD.26.pdf >.

Moore, RW 2007, *Building Preservation Environments with Data Grid Technology†*, viewed 5 February 2012, < http://www.interpares.org/display_file.cfm?doc= ip2_gs01_final_report.pdf >.

National Information Standards Organization 2004, *Understanding Metadata*, viewed 15 February 2012, < http://www.niso.org/publications/press/UnderstandingMetadata.pdf >.

Rothenberg, J 1999, *Emulation Solution, Avoiding Technological Quicksand:Find a Viable Technical Foundation for Digital Preservation*, viewed 2 February 2012, < http://www.clir.org/pubs/reports/rothenberg/contents.html/criteria.html#emulation >.

Wache, H 2001, *Ontology-Based Integration of Information - A Survey of Existing Approaches*, viewed 06 February 2012, < www.let.uu.nl/~Paola.Monachesi/personal/papers/wache.pdf >.

White, M 2005, *Metadata, The Content Management Handbook*, Facet Publishing, London.

Endnote

[i] Persistent Archive Concepts, Reagan W. Moore, 2003, Global Grid Forum

[ii] Understanding Metadata, NISO Press, National Information Standards Organization, 2004

[iii] Principles of Metadata Registries, A white paper of the DELOS working group on Registries.

[iv] The long-term preservation of authentic and accurate digital data: The InterPARES Project, Data science journal, Volume 4, 25th October, 2005.

[v] Rothenberg, J., Emulation Solution, Avoiding technological quicksand: Finding a viable technical foundation for digital preservation, (January, 1999)