A Dissertation submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

Digital Data Heritage Preservation (DHP) Modelling and Design

Submitted for the Degree of Doctor of Philosophy
Computer Systems

Lucia Cristina Carrion Gordon

Autumn 2018

University of Technology Sydney (UTS)
Faculty of Engineering and Information Technology (FEIT)
School of Electrical and Data Engineering (SoEDE)
Global Big Data Technologies Centre (GBDTC)

Supervisor

Dr. Zenon Chaczko

Date of the submission

19 January 2018

Dedication

I dedicate to my beloved family my son Benjamin and my husband Roger. My little angel, my love miracle who shows me what life is and how to be a researcher from his birth. For their continuous love, sacrifices and patience from the beginning to the end of my PhD. Thank you for all the important encouragement over this time. To my parents Lucia and Hugo, for their sincere prayers and encouragement throughout my life and my PhD. The constant support from my supervisor and best advisor on my research path: Zenon Chaczko.

Originality Statement

Certificate Of Original Authorship

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Production Note:
Signature removed prior to publication.

Signed ....

Date: 1st June 2018

Acknowledgments

I would like to acknowledge all the support and encouragement received during my PhD research. Firstly, I am indebted to my supervisor Dr. Zenon Chaczko who have been my supervisor and very good friend. For his continual guidance, support and suggestions. Without his insight, this work would not be possible. I would like also to thank him for his suggestions throughout the reviewing stage. His participation was vital to the completion of this thesis. His valuable guidance through this research was the opportunity to learn from his knowledge and experience. His frequent insights and patience with me are always appreciated.

I would like to express my sincere and deep gratitude to Dr Zenon Chaczko and his wife Margaret Wajs, who have been present during this journey. Thanks for always being there for me when I needed and for the nice hours of chats. Also, I would like to express my special appreciation and thanks to Professor Robin Braun, Associate Professor Dr Qiang Wu and Dr. Karla Navarro Felix at UTS University of Technology Sydney for their support in my Candidature Assessment. During the course of this research, I also benefited greatly from interactions and technical discussions with other talented and warm-hearted members from the CRIN centre. I would like to thank to Frank Zeichner, Industry Associate Professor, School of Systems, Management and Leadership for his special criteria around my innovative topic. His contributions guide me in the last period of my thesis. I must also express my thanks to my colleagues and friends in the Department of Computing and Communications at UTS University of Technology Sydney, Raniyah Wazirali, Pakawat Pupatwibul, Wael Alenazy and Jan Szymanski. For academic and experimental contributions through my research in publications to my friend Gabriela Lopez, Jaime Reinoso and Jason Gavriel.

Throughout the course of this Thesis, there have been several people who have afforded me with direction, motivation and provision for finishing this work. I would like to thank University of York, Prof Tony Ward and Bidyut Baruah. I spent great time in York during the conference in 2014. Also my colleague Annett Pfeffer from

TU Dresden, Germany. We always spent good time together in any place of the World.

On 2016 I had the opportunity to be a Lecturer in Hagenberg, Austria. Thank you to the people at the University of Applied Sciences Upper Austria (FH Oberösterreich), School of Informatics, Communications and Media Softwarepark, in Hagenberg, Australia where I spent my research visit. My sincere thanks in special Univ.-Prof. DI Dr. Witold Jacak (Academic Head), FH-Prof. DI Dr. Heinz Dobler , FH-Prof. PD DI Dr. Michael Affenzeller (Head of research group HEAL), DI (FH) Thomas Kern (Head of Research Center Hagenberg Campus) and FH-Prof. Mag. Dr. Berthold Kerschbaumer (Dean of School of Informatics, Communications and Media, Hagenberg). I want to thank them for their brilliant comments and suggestions and for their support during that time. I really enjoyed sharing those days, learning new skills and technologies. Hagenberg kept part of my heart. This town university teach me how the simple and wonderful life can be combined with learning and innovation.

Special thanks to the Faculty of Computer Science University of Vienna Prof. Dr. Dimitris Karagiannis Head of Research Group Universität Wien and Dr. Dominik Bork Postdoctoral Researcher - Fakultät für Informatik - Universität Wienthe Manager for their exceptional contributions to the project. Far or close we engage our common interest in Data Modelling.

On the personal front, I would like to express the wholehearted gratitude to my beloved son Benjamin and my dear husband Roger. Thanks for your unending love, endless support, dedicated time, patience, encouragement and understanding in every possible way. My parents Lucia and Hugo for their nonstop love, truthful prayers and support during the research. I would like to thank my best friends Laura and Lorena, my little daughters. Although, the circumstances of the life have kept us geographically distant in the past years, our hearts have always been and will continue to be together caring for each other. My special recognition to Yoli Mancero, my special friend and buddy. Her support and help in the precise moments will be always remembered in the bottom of my heart. To close friends who have always supported in all my endeavors, to all my sincere friends. My appreciation to my parents-in-law Mariana and Victor. They always are present, no matters the distance. Thanks for your love.

To my country Ecuador for providing a chance and funding to engage in this research. The trust on my skills and the conviction that the education is first makes me feel proud of what I have achieved. It is the opportunity to contribute from the academia to the development of my country. I would like to thank SENESCYT (Secretaria de Educación Superior, Ciencia, Tecnología e Innovación) and IFTH (Instituto de Fomento al Talento Humano) for supporting my PhD degree financially. They generously funded my candidature and provided the financial resources for the completion of this project. This research work would not have been possible without their assistance. Thanks for the trust on my skills, I am very proud to achieve the goal of achieving the Doctoral degree.

Nomenclature

BD	Big Data
BP	Business Process
BPM	Business Process Management
D2HP	Droid-based implementation of Digital Heritage Preservation Model
DHP	Data Heritage Preservation
DM	Data Management
DP	Digital Preservation
DROID	Digital Record and Object Identification
DS	Data Set
DT	Digital Trasformation
EA	Enterprise Architecture
IT	Information Technology
METS	Metadata Encoding and Transmission Standard
OLTP	OnLine Transaction Processing
PAIS	Process Aware Information Systems
SH	Serendipitous Heritage
SOTA	State of the Art
XENA	Xml Electronic Normalizing for Archives

Related Publications

The technical contributions and discussions in the thesis are mainly based on the following publications written by the author:

A) International Journal Publications

[J1] Carrion L., Sanchez J. “Architectural Framework to Preserve Information of Cardiac Valve Control”, World Academy of Science, Engineering and Technology International Journal of Information and Communication Engineering Vol:2, No:11, 2015

[J2] Carrion L., Lopez M., “Preservation Model to Process ’La Bomba Del Chota’ as a Living Cultural Heritage”, World Academy of Science, Engineering and Technology International Journal of Information and Communication Engineering Vol:2, No:11, 2015

[J3] Flores M., Horna L., Escobar A. and Carrión L., “Estimation of the Contaminant Risk Level of Petroleum Residues Applying FDA Techniques” LAJC Latin American Journal of Computing, Faculty of Systems Engineering Escuela Politécnica Nacional Quito-Ecuador, 2017

B) International Conference Publications

[C1] Carrion L. & Chaczko Z.,“Digital preservation strategy: Planning procedure to preserve Critical Information of Heritage”, In Proceedings of the 2nd Asia - Pacific Conference on Computer Aided System Engineering, APCASE 2014 Extended Abstracts, 10th -12th February 2014, South Kuta, Bali, Indonesia, page(s) 71 -72, APCASE Foundation, ISBN 978-0-9924518-0-6

[C2] Wazirali R, Carrion L. & Chaczko Z.,“ MultiIlayers DNA - QR Based Steganography” In Proceedings of the 2nd Asia - Pacific Conference on Computer Aided System Engineering, APCASE 2014 Extended Abstracts, 10th -12th February 2014, South Kuta, Bali, Indonesia, page(s) 50-51, APCASE Foundation 2014 ISBN 978-0-9924518-0-6

[C3] Carrion L., Chaczko Z., Alenazy W & Mu M1 “Development of an Expert System to assist in Resource Management”, IEEE Xplore ITHET 2014 York, England, 11-13 September 2014

[C4] Carrion L., Chaczko Z., Braun R., Dagher J. “ Design of Unit Testing using xUnit.net”, IEEE Xplore ITHET 2014 York, England, 11-13 September 2014

[C5] Carrion L., Chaczko Z., Alenazy W, & Tran A.” Augmented Reality Based Monitoring of the Remote-Lab”, IEEE Xplore

[C6] Wazirali R., Chaczko Z., Carrion L. and Slehat S., “Review of Quality Metrics Assessment for Steganography” pp. 1–8. APCASE 2015, Quito, Ecuador 14-16 July 2015

[C7] Alenazy W., Chaczko Z., Carrion L, and Chan C.Y., “Haptic Middleware Based Software Architecture for Smart Learning”, APCASE 2015, Quito, Ecuador 14-16 July 2015

[C8] Carrion L., Sanchez J., “Architectural Framework to Preserve Information of Cardiac Valve Control”, The 26’International Business Information Management Association Conference. November 11-12, 2015. Madrid, Spain. ISBN: 978-0-9860419-5-2

[C9] Chaczko Z., Carrion L., “Standardized Mapping Model for Heritage Preservation and Serendipity in Cloud”, The 26’International Business Information Management Association Conference. November 11-12, 2015. Madrid, Spain. ISBN: 978-0-9860419-5-2

[C10] Chaczko Z., Carrion L. “Towards Digital Heritage Preservation Framework for Situation and Context Aware for Information Management”, ICCS 2016: 18th International Conference on Computational Science Barcelona, Spain August 11 - 12, 2016, International Science Index vol:10 no:08

[C11] Carrion L., Chaczko Z., Wazirali R., “Ontological Metamodel for Consistency of Digital Heritage Preservation (DHP)”, 2017 25th International Conference on Systems Engineering (ICSEng), vol. 00, no. , doi:10.1109/ICSEng.2017.67 Las Vegas, USA, August 22-24, 2017, ISBN: 978-1-5386-0610-0 pp: 438-442

C) International Book Chapter Publications

[B1] Carrion L. & Chaczko Z.,“ Digital Patterns for Heritage and Data Preservation Standards ”, Springer Series: Studies in Computational Intelligence page(s) 47 -58, APCASE Foundation, ISBN 978-0-9924518-0-6, Series Ed.: Kacprzyk, Janusz, (ISSN: 1860-949X), Springer book is titled: Computational Intelligence and Efficiency in Engineering Systems, and edited by: Borowik G., Chaczko Z., Ford L.G., Jacak W., Luba T., Part I Computational Models and Knowledge Discovery

[B2] Carrion L. & Chaczko Z.,“ Bio-informatics with Genetic Steganography Technique”, Springer Series: Studies in Computational Intelligence page(s) 339 - 350, AP-CASE Foundation, ISBN 978-0-9924518-0-6, Series Ed.: Kacprzyk, Janusz, (ISSN: 1860-949X), Springer book is titled: Computational Intelligence and Efficiency in Engineering Systems, and edited by: Borowik G., Chaczko Z., Ford L.G., Jacak W., Luba T., Part IV Data-Oriented and Software-Intensive Systems

[B3] Carrion L., Chaczko Z., Resconi G., “Serendipity in Context of Digital Preservation and Artifacts in Large Infrastructure Oriented Systems”, International Conference on Computer Aided Systems Theory, EUROCAST 2015 15th International Conference, Las Palmas de Gran Canaria, Spain, February 8-13, 2015, Pages 110-117, Print ISBN 978-3-319-27339-6 Online ISBN 978-3-319-27340-2

[B4] Z. Chaczko, L. Carrion-Gordon, W. Bożejko, “The Metamodel of Heritage Preservation for Medical Big Data”

[B5] Z. Chaczko, R. Wazirali, L. Carrion-Gordon, W. Bożejko, “Steganographic Data Heritage Preservation Using Sharing Images App”

[B6] Z. Chaczko, R. Klempous, L. Carrion-Gordon, “Enabling Design of Biomimetic Middleware for Large Scale IOT-Based CyberMedical Systems”

Abstract

Motto: 'When the origin and heritage of data are lost, data has no meaning'

Dr. Zenon Chaczko

This study focuses on heritage concepts and their importance in an ever-evolving and changing Digital Era where system solutions have to be sustainable. The main idea of this research is related to the management of Heritage, directly in terms of preservation. It exposes an experimental methodology and a valid analysis of the results. The reliability and accuracy of data are very strategic points as the knowledge base is wide and complete.

This research proposes a model to explain how to build an accurate framework for Data Preservation. The relation between Preservation and Digital patterns of Heritage is well related due to the aspects of accessibility and context. They cover the conceptualization of real digital preservation. However, the availability, contextualization and value of the information remain the principal matters on which this research focuses. Firstly, in the introduction, the context is presented and the description of the initial scenario. Secondly, it addresses the process of preservation with modeling applications and the implementation of patterns. Finally, it presents the conclusions and future projects based on the findings. The principal objective of the research is the integration between models and standardization as a sustainable solution. A primary concern is the explosive amount of information. The complexity of the classification is how to manage the principal characteristics of data.

Digital Data and Heritage Preservation, as concepts, are related to data management, contextualization and storage. There are many issues and concerns related to this. This research explores the precise definition, context and need for patterns of heritage. The relations, interpretation and context provide us the appropriate methods to keep information for long-term use. The management of massive amounts of critical data involves designing, modeling, processing and the implementation of accurate systems. The methods to understand data have to consider two dimensions that this research has to focus on: access dimension and cognitive dimension.

Our cultural heritage, documents and artifacts increase regularly and place Data Management as a crucial issue. It involves exploration and approaches based on the review of recent advances. The adaptation of the architectural framework and the development of software system architecture, in order to build the system prototype, increases regulatory compliance mandates which are forcing enterprises to seek new approaches to managing reference data. Sometimes, the approach of tracking reference data in spreadsheets and doing manual reconciliation is both time-consuming and prone to human error. As organizations merge and businesses evolve, reference data must be continually mapped and merged as applications are linked and integrated. Accuracy and consistency, realize improved data quality and together let the organizations adapt reference data as the business evolves.

The challenge is to know how to keep the attributes of the data and how to preserve originality. Heritage is the concrete data, it gives interconnection with other aspects of reality.

(Extracted parts from Carrion Gordon (2014) and Gordon et al. (2015))

Terms and Definitions

Born Digital Defined as digital materials that are not intended to designate an analogue equivalent even the origin is form analogue source. For instance from digital source to have been printed to paper, physical way. Adapted from Commission et al. (2002)

DHP Access It addresses the usability of digital source retaining the qualities such as accuracy, functionality and authenticity primordial to generate digital material. Adapted from Commission et al. (2002)

DHP Authentication It establishes the authenticity of digital materials in a specific time. For instance, digital signatures. Adapted from Commission et al. (2002)

DHP Authenticity It refers to the trustworthiness of the digital artifact as a record. It includes confidentiality when the authentication of digital materials is necessary. Adapted from Commission et al. (2002)

DHP Electronic *Records* Defined as digital items. It could include for instance, emails, word processing documents, databases, or web pages. Adapted from Commission et al. (2002)

DHP Emulation A means of overcoming technological obsolescence of hardware and software by developing techniques for imitating obsolete systems.

DHP Heritage The concept is valued by the community because it provides links to the past and contributes to our sense of place and time. Adapted from Commission et al. (2002)

DHP Life-cycle Management The life-cycle management of digital resources is the need to manage the resource at each stage of its life-cycle giving relation between each stage and start preservation activities in practicable manner. Adapted from Commission et al. (2002)

DHP Metadata Model It contains information that describes significant aspects of a resource. The metadata is required to manage and preserve electronic items through the time over to ensure context and technical information.

Adapted from Commission et al. (2002)

DHP Migration To refresh of the media storage it is possible to make an exact digital copy or replicate original features and appearance and still maintain the compatibility of the resource with the new generation of technology. Adapted from Commission et al. (2002)

DHP Reformatting Copying information content from one storage medium to a different storage medium (media reformatting) or converting from one file format to a different file format (file re-formatting). Adapted from Commission et al. (2002)

DHP Refreshing Copying information content from one storage media to the same storage media. Adapted from Commission et al. (2002)

Digital Archiving The process of backup and ongoing maintenance as opposed to strategies for long-term digital preservation. Adapted from Commission et al. (2002)

Digital Materials Created as a result of converting analogue materials to digital form (digitisation), and 'born digital' for which there has never been and is never intended to be an analogue equivalent, and digital records. Adapted from Commission et al. (2002)

Digital Preservation Refers to the series of managed activities necessary to ensure continued access to digital materials for as long as necessary. Long-term preservation denotes continued access to digital materials, or at least to the information contained in them, indefinitely. Medium-term preservation defines continued access to digital materials beyond changes in technology for a defined period but not indefinitely. Short-term preservation describes access to digital materials either for a defined period while use is predicted but which does not extend beyond the foreseeable future and/or until it becomes inaccessible because of changes in technology. Adapted from Commission et al. (2002)

Digital Publications Refers to *Born Digital* objects which have been released for public access and either made available or distributed free of charge or for a fee. They may consist of networked publications, available over a communications network or physical format publications which are distributed on formats such as floppy or optical disks. They may also be either static or dynamic. Adapted from Commission et al. (2002)

Digitisation The process of creating digital files by scanning or otherwise converting analogue materials. The resulting digital copy, or digital surrogate, would then be classed as digital material and then subject to the same broad challenges involved in preserving access to it, as ’born digital’ materials. Adapted from Commission et al. (2002)

Digital Documentation The information provided by a creator and the repository, which provides enough information to establish provenance, history and context and to enable its use by others. To be comprehensive enough to enable others to explore the resource fully, and detailed enough to allow someone who has not been involved in the data creation process to understand the data collection and the process by which it was created.’ Adapted from Commission et al. (2002)

Digital Heritage It ia considered as a natural process. Heritage will properly analyse the context of the Data Set. The extended dimension of the value is relevant in this study. Most of the time Heritage has importance in traditional domains. It will carry the abstracts and the essence of the concepts. The concepts in Information Technology relate to Historical Values. It is important to carry on the aspects related with Inheritance and Heritage in the artifacts. The dimension for future generations of using data in context, presents opportunities for education and research. Adapted from Commission et al. (2002)

The Key words are based on (Digital Preservation Coalition, 2012) Verheul (2006) Adapted from Commission et al. (2002)

Contents

List of Figures

List of Tables

Part I.

Investigation of Digital Data Heritage Preservation (DHP)

1. General Introduction

The Digital World considers horizontal topics that cross many implicated sectors, which deal with the economic growth model. Many private and public industries are struggling from shortage of global and national regulatory requirements, Digital Transformation standards and policies. Digital Transformation findings address how to manage Digital Technologies in industries and enterprises. Some Industries have advanced digitalisation initiatives, which nowadays, are revolutionising one industry after another. In addition, organisations are facing continuously increasing competitive pressures due to new business models and rapidly changing customer needs. At the same time, they are battling frequent increases in costs that occur because of more and more legal requirements. In essence, companies are required to adapt to the pace of these shifts - organisationally, procedurally and technologically.

Throughout this research, although various models are proposed, the user-centric concept remains key.

Concrete data and how to keep heritage has led us to consider the component of each organisation in context of the available technology. The questions of how heritage is created is fundamental. Heritage is created based on social aspects. A need to move away from the traditional understanding of Heritage reflects the real meaning of the data. Artifacts and the tendency is to have less physical representation in the World of Logic. The representation of items refers to the tendency of more things non-physical and refers to how heritage over time passes on certain attributes. The challenge is to preserve the design and its origins. The principal idea of this research is to be able to classify a mass scale metamodel and framework. The integration of commercialisation and research is one of the topics that drives this investigation.

The following questions arise: How it can be applied in the improvement of technological field? Will it have relevance for the community?

Who will be the benefactors? Are there any alternatives to do that? Furthermore, these considerations are very applicable. At this point, the approaches like concepts and case studies give us a wide appreciation of the environment and show how can be the managing and handling of the difficulties. The tendencies and methodologies of commercialization are more relevant than 'recipes'. However, the realistic point of view of business and marketing provides the probability of a specific behavior happening in the marketplace. Undoubtedly, being aware of the different reactions and contextualization of the diverse fields, should teach us the principles of commercialisation.

Any effective method of Data Management in covering the primary concepts is through Digital Data and Historical Analytics. The contextualisation and data relationships are key concepts around DHP.

Data Heritage Preservation can be defined as the perfect combination the four stages below:

1. Problem
2. Situation
3. Context
4. Solution

The following main definitions are combined with a Standardized Mapping Model, as a common practice to create, analyse and develop the architecture domain in the IoT (Internet of Things) community.

A breakdown of the term DHP is the following:

- Digital Data is discrete
- Heritage: the legacy of physical artifacts and intangible attributes of a group or society: man-made heritage.
- Preservation: the prolongation of the life of a record and relevant metadata which enhances its value, or improves access

Around these definitions, the following questions need to be answered:

How can the model generate metadata or metadata models? How is the data presented? How can one represent the rules? How can facilitate the retrieval of information as well as find and allow the connection of the data?.

In relation to concepts about Data Aging, our proposal is to generate New Relations and New Mechanisms.

The main concern around this challenge are how to pack information in repositories, how metadata is used in Industry and how the client can look for the information as well as the issues around Research.

The identification of a Metadata model and how to implement the standardised model shows us the best way to define Metadata. One of the important issues is to define the navigation tools.

It is important to check the techniques and review the concepts of Metadata of the model to easily find data and the relationships between heritage and preservation.

There is a new perspective of Digital Artifacts. The new definitions specify heritage elements and how to preserve heritage in the mainstream. These include the origin and the meaning. These concepts give us a new perspective of the definition of the ontology of Metadata, Epistemology and how to relate with Heuristics. The process of the prediction of data is important. However, there are concerns about mapping. The exploration of the Metadata Model for Dynamic Representation and the definitions of data needs to be decided in order to know which information is relevant. Predictions around data generation and classification of the impact of small changes and how to rearrange data are statements to be considered for the model representation. Digital Heritage guides us to monitor and modeling. The considerations around heritage including Heritage Aspects such as data classification and Heritage Elements, which show value to the user.

The goal is to create a Predictive Model where algorithms organise the metamodels.

1.1. Background

The rising awareness of the challenges of preserving information over the long term has led to a wealth of initiatives developing economic models, methods, tools, systems, guidelines and standards for digital preservation. The challenge of digital preservation is to assure that information, which nowadays is coded and stored in digital formats, can be read and be used in an unforeseen future. This is an interdisciplinary problem combining organizational and technical challenges.

1.1.1. Research Motivation

The research motivation addresses Serendipitous Heritage and proper analysis in the context of datasets (DS) which extend Dimension of value. The importance of Heritage in traditional domains will read to the abstraction of the essence of the term.

This concept in Information Technology (IT) is defined in relation to Historical items. In this way, heritage is the essence of the artifacts and expresses a deep definition of data.

The goal is for future generations to analyse the data in context of their environment. In order to design a practical system that manages massive amounts of critical data to keep in a long-term, one needs the usage of media storage, conversion, migration and management strategies.

To define the nature of data and its challenges it is necessary to apply for a Business Context of Data preservation of Digital Artifacts. For the motivation for Data Heritage Preservation, the definitions are diverse and depends on the domain.

It defines the motivations behind this research in terms of its significance in the world of Digital Preservation, the capability of using both a Research Problem and Motivation to create new types of preserve of Heritage in the World.

The main goal is to define a dynamic solution for the use of Serendipitous Heritage, based on Software Architecture and to integrate the definition of Heritage for Data Preservation.

1.1.2. Research Aims and Objectives

The objectives of determine the framework, methodology and patterns for Preservation and Applications of Digital Preservation applied in Big Data architecture.

The three objectives are the following:

O1: To design a process model which can analyse the whole steps of digitalisation in order to maintain the accessibility and contextualization of data.

O2: To increase the prevalence of preservation in the transactions of government entities and business activities through digital resources and metadata information.

O3: To improve the results of digitalization with lifecycle management and to propose an enterprise software architecture for the standardization of data.

The aims can be discussed as follows:

A1: To understand the process and the challenges of changing from Physical Information to Digital Age, related to “Born Digital” information.

A2: To assume social responsibility related to improvements in the technology and electronic records in documentation.

A3: To prove and achieve an efficient architecture for Heritage Preservation through the management of Digital Archiving, Materials, Records and Publications.

1.1.3. Research Scope

The definition of the Research Scope is strongly shaped by the stakeholders who are influenced by this research.

According to the concept of the Digital Preservation of the development of technologies as well as the diffusion of Digital Transformation, the applications of this research have diversity of fields and case studies.

The importance to focus on stakeholders is reflected in the development of objectives. The key stakeholders of Digital transformation, shown in Figure 1.1, are the following:

1. Government Entities
2. Business Enterprises
3. Socio-Technical Cultural Domain
4. Digital Society

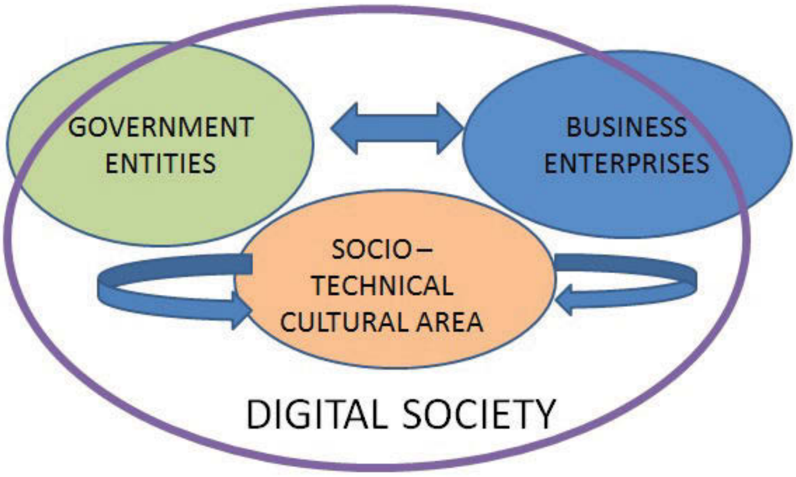


Figure 1.1.: Stakeholders Digital Transformation

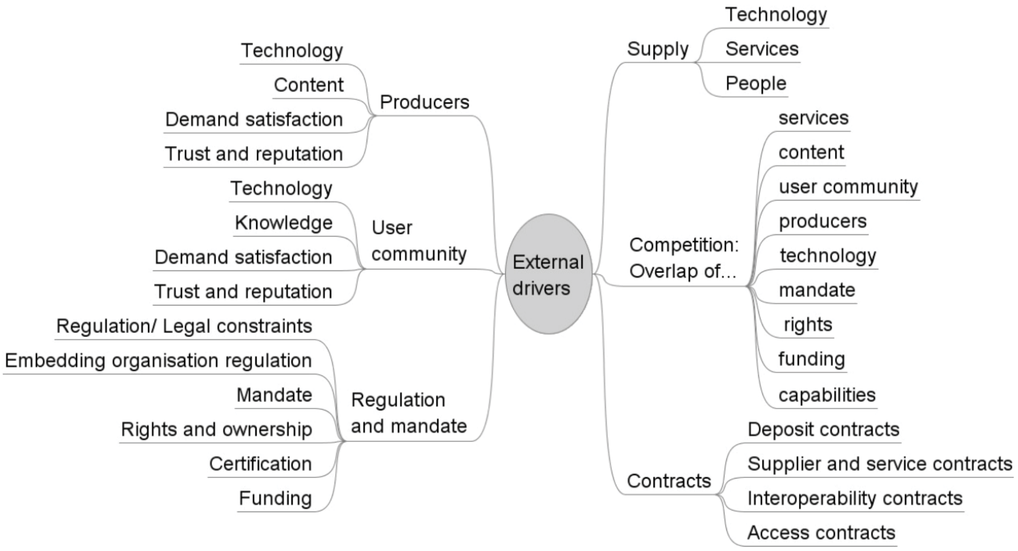


Figure 1.2.: Main external drivers. Adapted from (Becker, 2011)

Stakeholders: IoT Community

The idea of an IoT Society begins in essence with the general definition of the stakeholders. Nowadays, people and users are more connected than ever. The Internet is the common connector between technology and human beings. However, due to the heterogeneous nature of the IoT devices, data is distributed, aggregated, and processed which in turn present challenges to digital forensics investigations in many ways. The need to implement new techniques is required to assume these challenges. To realise the approaches proposed in this research, an orientation to the implementation, deployment, analysis and evaluation of the IoT is important.

1.1.4. Research Questions

The research questions are include lookup at how to achieve Business Process Reengineering for data preservation by a virtual framework as well as how to develop a digital oriented architecture focused on solutions to ensure the mechanism of preservation. They require a study of the use of the mechanism of Metadata for improving the quality of storage repository and addressing whether we can incorporate them in Systems Engineering.

All we can embedded while lookup at the engineering practice at considering the correlation between Best Practices and Engineering standards with Digital Heritage.

Based on these considerations the final Research Questions come down to the following three:

RQ1: How the process of Digital Data preservation can be improved?

RQ2: How DHP elements can benefit to Heritage Context?

RQ3: How the quality of preservation and access to critical data can be enhanced?

1.2. Significance and Justification

Significance

Digital preservation as an innovative idea tries to develop a standard framework that keeps generations of data for the long-term. The tendencies related with Digital Archiving, Digital Materials, Digital Preservation, Digital Publications and Digital Records gives us the prospect of developing an architectural solution. Based on the challenge of the Digital Era, concepts of accessibility and contextualization of data are crucial.

Through the development of procedures and processes, modelling embedded systems, proposed an architectural solution for digital preservation to achieve an accurate designing.

The next generations could access and understand Historical Heritage. Business enterprises could improve their activities through digitalization and include Lifecycle Management in the Supply Chain as the power of knowledge is based in the content, meaning and interpretation of data will be solved with Digital Preservation, which will be based on techniques and an architectural proposal. The relationship between Software and Data Management is the basis for best practices in Digitalisation. The reformatting of the information is directly related to the standardisation of data and with Digital Preservation and Heritage Patterns. The Digital Society, as it is called, has developed a relationship with technological matters over time. The challenge of this research is to manage information, without including the experience of any specific field. The generalisation and the expansion of this model and framework is a principal element of the development of efficient Enterprise Architecture.

The significance or the relevance in the research process is to design a practical system that can manage massive amounts of critical data or information. The storage media, migration, conversion, and overall management strategies are needed to define the appropriate adaptation of DHP model.

Justification

The principal justification for this research is the management of real data. In order to do this, we need to know the issues and the motivation of the generation of the information, which we are dealing.

The justification behind this research, in terms of its significance, is the capability of using a Research Problem and Motivation to create new types of preserved Heritage in the world of Digital Preservation.

The increase in digital content being collected and created by and for IoT users has resulted in associated technological and organisational challenges. The users will respond to these challenges through the development of new strategies to ensure the ongoing preservation and access of its digital content.

1.3. Hypothesis

Our Hypothesis is that:

1.3 Hypothesis

The Data Heritage Preservation model based solution is able to improve Dependability, Manageability and Usability of information in Digital Transformation processes.

This research focus on the application of a DHP model in Digital Transformation Processes can improve dependability, manageability, scalability and performance of the Big Data and IoT System Architecture.

Manageability refers to Data Collection, Data Storage and Data Updating. It requires the data to be implementable, controllable, trackable and contrivable. In this research, it is intended to validate the dependability using qualitative methods. However, manageability and usability will be validated and tested using quantitative approaches and action studies. This enforces the trustworthiness of the framework that ensures the information is consistent and repeatable.

Terms such as ’data’, ’information’, ’knowledge’ and ’knowledge management’ are frequently confused. ’Data is a set of discrete, objective facts’. ’Information is data that has been organized placed in context with meaning’. ’Knowledge is information combined with experience, context for actions and decisions’ Tuomi (1999).

A new point of view based on a standard architectural approach is related to Data Management and the principal tendency is to have results focused on: Application, Platform, Media and Long-term access.

Important knowledge is based on the relationship between situation and context. One of the best ways to validate the concepts is to demonstrate the benefits for the real stakeholders.

Preservation of information relates to the growth in the massive amount of data, driving the society, to solve Big Data issues and information loss. The ability to keep data long-term and open date, are the principal concerns we need to focus on management of information.

It is necessary to demonstrate, validate and verify the benefits of the model. On the other hand, these concepts are important in the process of the case studies.

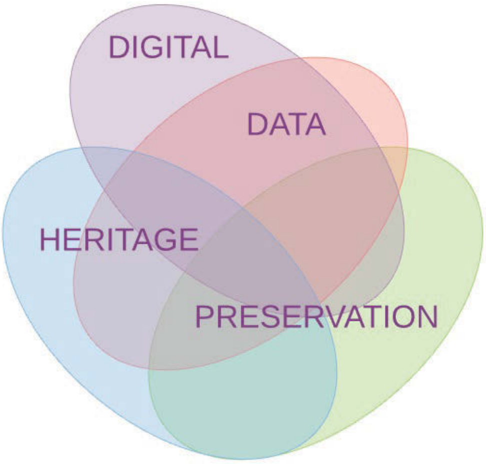
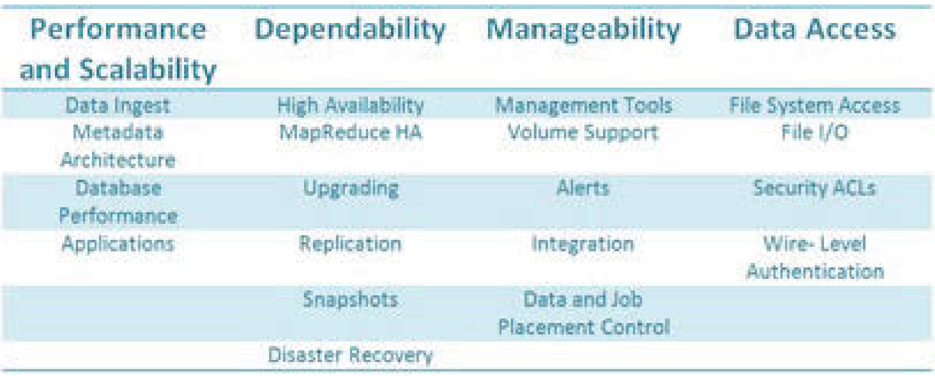


Figure 1.3.: Digital Data Heritage Preservation (DHP)

1.4. Problem Statement, Challenges and Validation

The challenges around Data Heritage Preservation are to define the correct model and the accurate way of validation. For the validation process, the model designing is based on a known tool for format identification called DROID (Digital Record Object Identification). It will be explained on detailed on chapter 5.

Table 1.1.: Quality Attributes



Performance and Scalability	Dependability	Manageability	Data Access
Data Ingest	High Availability	Management Tools	File System Access
Metadata Architecture	MapReduce HA	Volume Support	File I/O
Database Performance	Upgrading	Alerts	Security ACLs
Applications	Replication	Integration	Wire- Level Authentication
	Snapshots	Data and Job Placement Control	
	Disaster Recovery		

1.5. Contributions

The following is the list of publications we found as a result of our investigations. They are divided by Case Studies or Areas of study. The general research is in Preservation, this is divided in specific topics.

Preservation of Medical Records

The relation between the case studies and the preservation is essential. Through the following publications the initiatives about medical records, for instance, Cardiac Valve Control, Bio-informatics and Medical Big Data, were developed showing the importance of the Boga Data Management through the appropriate implementation of DHP Model.

[J1] Carrion L., Sanchez J. "Architectural Framework to Preserve Information of Cardiac Valve Control"

[B2] Carrion L. & Chaczko Z.," Bio-informatics with Genetic Steganography Technique"

[B4] Z. Chaczko, L. Carrion-Gordon, W. Bożejko, "The Metamodel of Heritage Preservation for Medical Big Data"

Preservation in General Terms

Based on the following publications, the objective about planning, strategy, ontology, definition of patterns and the relationship with Serendipity. As a Big Data research, DHP modelling and designing addresses many terms. These are examples of the principles definitions that have relevance in the correct development of the framework.

[C1] Carrion L. & Chaczko Z.,"Digital preservation strategy: Planning procedure to preserve Critical Information of Heritage"

[C9] Chaczko Z., Carrion L., "Standardized Mapping Model for Heritage Preservation and Serendipity in Cloud"

[C10] Chaczko Z., Carrion L. "Towards Digital Heritage Preservation Framework for Situation and Context Aware for Information Management"

[C11] Carrion L., Chaczko Z., Wazirali R., "Ontological Metamodel for Consistency of Digital Heritage Preservation (DHP)"

[B1] Carrion L. & Chaczko Z.," Digital Patterns for Heritage and Data Preservation Standards"

[B3] Carrion L., Chaczko Z., Resconi G., "Serendipity in Context of Digital Preservation and Artifacts in Large Infrastructure Oriented Systems"

Preservation and Development

Under this classification, these publications are the most external, in order to mention other topics, but not relevant to the study. For example, the study of Petroleum Residues through statistical approach, Resource Management and XUnit.net, seem to be outside of the scope, but the outcome is an interesting outcome for the study. These publications help us for the clarification of the domain of study.

[J3] Flores M., Horna L., Escobar A. and Carrión L., "Estimation of the Contaminant Risk Level of Petroleum Residues Applying FDA Techniques"

[C3] Carrion L., Chaczko Z., Alenazy W & Mu M1 "Development of an Expert System to assist in Resource Management"

[C4] Carrion L., Chaczko Z., Braun R., Dagher J. " Design of Unit Testing using xUnit.net"

Preservation and Cultural Approach

Although one of the aims in the research is to differentiate from the cultural domain, this publication is an interesting approach to 'La Bomba' Ecuadorian dance and expresses, how the technology can keep the roots of it. The Preservation is inherent to this publication for the appropriate expression of the cultural manners in the explanation of the dance background.

[J2] Carrion L., Lopez M., "Preservation Model to Process 'La Bomba Del Chota' as a Living Cultural Heritage"

Preservation and Techniques: Steganography and Smart Classroom

Education domain and topics related to Steganography and Smart Classroom are the correct connection with different case studies detailed in the experimentation chapter. These are complex concepts that give sense to the development of preservation solutions. These publications contribute to spread the concept of preservation, but better, the approach for the DHP modelling and design.

[C2] Wazirali R, Carrion L. & Chaczko Z.,“ MultiIlayers DNA - QR Based Steganography”

[C5] Carrion L., Chaczko Z., Alenazy W, & Tran A.” Augmented Reality Based Monitoring of the Remote-Lab”,

[C6] Wazirali R., Chaczko Z., Carrion L. and Slehat S., “Review of Quality Metrics Assessment for Steganography”

[C7] Alenazy W., Chaczko Z., Carrion L, and Chan C.Y., “Haptic Middleware Based Software Architecture for Smart Learning”

[B5] Z. Chaczko, R. Wazirali, L. Carrion-Gordon, W. Bożejko, “Steganographic Data Heritage Preservation Using Sharing Images App”

[B6] Z. Chaczko, R. Klempous, L. Carrion-Gordon, “Enabling Design of Biomimetic Middleware for Large Scale IOT-Based CyberMedical Systems”

1.6. Thesis Structure

Throughout the text various models are proposed but the user-centric concept remains key.

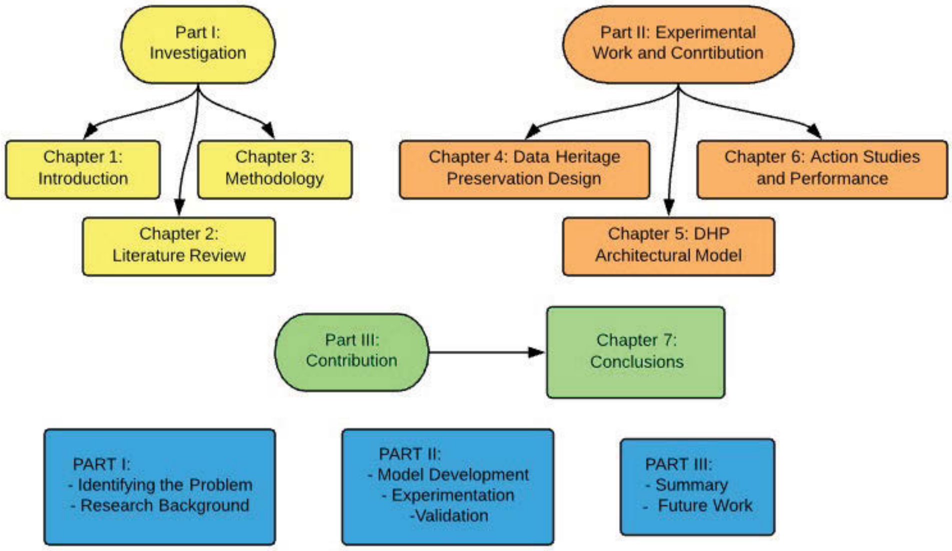

Part I: Investigation
Chapter 1: Introduction
Chapter 3: Methodology
Chapter 2: Literature Review
Part II: Experimental Work and Conrtibution
Chapter 4: Data Heritage Preservation Design
Chapter 6: Action Studies and Performance
Chapter 5: DHP Architectural Model
Part III: Contribution
Chapter 7: Conclusions
PART I: - Identifying the Problem - Research Background
PART II: - Model Development - Experimentation -Validation
PART III: - Summary - Future Work

Figure 1.4.: Organisation of the Thesis

The thesis is organised in the three main parts (Figure 1.3) as follows:

1. Investigation
2. Experimentation
3. Contribution

The Investigation part consists of the following chapters:

- General Introduction
- Literture Review: SOTA of Digital Transformation
- Methodological Perspective

The Experimental Work is covered in the following chapters:

- *Data Heritage Preservation Design:* DHP Modelling and Design: Explaination of a Proposed model in terms of Designing and Processing.
- *Architectural Model:* DHP Implementation and Simulation: Inspired by DROID as it explains the implementation of the software.
- *Action Studies and Performance:* DHP Vision and Benchmarking Approach: Comparison between case studies and different solutions.

Explained in the Figure 1.5 with the following details.

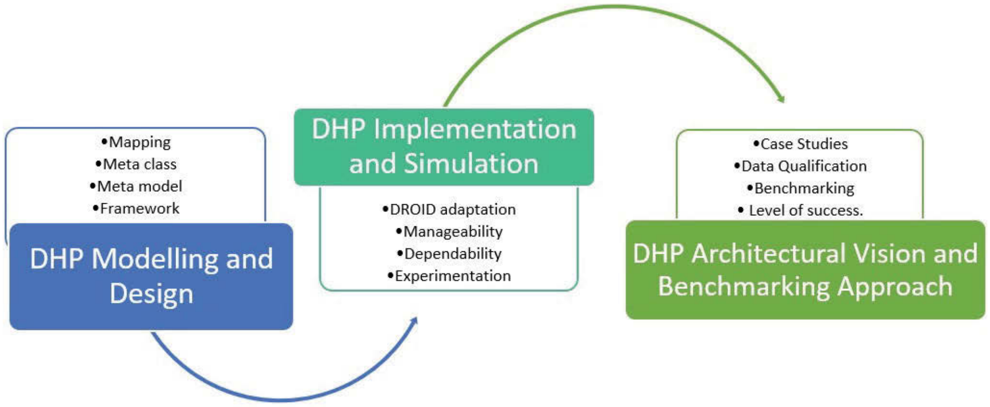


Figure 1.5.: Process for DHP Research Contribution

The last part with the Contribution, it is integrated with the conclusions which shows the current achievements by the implementation of the model and provides conclusions related to Digital Preservation.

2. SOTA of Digital Transformation

A most common definition for the concept of Digital Transformation provided by Berman (2012*a*) as: 'Digital transformation is the application of digital technologies to fundamentally impact all aspects of business and society'. One of the most important requirements for (DT) development is to ensure that both the data reliability and data authenticity meet the needs of a diverse group of data custodians. An additional requirement is to provide a sufficient level of data security to prevent any possible loss or corruption of important digital artefacts. Data security may include data encryption/decryption mechanisms and other techniques such as steganography or watermarking. Knowledge management and ontology models of Digital Transformation play important roles in the process. A good understanding of various ontology components, their relations and interactions, help us to analyse and assess the tendencies of change. Modelling of the architecture and its representation(s) is a useful tool to develop and analyse the best prototypes. Nowadays, some of the most advanced ICT technology solutions mentioned by Zysman et al. (2011) can be met by the requirements for sufficient and effective data storage. However, a specific need for preservation of the data authenticity and originality during Digital Transformation requires us to consider an efficient computing environment (i.e., methodologies, software frameworks, and design patterns, applications of digital preservation), where the original digital artefacts and related knowledge are not lost or corrupted during the process of change. The real challenge of the Digital Transformation Era is how intelligent and effective data preservation can be implemented and maintained.

Digital Transformation, which takes place in government entities such as the *National Library of Australia* (NLA), offers a useful case study to evaluate the effectiveness of Digital Preservation and archiving mechanisms. The study and experience gained when analysing NLA mechanisms, motivated this research and allowed for the identification of areas of possible improvement to the process model of data preservation and archiving.

Knowledge Preservation can be applied in diverse fields and this research study focuses on the adaptation of standards, models and patterns that could ensure the design of scalable, adaptive and survivable DT applications.

It is hoped that the findings of this research can create future opportunities for collaboration with both public and private institutions that undergo Digital Transformation phases. There is a growing narrative around uncertainly and risk associated with a rapid pace of digitalisation. Some of the thinking around DT in large organisations gives a sense of orientation for prime objectives and priorities for the building of effective data models and the data preservation framework(s). The benefits of this research should serve the customers and the community as a whole. The problems covered in this research address both global and local organisations issues and challenges.

Information preservation is one of the most important issues in human history, culture, and economics, as well as the development of our civilization. While earliest information was recorded in carvings on stone, ceramic, bamboo, or wood, the development of civilization paved the way for new storage media and techniques for recording information, such as writing on silk or printing on paper. Eventually we were able to put photographic images on film and music on records. A revolutionary change occurred in the information storage field with the invention of electronic storage media. With the advent of high-performance computing and high-speed networks, the use of digital technologies is increasing rapidly. Digital technologies enable information to be created, manipulated, disseminated, located, and stored with increasing ease. Ensuring long-term access to the digitally stored information poses a significant challenge and is increasingly recognized as an important part of digital data management.

2.1. The Era of Digital Transformation

Hedges et al. (2007*a*) addresses the statement, 'Strategy, not technology, drives digital transformation' Kane et al. (2015). Digital Transformation is the tendency in terms of Technology. The research questions are related to the integrity and authenticity of the information we manage. The completeness of these definitions makes it relevant.

Digital Preservation has relevance according to ontology and knowledge management Nasir & Noor (2010).

The management of heritage inside the cultural method, with the integration of Knowledge Management and Ontology, integrates a complete methodology for better comprehension. The relationship with the field of culture relates directly with the terms of preservation.

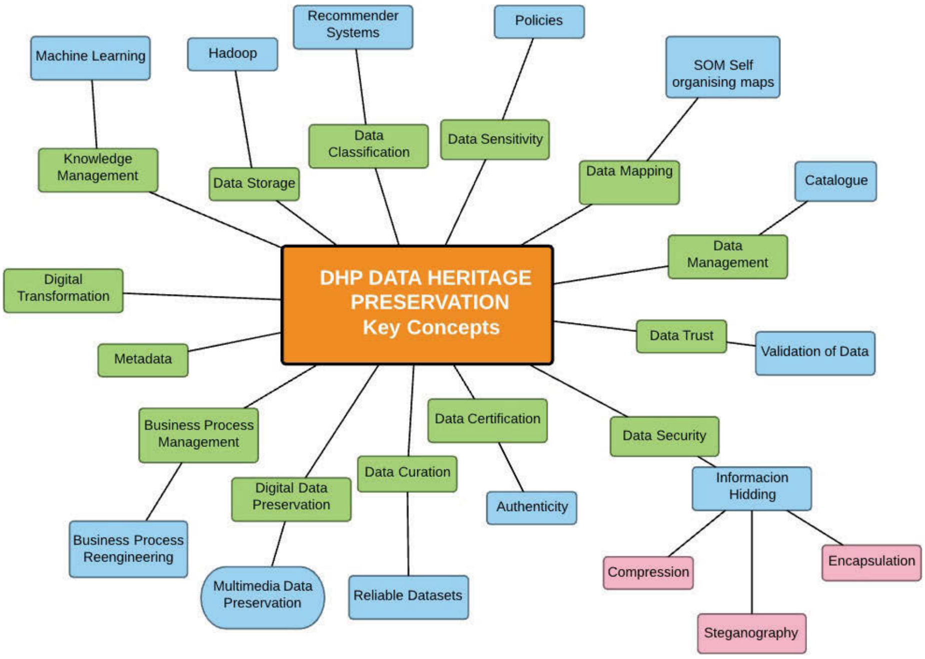


Figure 2.1.: DHP Key Concepts

As Figure 2.1 shows there are many ideas around DT concepts to analyse in SOTA. According to Thibodeau (2002) there are many important ideas projected in the publication mentioned about the integration and relation between key concepts.

According to Prakash et al. (2012) for 'Privacy-preserving data mining (PPDM)' clustering is defined as an alternative of preservation. The contextualization is much related to the customers or the users of the new framework. The classification of data as integrating Digital Communities to have access to the preserved information is stated by Becker et al. (2011) that in a digital information age one must understand the mission to ensure a knowledgeable society in the long-term.

Rogers (2015) addresses that digital records consist of data (content) user-generated and metadata generated by the source and location, focus on management of performance of the record for instance the native file format, describing the data and metadata is generated by the user.

Zysman et al. (2011), Patel & McCarthy (2000) explain that Digital Transformation as a Service has relationship with leadership and e-businesses. This field is one of the basis for the cases studies in the research.

2.2. Conceptualisation

The DHP approach needs clarification of key concepts that are involved. The principal terms to define are Metadata, Data Heritage, Preservation, Serendipity and Big Data Concepts. Not only are the concepts important, their context is also an important requirement. These terms could be conceived as general but the idea is to contextualize them based on the field. This means keeping a balance between a general and a specific approach. Harvey (2005) observes that many terms have been used interchangeably to describe the characteristics/ elements of digital objects that must be retained for long-term preservation. The following Figure 2.2 explain on detail the principal Concepts.

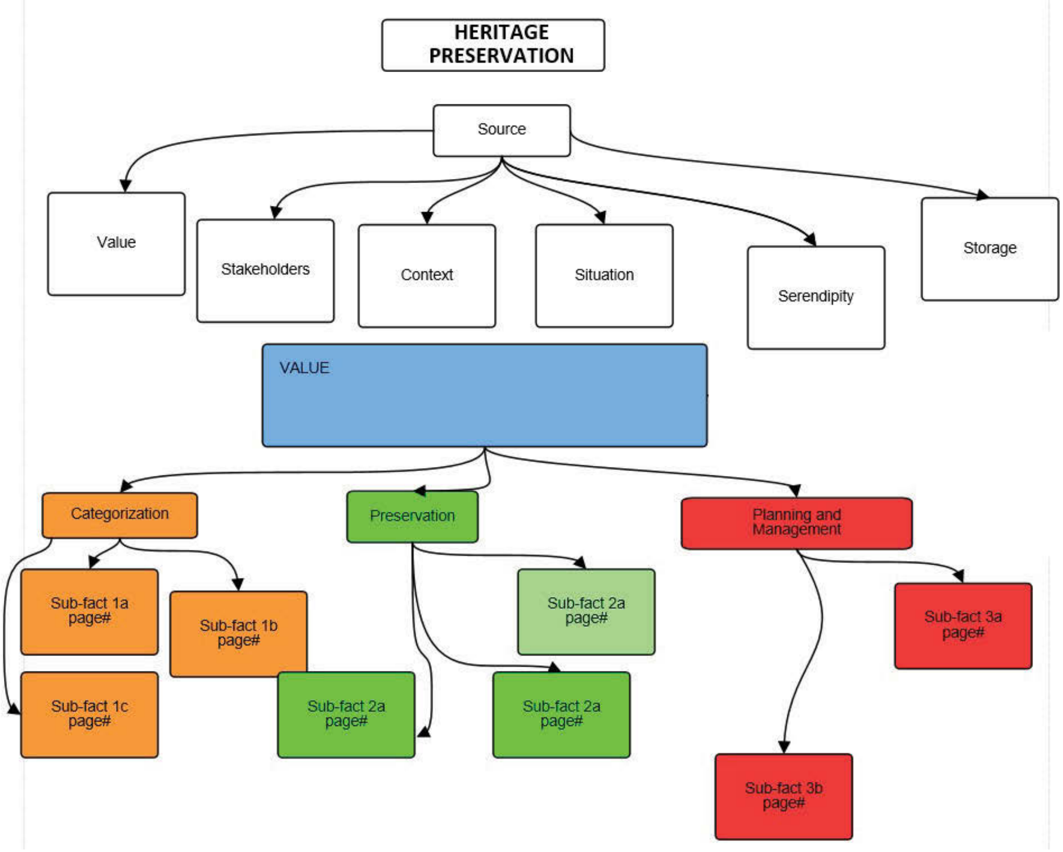


Figure 2.2.: DHP Orientation Concepts

2.2.1. Metadata Definition

In recent years, the concept of descriptive metadata for electronic resources has received much attention. According to Chang et al. (2001) the metadata is defined under the following characteristics: Definition of metadata preservation in the items, Requirements at High-level, A reference Model (OAIS) Open Archival Information System, A review of existing preservation approaches, the identification of convergence/divergence. The multidimensional connections between the SOM Maps clarify Metadata definition. The principal characteristics are Proximity, Dates, Execution, Content, Authorship, and Context. The principal objective of Metadata is to keep the important features of the information. It gives weight of accurate preservation. An effective mechanism is to identify if metadata changes, to the SOM algorithm, Self-Organising code.

2.2.2. Data Heritage

Becker et al. (2011) suggested to extend the metadata by using the concept of the stakeholder as this would allow for analysis of user concerns and help to define the metadata model drivers and constraints.

The Digital Data Heritage concept should not be considered as an automatic mechanism. It is related to syntax, context, meaning and behavior. This is a misunderstood concept of the heritage. In popular perception, heritage is always about passing the assets of one generation to a future generation. In the context of digital data, heritage should not be seen as a one-dimensional and isolated element, but rather as a linked asset with multiple relations (dimensions).

Digital data needs to be considered in context of its connections and relationships as these give the true meaning to the information. If the heritage of the digital data is not adequately preserved for the future, vital components of the information can often be lost forever. At times, collected data may have serendipituos aspects, without any specific purpose, immediate use or pre-defined format and yet, there might still be a certain value to keep such data. In the context of Big Data it may aggravate issues with the handling of large volumes of data. Thus, one has to be very careful defining the seredipitous elements of digital data.

Considerations should be given to a proper justification for maintaining serendipitous data. When defining the Metadata record, one of the main questions to ask is, how one can incorporate the contextual elements into the Metadata model. Keeping the Digital Heritage data can pose a dilemma in the context of Big Data . Hence, the second key question is, should one keep the digital data with or without Digital Heritage?. There are some risks associated with maintaining the additional data components. These are because the more data that is collected and stored, the more data that could be potentially lost. The trade-offs of analysis involving the maintenance costs of digital heritage data and the value of a new knowledge that can be created in the future need to be considered. There are many other questions that to be asked about the need for keeping the digital heritage data. For example, the new value created in the future and the digital data change in time should add value to heritage records. It is important to define the cognitive, contextual and physical dimensions for the digital data to create value in the future? Heritage, as a general term, has many definitions, most of the time oriented to cultural influence. Kaplan (2015*a*) stated that it is considered 'helpful' in the organisation of several dimensions in the context of 'Big Data Digital Humanities research'. The rationale for this is to define the nature of and limits for Digital Society. Heritage generates much discussion and complex issues to consider. Heritage as a concept is related with a deep transformation in the Digital Society. These findings are fundamental concepts for main consideration Kaplan (2015*a*). The process of data collection and the interpretation of data must take place within the context of Digital Culture. Figure 2.3 shows one example for digitalisation of the physical cultural item. The technological approach proposes a real look and feel of the construction.

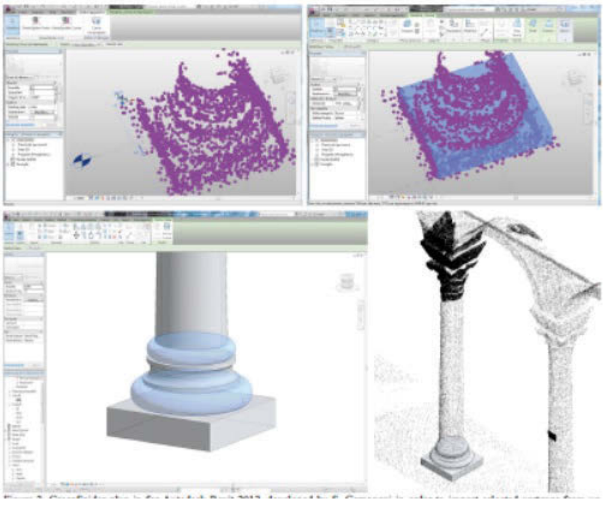


Figure 2.3.: Data Heritage (Garagnani, 2013)

2.2.3. Preservation

There are generally two approaches to long-term preservation of digital materials:

preserving the object in its original form as much as possible along with the accompanying systems, migration or transformation and transforming the object to make it compatible with more current systems but retaining the original 'look and feel'.

Migration is the most widely used method, but there can be changes to the original. If some of the original properties are lost, what then are the essential properties to maintaining its integrity? Currently there are no formal or objective ways to help stakeholders decide what the significant properties of the objects are.

Certain characteristics of digital objects must be preserved over time in order to ensure the continued accessibility, usability, and meaning of the objects, and their capacity to be accepted as evidence of what they purport to record.

An important goal of digital preservation is more than just retrieving the objects; it is to ensure the authenticity of the information. A digital object can change as long as the final output is what it is expected to be. The properties to preserve come from the purpose of the object, and at least one purpose for the object needs to be defined. Archivists have created standards that look at records in the context of their creation, intended use and preservation. It is important to ask what features of the object is important when delivering to the user. There may be many uses to many communities that were not intended by the object creator, so we should not let 'the ideal' limit 'the reasonable'.

The classification for Preservation is defined as Active and Passive preservation. It is considered according the nature of the generation of data.

Active preservation tries to ensure accessibility to records and items, continuously over time. The interaction with them has to be in active way in terms of storage and management, if there is movement of the item into a new storage environment. Both passive and active preservation require protection by the integrity of the original item as postulated by Gracy and Kahn (2012).

Passive preservation is considered as a process which keeps the integrity and accessibility of data in a continuous manner, so the digital items can generate metadata. Basically, passive preservation pretends to maintain the original item with less changes in the used technologies which are used for storage and processing (Gracy and Kahn 2012).

2.2.4. Serendipity Concepts

One manner how to keep the context as an important consideration is to define serendipity. It is a popular word that captures a rich phenomenon with potential implications starting from the personal through the global level. According to McCay-Peet (2014*a*) it is associated with revelations, discoveries, life events and innovations.

McCay-Peet (2014*a*) stresses the importance of serendipity:

When we fund basic research, we are funding serendipity. Even a sober, frugal, post-recession United States must invest in serendipity, because without it, there is no vitality in the innovation ecosystem. Indeed, there is no innovation (Jackson, 2012).

In the context of business and research, serendipity is often credited with contributing to innovation, the acceleration of business growth and the development of new research directions. Thus, in the big picture, it is important to understand serendipity, what influences it, and what may facilitate it. While the impact of serendipity has the potential to reverberate beyond the individual, we can also see the importance of serendipity at the micro level – the affective response serendipity evokes and the learning it sparks. Serendipity can bring simple pleasures and happiness to a person's everyday life. Thus, when we view serendipity in terms of its potential to positively influence a person's emotion or affect, we can see how serendipity that occurs within digital environments may have an impact on user engagement, 'a quality of user experience with technology'.

2.2.5. Internet of Things IOT

The connection between DHP and IoT has big Data Problem. It should be sourced for the future for capturing data DHP to ensure the originality of Data.

Information and Communications Technology (ICT) controls our daily behaviors. It becomes a main part of our life's critical infrastructure bringing interconnection between heterogeneous devices in different aspects. Personal computing, sensing, surveillance, smart homes, entertainment, transportation and video streaming are examples, to name a few. As a critical living entity, the internet pretend to develop the changing and evolving leading to emerging new technologies, applications, protocols and algorithms. Acceleration of wireless communication trends brings an ever-growing innovation in the internet connectivity and mobile broadband. Infrastructure less communication devices become ubiquitous, smart, powerful, connectible, smaller, cheaper, and easier to deploy and install. This opens a new future direction in the society of the ICT: the Internet of Things (IoT). Nowadays, the IoT, earlier defined as Machine-to-Machine (M2M) communications, has become a key concern of the ICT world and research communities. In this paper, we provide an overview study of the IoT paradigm, its concepts, principles and potential benefits. Specifically, we focus on the IoT major technologies, emerging protocols, and widespread applications. This overview can help those who start approaching the IoT world aiming to understand and participate in development (Abdul-Qawy et al, 2015).

The following Figure 2.4 shows the different categories of units in IoT environment. This projection states the accurate idea of the future implementations. In the Figures below Figure 2.4, Figure 2.5, Figure 2.6, Figure 2.8, and Figure 2.9 address more details about IoT. It states the Internet of Things (IoT)* units installed base by category from 2014 to 2020 (in billions).

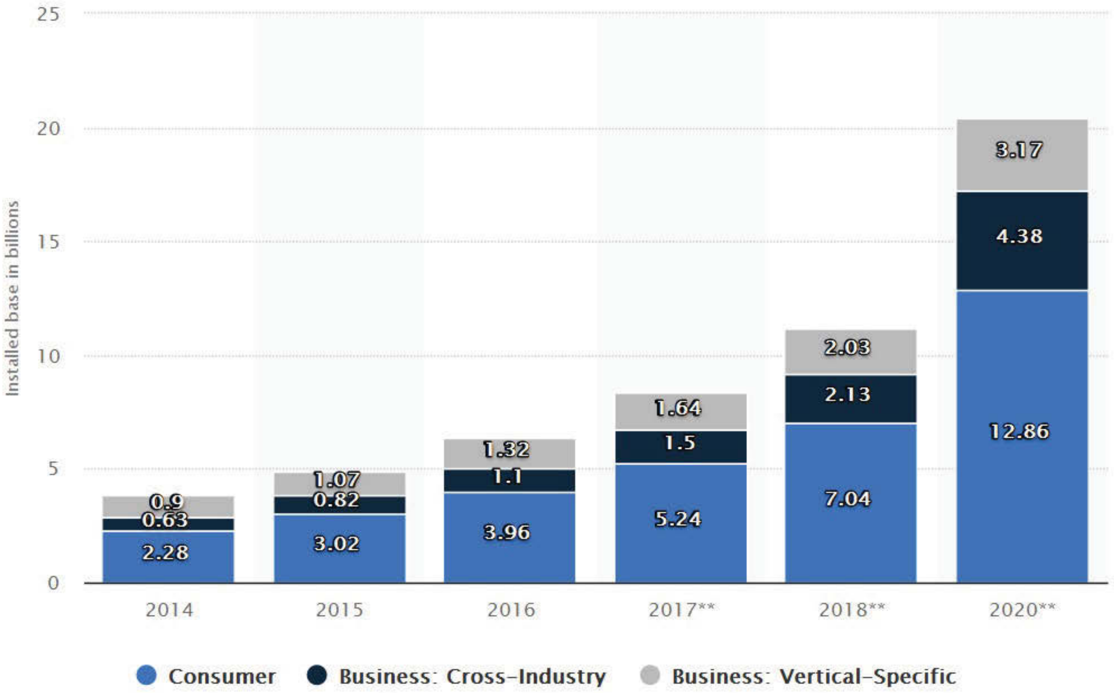


Figure 2.4.: (IoT) units installed from 2014 to 2020 (Statista, 2018)

In Figure 2.4 there is the explanation of accurate projection according to Gartner for the next years. There are five principal categories in this classification.

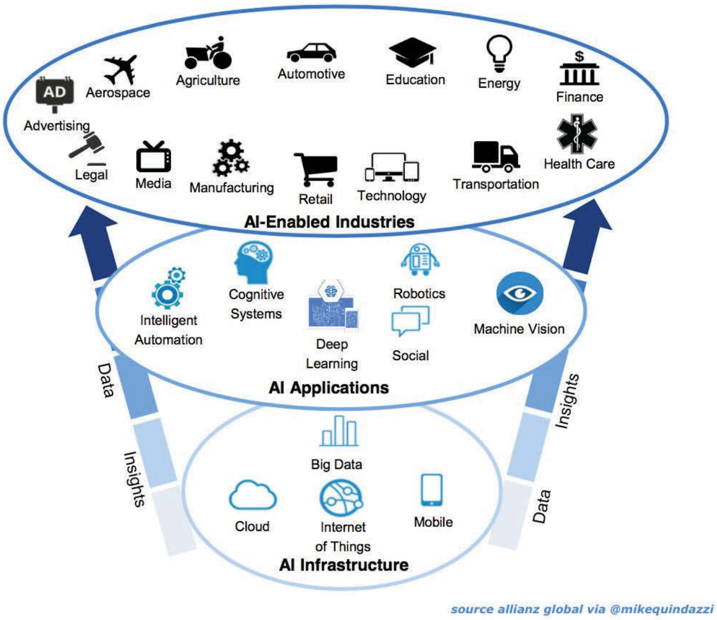
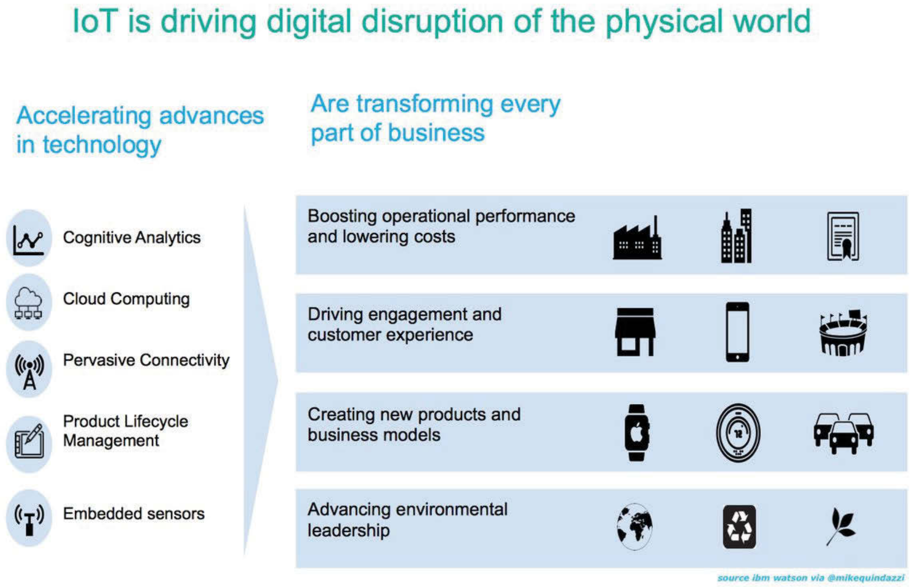


Figure 2.5.: Technological Levels (Quindazzi, 2017)


IoT is driving digital disruption of the physical world

Accelerating advances in technology

- Cognitive Analytics
- Cloud Computing
- Pervasive Connectivity
- Product Lifecycle Management
- Embedded sensors

Are transforming every part of business

- Boosting operational performance and lowering costs
- Driving engagement and customer experience
- Creating new products and business models
- Advancing environmental leadership

source ibm watson via @mikequindazzi

Figure 2.6.: Internet of Things next steps (Quindazzi, 2017)

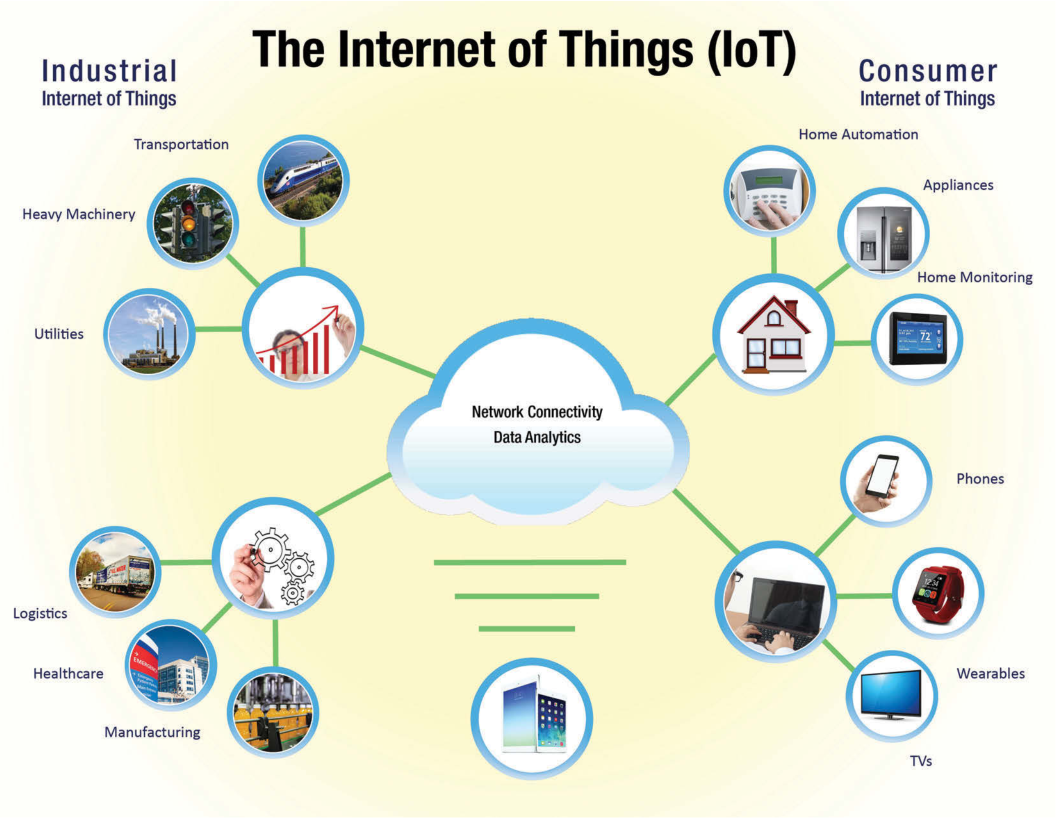


Figure 2.7.: IoT: Industrial and Consumer (Quindazzi, 2017)

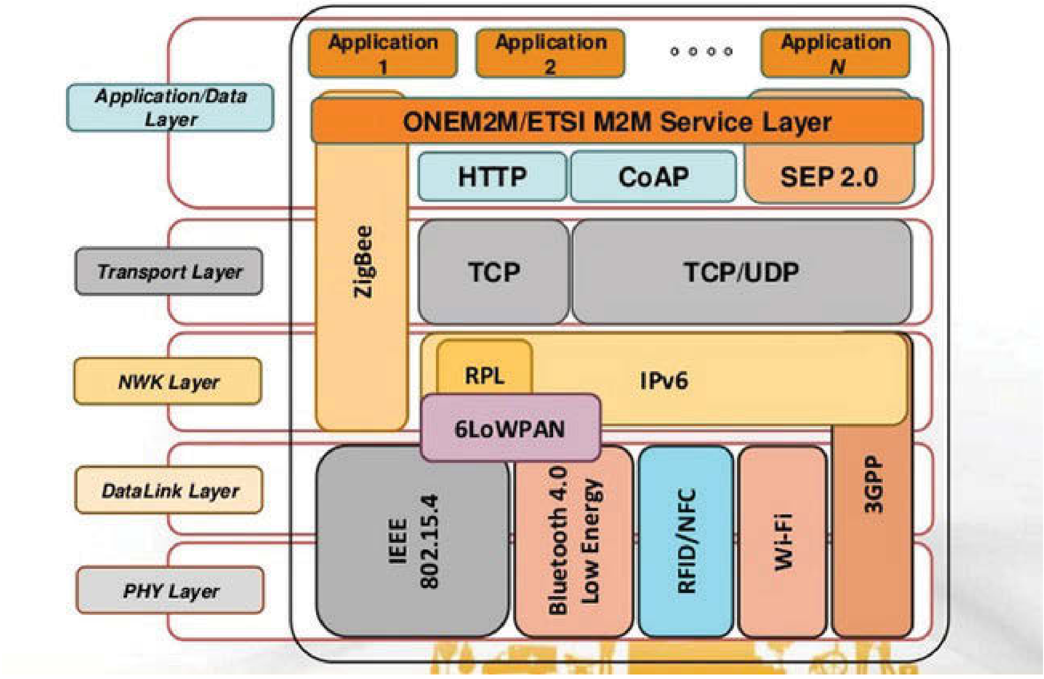


Figure 2.8.: Technological Layers

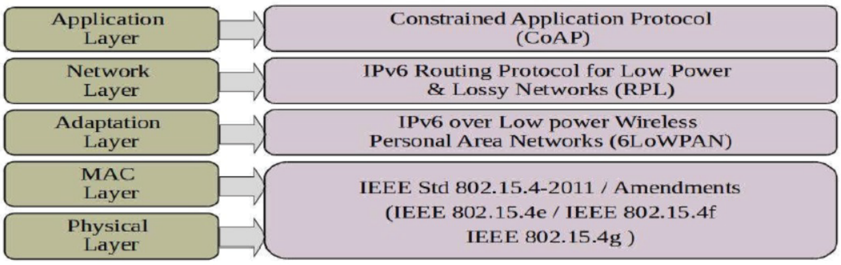

Application Layer
Constrained Application Protocol (CoAP)
Network Layer
IPv6 Routing Protocol for Low Power & Lossy Networks (RPL)
Adaptation Layer
IPv6 over Low power Wireless Personal Area Networks (6LoWPAN)
MAC Layer
Physical Layer
IEEE Std 802.15.4-2011 / Amendments (IEEE 802.15.4e / IEEE 802.15.4f IEEE 802.15.4g)

Figure 2.9.: Emerging Protocols for the IoT. Source: (Abdul 2015)

2.2.6. Big Data Concepts

The relationship between data processing pipelines and large cultural datasets are critical. Heritage and Big Data face challenges in terms of data management and quality of the information (Reiter et al. 2011).

Digital computers entered our homes, landed on our desktops, slipped into our pockets, and have seemingly become ubiquitous. At an ever faster pace, these devices have become highly interconnected and interoperable. Consequently, our archives, our work, our actions, and our interactions are increasingly digitalized and stored in databases or made accessible via the Internet. This data, generally characterized by high volume, variety, and velocity (i.e., accumulation rate), has come to be called "Big Data". As of yet, Big Data has seldom been utilized in management research. Therefore, this dissertation explores the opportunities that Big Data brings for management scholars and describes three distinct projects that show how Big Data can be utilized in management research.

- **Blockchain Definitions:** The main and basic concept of blockchain, is considered as a permissioned ledger technology for sharing and replication that access business networks. This concept improves efficiency and increases access. It highlights the development of business challenges in any industry. The statement about 'Trusted timestamping is a process for proving that certain digital information existed at a given point in time' Gipp et al. (2016) shows how critical the certification of the information through this methodology is. Through the figures specified in Connected Markets Figure 2.10, Nakomoto Figure 2.11, Ripple Figure 2.12, Quorums from the source Figure 2.13 (IBM, 2016) the explanation of the Blockchain organisation have different representations. The nodes are accurate connected and determine the efficiency of the method.

  The original orientation of the Blockchain concept come from the idea to have something represented as a digital entity that can be passed securely from one party to another. Imagine you have some money in your bank account, and you would like to securely transfer some of it to another person. This usually goes through a digital transaction without physical money intervention. This is nice and simple. But, it requires trusting the bank. Is it possible for some digital entity stored on my computer, representing money, to be passed securely to someone's else computer, without a bank? Yes it is possible. However, regardless of representation any digital sequence can be copied.

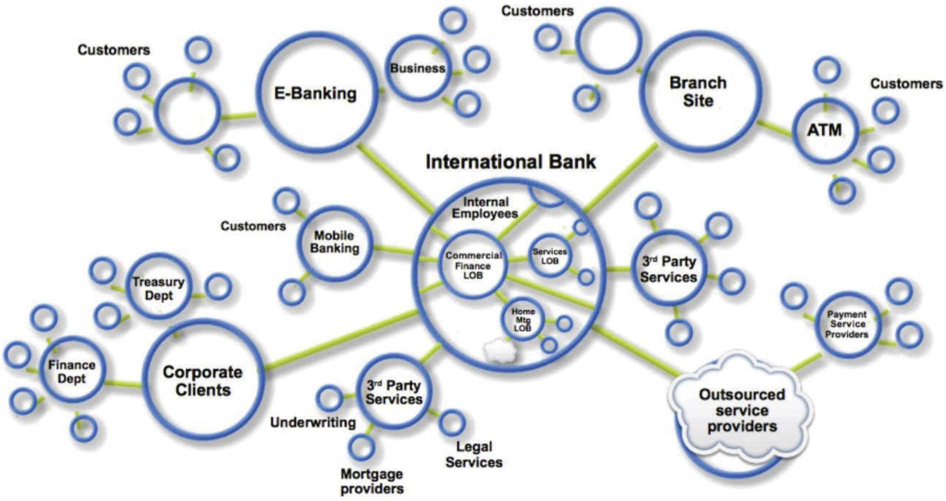


Figure 2.10.: Connected Markets Source: (IBM 2016)

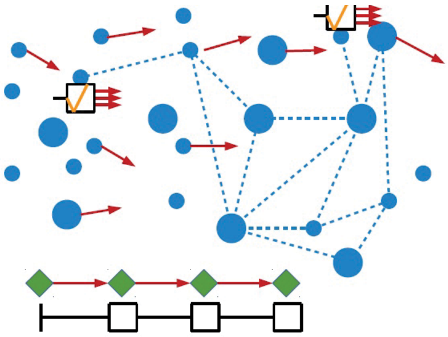


Figure 2.11.: Nakomoto (IBM 2016)

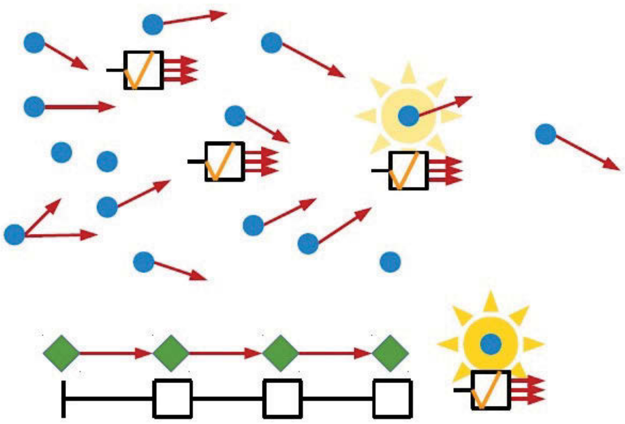


Figure 2.12.: Ripple (IBM 2016)

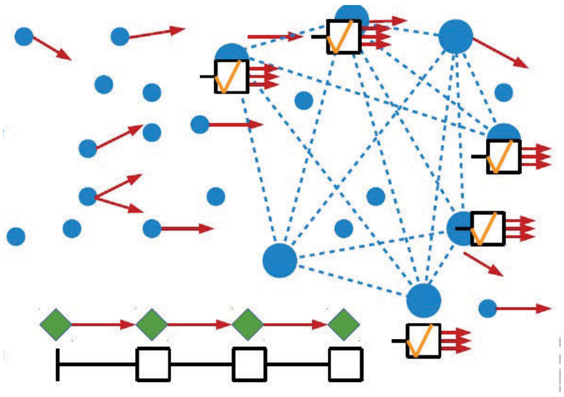


Figure 2.13.: Quorums (IBM 2016)

- **Business Process Management BPM:** This creates a bridge between business analysts, developers and end users. It offers process management features and tools in a way that both business users and developers can improve on. Domain-specific nodes can be plugged into the palette, making the processes more easily understood by business users (Grefen et al. 2009, Chowdhary et al. 2006).
- **Deep Learning:** (deep machine learning, or deep structured learning, or hierarchical learning, or sometimes DL) is machine learning based on a set of algorithms that attempt to model high-level abstractions in data by using multiple processing layers with complex structures, or otherwise composed of multiple non-linear transformations. Bengio et al. (2012) suggested deep

learning is part of a broader family of machine learning methods based on learning representations of data. An observation (e.g., an image) can be represented in many ways such as a vector of intensity values per pixel, or in a more abstract way as a set of edges, regions of particular shape, etc.. Some representations make it easier to learn tasks (e.g., face recognition or facial expression recognition) (Guo et al. 2016).

One of the promises of deep learning is replacing handcrafted features with efficient algorithms for unsupervised or semi-supervised feature learning and hierarchical feature extraction. Research in this area attempts to make better representations and create models to learn these representations from large-scale unlabeled data. Some of the representations are inspired by advances in neuroscience and are loosely based on interpretation of information processing and communication patterns in the nervous system, such as neural coding which attempts to define a relationship between various stimuli and associated neuronal responses in the brain (Guo et al. 2016). Various deep learning architectures such as deep neural networks, convolutional deep neural networks, deep belief networks and recurrent neural networks have been applied to fields such as computer vision, automatic speech recognition, natural language processing, audio recognition and bioinformatics where they have been shown to produce state-of-the-art results in various tasks.

2.2.7. Steganography

Regarding Steganography as a method, Kaur et al. (2015) mentioned that spread spectrum method spreads the secret message bits across the digital audio signals frequency spectrum. The spectrum uses a code that is independent of actual signal. The principal characteristic is capacity defined as the amount of secret information that can be embedded within the host message or host audio signal.

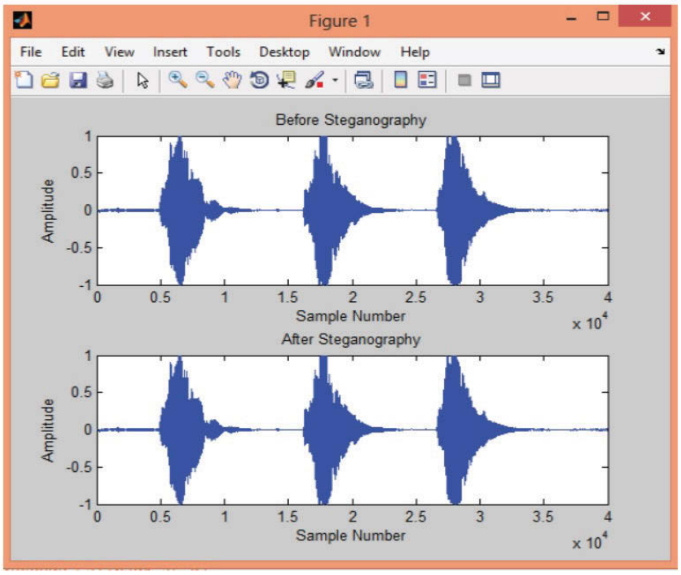
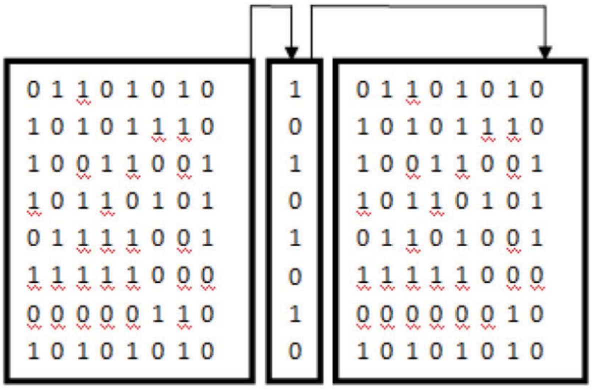


Figure 2.14.: Output encoding Steganography. Source: (Kaur, 2015)

$$Capacity = \frac{actual\ output}{expected\ output}$$

Figure 2.15.: Capacity in Steganography. Source: (Kaur, 2015)

The benefit of replacing the least significant bit (LSB) is that the error is the less as compared to higher bits. See in Figure 2.16 the comparison between bits and the replacement method between the higher and the lower.



0 1 1 0 1 0 1 0	1	0 1 1 0 1 0 1 0
1 0 1 0 1 1 1 0	0	1 0 1 0 1 1 1 0
1 0 0 1 1 0 0 1	1	1 0 0 1 1 0 0 1
1 0 1 1 0 1 0 1	0	1 0 1 1 0 1 0 1
0 1 1 1 1 0 0 1	1	0 1 1 0 1 0 0 1
1 1 1 1 1 0 0 0	0	1 1 1 1 1 0 0 0
0 0 0 0 0 1 1 0	1	0 0 0 0 0 0 1 0
1 0 1 0 1 0 1 0	0	1 0 1 0 1 0 1 0

Figure 2.16.: LSB coding method (Kaur, 2015)

2.2.8. Software for Digital Preservation

There are several initiatives of DHP. When we addresses the solution for Digital Preservation and add Heritage components, in this field, the software solution can facilitate normalisation in terms of records or items. There are three interesting development in this domain These are examples:

XENA (Xml Electronic Normalizing for Archives)

A solution for Digital Transformation and digital preservation which is driven by the National Archives of Australia. The project was inspired by the national needs to support massive amounts (around 380 km shelf space) of long-term digital records (O'Donnell et al. 2010),(Cunliffe 2011). These catalogued records need to be securely archived as they often represent unique and important cultural artefacts. However, important dimensions and relations among the artefacts can often be missed. There is a need for a more integrative and multidimensional approach that would allow for the linking of various digital objects in order to produce new perspectives and insights derived from various sections of digital repositories.

PDF/A

Considered as a normalisation format, PDF/A is based on the PDF format. Adobe Systems in 1993 developed it. It keeps one of the characteristics of read-only PDF files. These normalised files, maintain the appearance of the item from the source but do not generate a record.

2.2.9. Standardization for Digital Preservation

METS (Metadata Encoding and Transmission Standard)

This standard is pointed out as a separate system and is related to the complex records in the parts and 'encapsulates' them in the .xml metadata to allow the reproduction of the item where it is similar to the 'original' form. It allows for open flexibility and improves interoperability

2.2.10. Models supporting Digital Preservation

(CMM) Analysis of Capability Maturity Model

The improvement in software development processes could be focused on the application of CMM. Considering Heritage Preservation as a process, through this model, it can measure and manage the levels as: Initialization, Repeat, Definition, Management and Optimisation. The stages are explained in the following Figure 2.17.

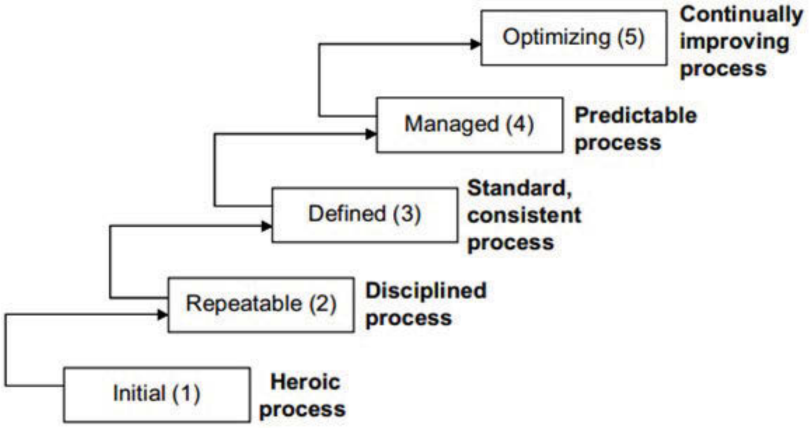


Figure 2.17.: The capability maturity model. Adapted from (Kaner and Karni, 2004)

In the processes of Preservation, the definition of the model introduce Serendipitious and Heritage aspects to consider. Below in Figure 2.18, the concepts are detailed.

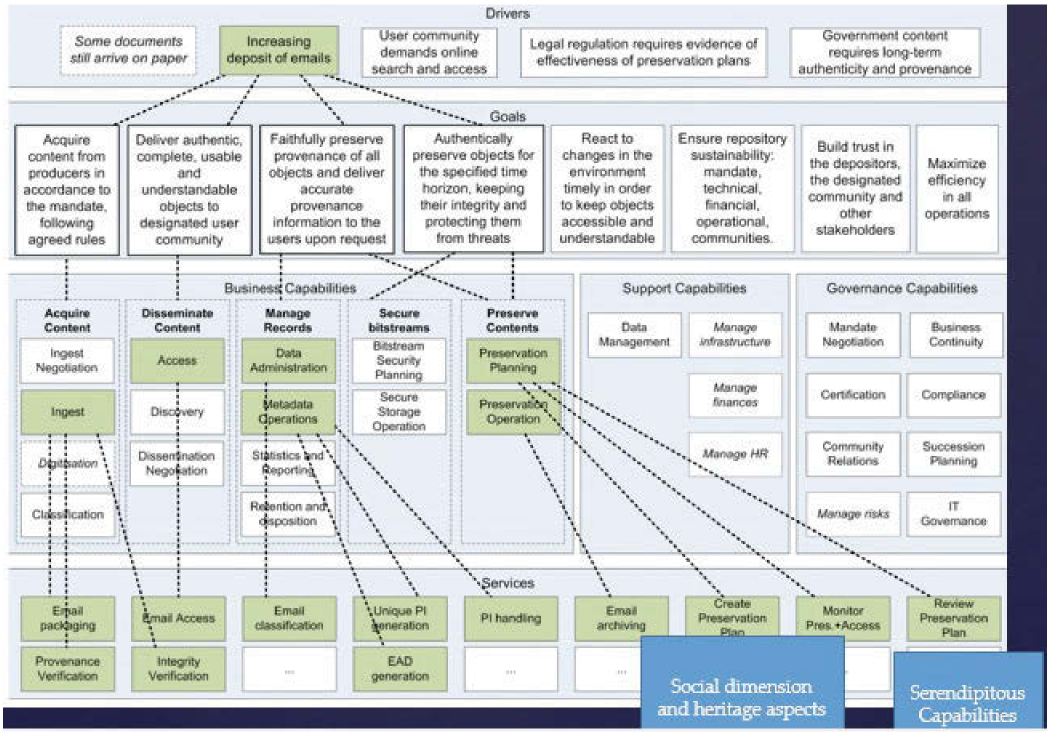

Drivers
Some documents still arrive on paper
Increasing deposit of emails
User community demands online search and access
Legal regulation requires evidence of effectiveness of preservation plans
Government content requires long-term authenticity and provenance
Goals
Acquire content from producers in accordance to the mandate, following agreed rules
Deliver authentic, complete, usable and understandable objects to designated user community
Faithfully preserve provenance of all objects and deliver accurate provenance information to the users upon request
Authentically preserve objects for the specified time horizon, keeping their integrity and protecting them from threats
React to changes in the environment timely in order to keep objects accessible and understandable
Ensure repository sustainability: mandate, technical, financial, operational, communities.
Build trust in the depositors, the designated community and other stakeholders
Maximize efficiency in all operations
Business Capabilities
Acquire Content
Ingest Negotiation
Ingest
Digitisation
Classification
Disseminate Content
Access
Discovery
Dissemination Negotiation
Manage Records
Data Administration
Metadata Operations
Statistics and Reporting
Retention and disposition
Secure bitstreams
Bitstream Security Planning
Secure Storage Operation
Preserve Contents
Preservation Planning
Preservation Operation
Support Capabilities
Data Management
Manage infrastructure
Manage finances
Manage HR
Governance Capabilities
Mandate Negotiation
Business Continuity
Certification
Compliance
Community Relations
Succession Planning
Manage risks
IT Governance
Services
Email packaging
Email Access
Email classification
Unique PI generation
PI handling
Email archiving
Create Preservation Plan
Monitor Pres.+Access
Review Preservation Plan
Provenance Verification
Integrity Verification
...
EAD generation
...
...
Social dimension and heritage aspects
Serendipitous Capabilities

Figure 2.18.: Part of a capability model. Adapted from (Becker, 2011)

2.3. Statistical Preservation Approach with R

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS.

The size of the bubble represents module size in the data subset used to assess module preservation.

R. capsulatus modules found to be conserved in R. sphaeroides are listed on the right side.

(a) Module preservation as a function of module quality

(b) Relationship between the two preservation statistics (Zsummary.pres and medianRank). Lower medianRank indicates higher preservation.

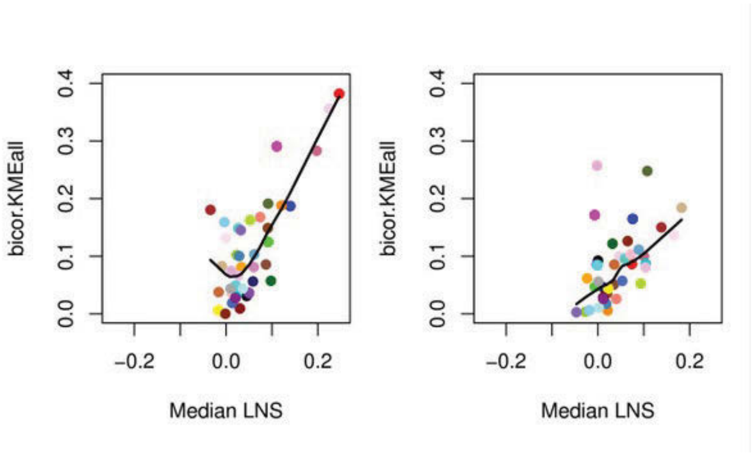


Figure 2.19.: Module Preservation in R (R Project)

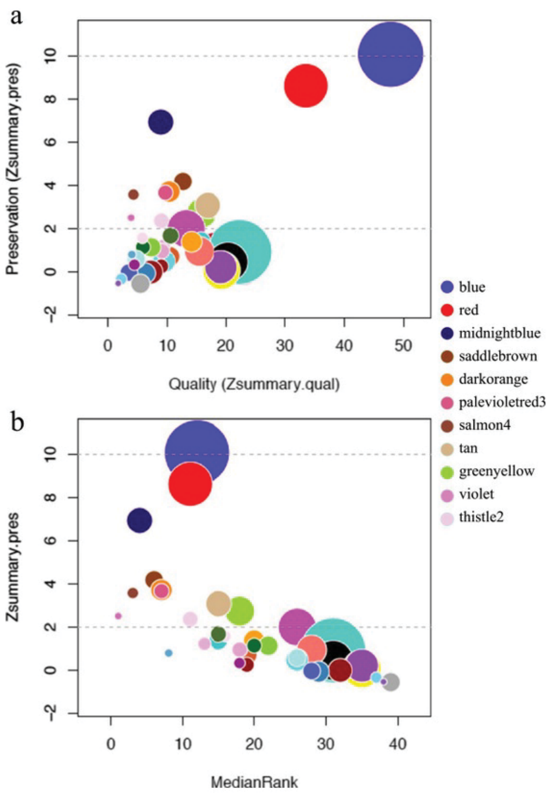


Figure 2.20.: Module Preservation in R. (R project)

The relationship between module preservation connectivity statistics. Total network correlation of the module connectivity (bicor.KMEall) as a function of the median-LNS per module for R.

Each point represents a module labeled by the color corresponding to the module name. The black line is the less smoothed line.

2.4. Database Approach

The Digital Preservation has an important relationship with Databases. In this stage the expression of the model in terms of database, gives us the idea of the model in an organised manner. The components of the Digital Preservation Model are: IoT user, Digital Dataset, Heritage Component, Preservation, Application, D2HP based on DROID, Case Studies: Medical Records, Government Data and Business Cases.

DROID as a tool for metadata idetification, will be explained on detail in chapter 5.

As Figure 2.21 shows, DHP could be represented as entities with relationships and connections. This is a modular manner to understand the different stages and levels of DHP model.

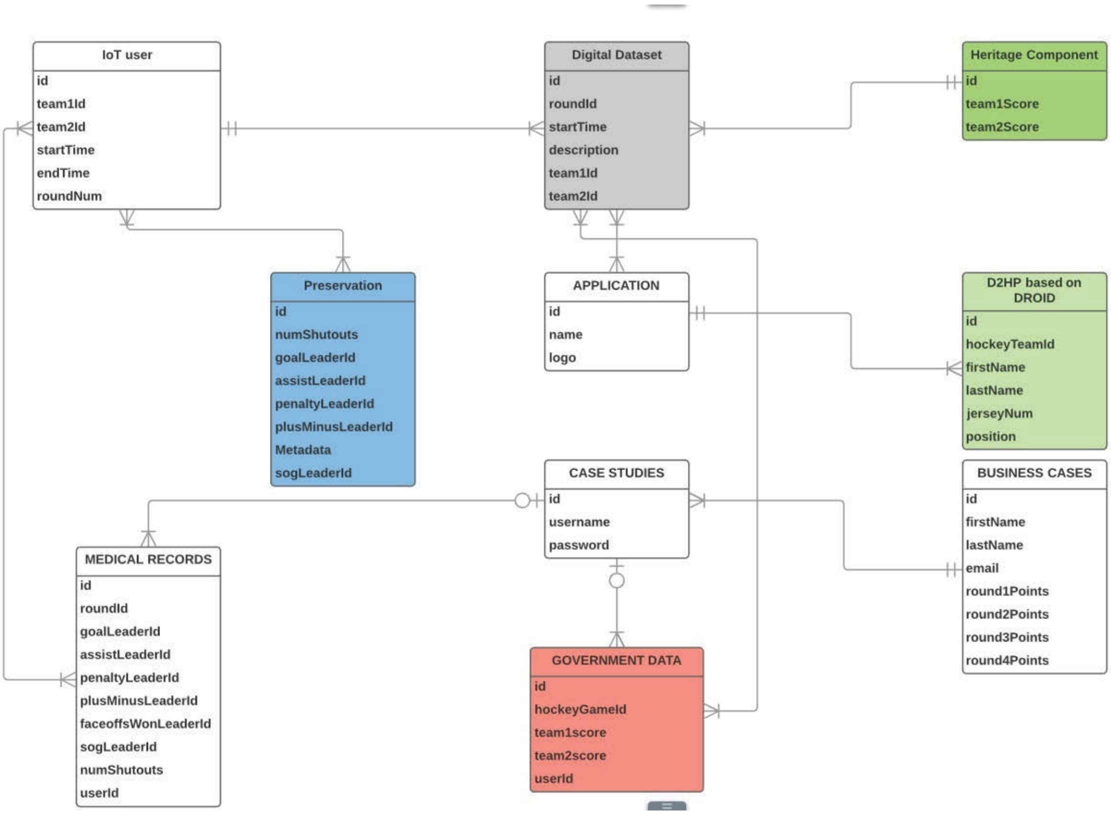


Figure 2.21.: DHP Database Approach using DROID

2.5. Mapping Description

There are three Mapping options for representing the DHP model. The Digital Preservation solutions are considered the best approach for the DHP solutions. They are not the same as in context of Heritage, but the similarities in applications are important. These are the following:

- Open Archival Information System (OAIS) Model
- Digital Curation Centre (DCC) Curation Lifecycle Model
- Digital Preservation Outreach & Education (DPOE) Modules

2.6. Organizational Assessment Frameworks for Digital Preservation

Mandates and business requirements are forcing many organizations to develop digital preservation capabilities to retain digital information over the long-term. Digital preservation (DP) aims to ensure that authentic digital objects remain accessible and usable over time, through a combination of people, technology, and processes. A fundamentally interdisciplinary field, DP has seen considerable contributions from information science, archival science, computer science, information systems, and other fields. As DP standards, methods, and concepts developed, organizations began to invest resources to meet their preservation needs. According to Maemura et al. (2017) interest and focus in recent years has subsequently extended from technical DP research and prototypes for individual tools, techniques, or system components to the deployment and operation of full-fledged archival information systems.

- (RQ1) How has organizational assessment developed in the field of DP?
- (RQ2) What types of research exist on assessment frameworks in DP?
- (RQ3) What gaps exist in the current landscape of research on assessment in DP, as compared to related fields, and what opportunities do these present for further research?

Maturity models themselves are thus just one central component of a comprehensive assessment framework consisting generally of three building blocks: a structured model, which is a ’formal description of some aspects of the physical or social reality for the purpose of understanding and communicating’ Mettler & Rohner (2009); a method, i.e. a systematic means for applying the model and achieving an objective with it; and a tool, a concrete or abstract thing that enables the completion of a particular task. The elements comprising each framework may vary, with the model alone serving as the minimum. Frameworks are also supported by documentation for one or all of these components.

The Figure 2.22 explains the timeline of the Digital Preservation frameworks and the relevant topics about this domain.

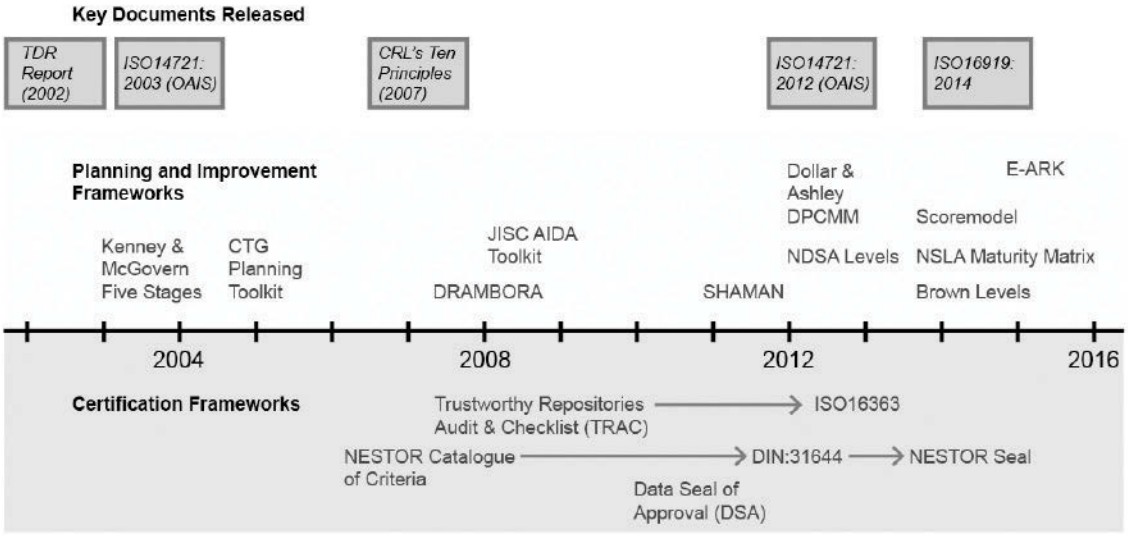
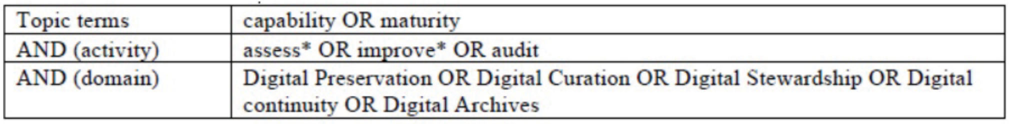


Figure 2.22.: Timeline of assessment frameworks. Source: (Maemura, 2017)

Table 2.1.: Key search terms. Source: (Maemura, 2017)



Topic terms	capability OR maturity
AND (activity)	assess* OR improve* OR audit
AND (domain)	Digital Preservation OR Digital Curation OR Digital Stewardship OR Digital continuity OR Digital Archives

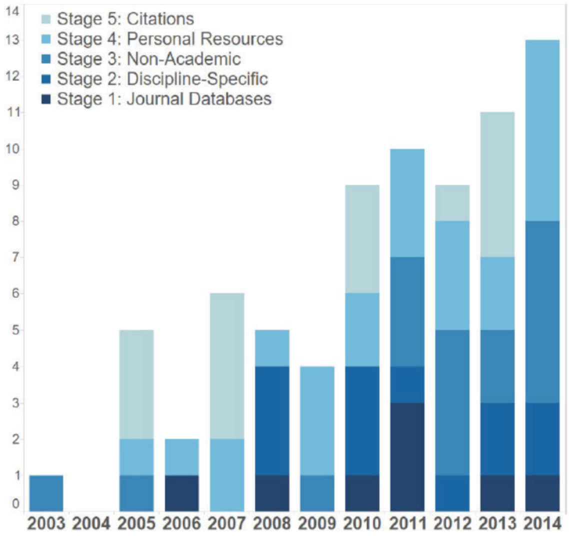


Figure 2.23.: Total publications of Digital Preservation per year by search. Source: (Maemura, 2017)

The terms and the publications per year about the framework are explained in the figures above, Table 2.1 and Figure 2.23. We divided the category of Development into three distinct sub-categories:

- Model Development includes work that introduces an assessment framework or provides full documentation of a model. This sub-category also includes work that substantially builds upon an existing model, such as the introduction of a new version with significant changes.
- Proto-Development includes work that lays the conceptual groundwork or vision for a model that is not yet developed.
- Model Presentation includes work that presents a previously developed model, introducing it to a new community or in a new venue.

The majority of publications in the Development category fall under the main sub-category of model development.

Assessment is a crucial concern for managing digital repositories and archives. It is often carried out as input for strategic decisions and resource investment. Certification may be a key motivation for assessment, as an external warrant of trustworthiness. However, assessment for improvement is just as central to the continued improvement of a repository's ability to preserve. The difference in focus for improvement-oriented assessment means a shift in perspective and raises rather different requirements for the assessment framework.

A critical perspective on design, evaluation and validation is important in that stage of the comparison of the frameworks.

Similar to a phenomenon noted by Wendler, application reports often take the form of “success stories” described by the authors of models. As noted above, the level of scrutiny that assessment frameworks have undergone, and the level of evidence that is available for others to evaluate and scrutinize, is surprisingly low. Similarly, notions of “best practice” are adopted uncritically across the board; while the term features 129 times across 50 of the publications surveyed, it is never itself defined, discussed or reviewed. The amount of evidence available to demonstrate that the practices espoused as “best practice” are consistently showing results superior to other practices is unclear. The question is, then, how valuable an assessment can be for an organization if the foundation of the model used, the clarity of the concepts employed, the suitability of the conceptual structure, and the effectiveness of the assessment method are unclear. Maemura et al. (2017).

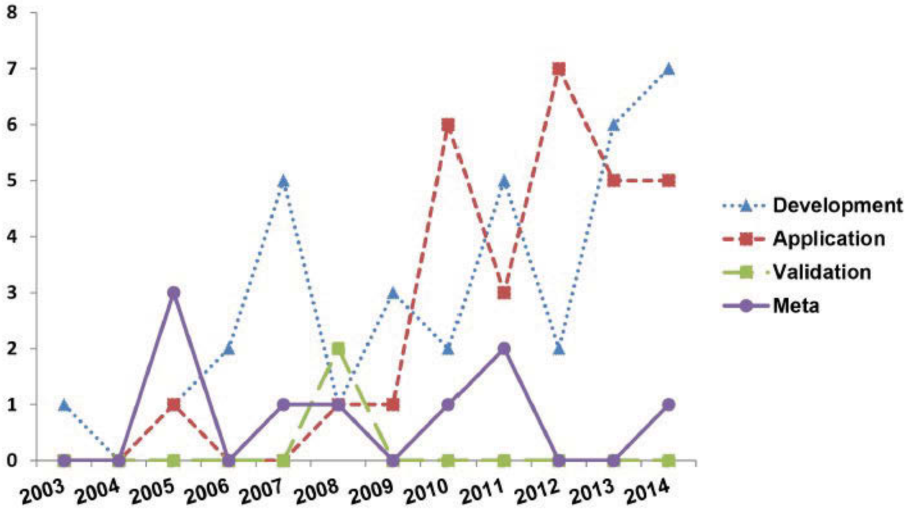


Figure 2.24.: Research type over time. Source: (Maemura, 2017)

2.6.1. Criteria for Benchmarking Solutions

The interest in benchmarking is focused mainly on the technical research community. They have identified three major benchmark components: motivating comparison, task sample and performance measures. It is open the order in which those components are developed.

- The comparison defines the study to be done and the benefits that comparison will bring in terms of the future research agenda. For example, Kienle and Sim motivate their benchmark for fact extraction from web sites by enabling the comparison of capabilities of different fact extractors. Heckman and Williams propose a benchmark for tools that detect anomalies in source code. The main motivation is to find tools with the best rate of anomaly detection.
- The task sample is a list of tests that the subject, to which a benchmark is applied, is expected to solve. Kienle and Sim use both artificial and real web sources as task samples for their web site extractors. Heckman and Williams divide their task sample into two parts: six real Java subject programs and a list of true and false anomalies in those programs.
- The performance measures are qualitative or quantitative measurements taken by a human or a machine to calculate how the subject has incidence for the task. For instance, Heckman and Williams provide a list of well-established measures from the area of data mining and software anomaly detection.

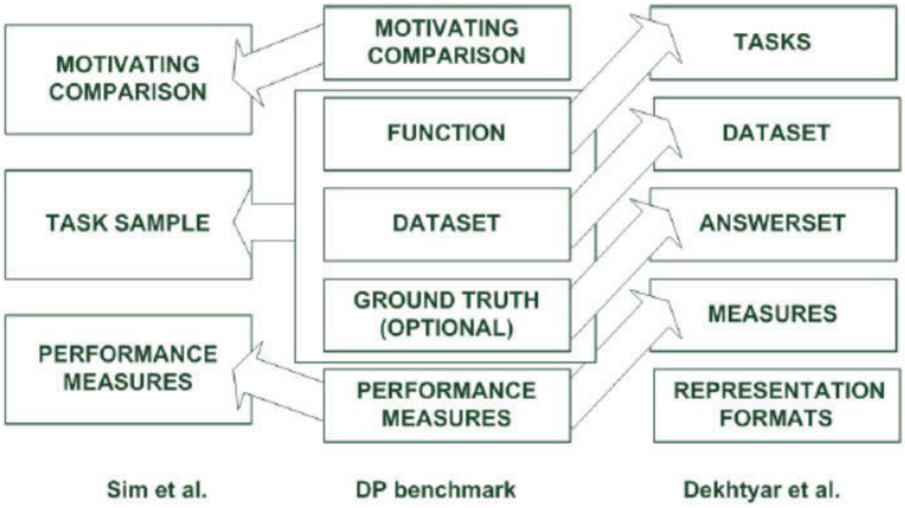


Figure 2.25.: The common benchmark model mapped to the models. Source: (Duretec, 2015)

Benchmarking

BenchmarkDP is developing the first coherent, systematic approach to assess and compare digital preservation processes, systems, and organizational capabilities.

Current theoretical frameworks, engineering approaches and technologies in ICT are not sufficiently capable of efficiently and effectively sustaining the authenticity, understandability and usability of digitally encoded information over time. From its origin in cultural heritage and eScience, digital preservation (DP) has emerged as a key challenge for information systems (IS). Yet, there is a profound lack of theoretically sound and verifiable frameworks to address digital longevity. Currently, numerous strands of applied DP research are hitting a glass ceiling.

In the object and content dimension, there is a profound lack of objective, standardised and comparable metrics and benchmark collections for experimentation. While fields such as Information Retrieval have for decades been able to rely on benchmark collections annotated with ground truth to enable systematic improvement of algorithms and systems along objective metrics, DP is yet unable to provide the necessary ground truth for such benchmarks. These objective indicators, however, are the key enabler for quantitative experimentation and innovation. In the dimension of systems, process and organisations, existing models are not sufficiently considering of digital longevity and information preservation over time. On the other hand, valuable domain knowledge in the area of DP is insufficiently anchored in systematic architecture and design principles. As a consequence, systematic comparison and improvement of processes, systems, and organisational capabilities is not possible. Finally, in the engineering dimension, there is an increasing move towards the ex-post preservation of complex systems and processes rather than solely information. However, up to now, digital longevity is not acknowledged as a fundamental design concern in the IS lifecycle.

Benchmark DP will advance the fundamental understanding of and capacity for digital longevity in ICT by creating theoretical frameworks and quantitative benchmarks on three interlinked dimensions:

For content, we will create a model-driven benchmark generation framework that provides realistic approximations of real-world digital information collections with fully known ground truth that enables systematic quantitative experimentation and measurement. For systems and processes, we will create and evaluate a Capability Maturity Model for DP that enables systematic process improvement and governance of ICT systems with respect to longevity over time. Finally, we will lay the groundwork for pushing longevity upstream and address it as a fundamental design concern in the IS lifecycle.

By providing the first systematic, quantified approach to measuring, controlling and improving DP processes and organisational capabilities, Benchmark DP will enhance the ability of organisations to control and improve processes for information management and IS design to achieve desired levels of longevity.

2.6.2. Collaborative benchmarking in digital preservation

Applied research and technology development efforts in Digital Preservation (DP) have invested substantial resources into the development and maintenance of tools to support key DP processes. These include file format identification, migration and emulation tools, quality assurance mechanisms, automated annotation, and digital forensics, to name but a few. The need to evaluate these tools systematically has become more pressing as they are increasingly being deployed in operational digital archives and repository systems.

Systematic evaluation enables the community of researchers, solution providers and content holders to establish the systematic sharing, aggregation and analysis of evidence. In the context of software systems, this makes rigorous experimentation a particularly relevant mode of inquiry.

Fields such as Information Retrieval (IR) and Software Engineering (SE) have adopted benchmarking, a specific mode of systematic experimentation, as a core component of their research agenda. In these communities, a systematic, well-defined process exists through which members of the community create and share evidence about specific products in an organized, rigorous process, in order to assess and compare these products according to accepted measures of success. A benchmark is 'a set of tests used to compare the performance of alternative tools or techniques'. This means that for different kinds of tools, different tests need to be developed. A key obstacle to the feasibility of such benchmarks has been the lack of well-annotated data sets available to facilitate comparative testing.

While the exact definition and structure of a benchmark varies across these disciplines, the underlying purpose is similar. A software benchmark is a systematic, repeatable method of comparing software tools reliably for a particular purpose. In digital preservation, this generally requires five main components: a motivating comparison that specifies the purpose of creating an evaluation and comparison test; the specific function to be performed, and a dataset on which it should be performed; ground truth, i.e. accurate and provably correct answers; and the performance measures to be collected. Note that performance here is not limited to speed, but denotes any success measures to be compared.

Increasing consensus is emerging on the types of tasks that need to be supported by DP tools. Our research aims to establish systematic evaluation and sharing of evidence about software tools as a technique in digital preservation. To be effective, this must be a community-driven process.

2.6.3. Objectives and participants

It covered key roles and perspectives on this topic:

1. Practitioners use DP systems and solutions to preserve digital material. The need to choose tools carefully in organizational decision making is addressed in preservation planning, and the systematic analysis of such decisions has yielded research priorities for automation. The direct involvement of practitioners in benchmark definitions ensures that priorities for benchmarking initiatives focus on those kinds of tools that are most in need of systematic comparison. This allows practitioners to influence the definition of benchmarks to reflect their decision needs, terminology, and practice.

2. Developers create and improve software tools. Involvement in benchmarking allows these tools to undergo rigorous testing and comparison and enables a well-supported demonstration of improvement over time. Robust data sets and comparison metrics make testing more effective and it's results more visible. This provides an opportunity to promote specific solutions and supports a deeper understanding of quality aspects and features that are most important in order to support the allocation of limited resources.

3. Researchers are driven by a variety of interests, focusing on interesting questions rather than answers. For some, benchmarking provides a rigorous mechanism to corroborate technology innovation. Others are interested in the social collaboration aspects of benchmarking, and in methodological questions of rigorous experimentation and evaluation.

4. For the community at large, a common ground and joint roadmap for systematic evaluation and comparison of software tools is important for the growth of an emerging discipline. Member organizations such as the Digital Preservation Coalition and the Open Preservation Foundation (OPF) are a natural meeting place and platform to advance this collaborative effort.

2.6.4. Benchmarking theory and practice

Provided an overview of benchmarking initiatives, providing examples and lessons learned in settings such as music information retrieval (MIREX), multimedia (LifeCLEF) and text retrieval settings (TREC).

Common theme about benchmarking arose that were particularly relevant for the DHP model:

- Benchmarking poses specific challenges to data quality. In particular, the availability of data and the evaluation of data quality can be difficult. Music IR faced legal challenges that prevented the release of open data for copyright reasons. Robust ground truth annotation, a crucial ingredient, is often challenging to create. Providing the ground truth through human evaluation is expensive. For many tasks, human evaluation (e.g. via the Evalutron 6000) is essential.
- Benchmarking benefits from central platforms. Managing the data sets centrally and reliably turned out to be a key factor and an essential requisite for reproducibility.
- Benchmarks may proliferate once the community gets started. After the initial establishment of benchmarking as a method and platform, the number of different tasks that are proposed and evaluated can grow drastically. Tasks can emerge through a community process of proposal and voting.
- Performance measures can be contentious. It is not easy to agree on clear and well-specified criteria for success.
- Timelines and joint actions need to be established. Benchmarking requires effective coordination of contributions from a wide range of community members. This means that clarity on timelines, expectations and commitments is needed.

A discussion ensued in which comparisons were made between IR, SE, and DP. In IR, the participants of benchmarking campaigns are research teams from industry and academia, and there is considerable competition and continuous improvement of algorithms. In contrast, many tools in DP are still used off-the-shelf. The community is in a middle ground, then, between IR and SE. This is reflected in the specification of the benchmark components as well. Legal barriers have also obstructed the wide availability of data sets in DP.

2.6.5. Benchmarks to consider

Three proposed starting points for these benchmark specifications were clearly scoped and aimed to be developed in further specifications.

1. Artur Kulmukhametov proposed the photo migration benchmark. It measures functional correctness to enable a ranking of migration tools using a dataset of raw photographs. The function to benchmark is the migration of photographs from proprietary raw formats to the Adobe Digital Negative (DNG) format. This is a practical problem for professionals and institutions when selecting the best tool for migration of raw photographs. The motivation for this benchmark is the need to compare the correctness of migrations done by software tools on the photograph dataset. To define when a migration is objectively correct, a set of performance measures is proposed. There is no robust data set at this point and no robust ground truth, but criteria for compiling a data set exist.
2. Krešimir Đuretec presented a document property extraction benchmark, designed to facilitate comparison of characterization tools for electronic document formats. The motivation for this benchmark is to enable comparison of the coverage and correctness of characterization tools for electronic documents, with a focus on the MS Word 2007 file format. The function thus is characterization, or feature extraction, and the performance measures combine the coverage of existing properties in a data set and the accuracy of characterization. The benchmark is highly specified in terms of success metrics, and narrowly focused in terms of the suggested data set. Conducting ground truth annotations is crucial and approaches to generating data to facilitate such ground truth are under development.
3. Bengt Neiss presented one of the three Preforma benchmarks. For each major scenario in the project, one benchmark is being created using the same structure as described above. Given the project focus on format conformance, the structure of these three benchmarks is identical. One motivation to define this benchmark is to facilitate comparison between different releases of the software tools developed within the Preforma project. The tools are expected to perform four functions:
   a) verify conformance of files to the specifications of a particular set of standards,
   b) verify compliance of files with institutional acceptance criteria,

c) report deviations in human-readable and machine-readable form,

d) fix certain basic errors in the metadata of the files that are evaluated.

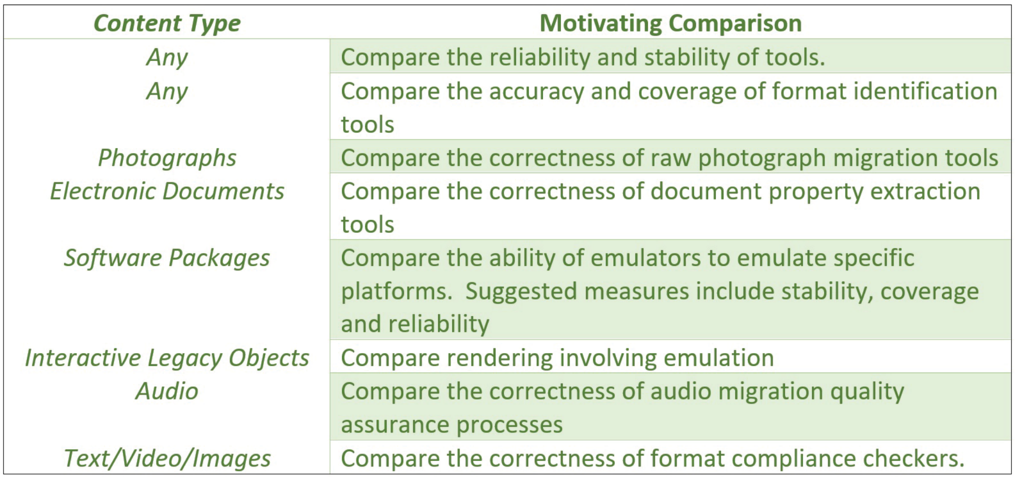
The **data set** necessary to do so is under development and envisioned to draw from three sources:

1. a set of files that is declared to be the reference representation,
2. synthetic files with particular conformance problems,
3. real 'live' files with unknown properties.

Open issues discussed and included performance measures, the composition of the data set, and possible approaches to compiling it.

Next, the group collaboratively assembled possible scenarios for benchmarking. The following Table 2.2 and Table 2.3 presents a summary of identified comparison choices that were collected.

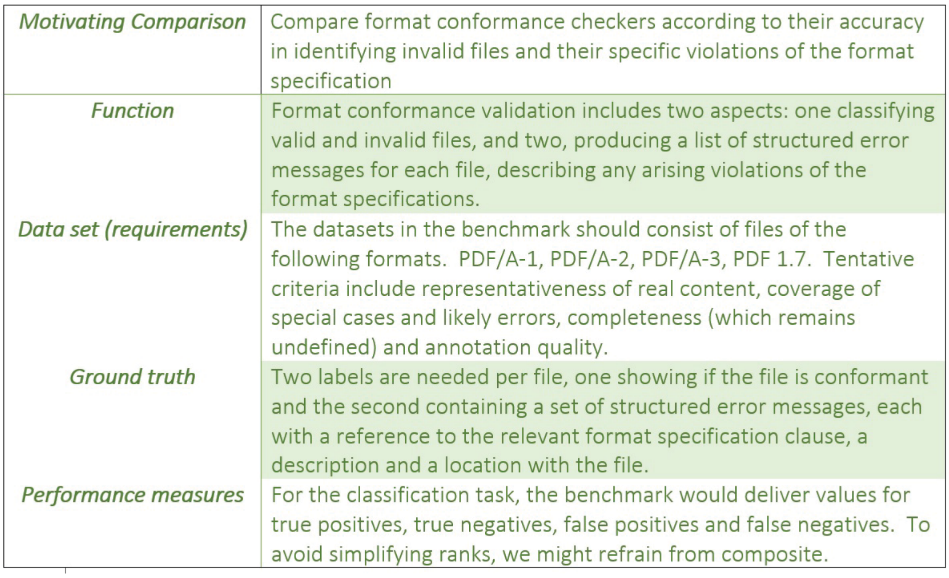
Table 2.2.: Identified Content and Comparison



Content Type	**Motivating Comparison**
Any	Compare the reliability and stability of tools.
Any	Compare the accuracy and coverage of format identification tools
Photographs	Compare the correctness of raw photograph migration tools
Electronic Documents	Compare the correctness of document property extraction tools
Software Packages	Compare the ability of emulators to emulate specific platforms. Suggested measures include stability, coverage and reliability
Interactive Legacy Objects	Compare rendering involving emulation
Audio	Compare the correctness of audio migration quality assurance processes
Text/Video/Images	Compare the correctness of format compliance checkers.

An interesting discussion revolves around the need to evaluate emulators as well as rendering tools. While sometimes seen as one specific evaluation scenario, there are two aspects to be distinguished. Evaluation of rendering is a critical and challenging need; evaluating the technical capabilities of emulators is similarly challenging, but different in focus. Evaluation of rendering is independent of the technical stack: the same approach is needed to evaluate migration and emulation.

Table 2.3.: Identified Data for Benchmarking



Motivating Comparison	Compare format conformance checkers according to their accuracy in identifying invalid files and their specific violations of the format specification
Function	Format conformance validation includes two aspects: one classifying valid and invalid files, and two, producing a list of structured error messages for each file, describing any arising violations of the format specifications.
Data set (requirements)	The datasets in the benchmark should consist of files of the following formats. PDF/A-1, PDF/A-2, PDF/A-3, PDF 1.7. Tentative criteria include representativeness of real content, coverage of special cases and likely errors, completeness (which remains undefined) and annotation quality.
Ground truth	Two labels are needed per file, one showing if the file is conformant and the second containing a set of structured error messages, each with a reference to the relevant format specification clause, a description and a location with the file.
Performance measures	For the classification task, the benchmark would deliver values for true positives, true negatives, false positives and false negatives. To avoid simplifying ranks, we might refrain from composite.

2.6.6. The Proforma conformance checking scenario, with a focus on PDF

A set of open issues and follow-up actions were defined that the Proforma and BenchmarkDP project teams will take forward.

The discussion built directly on the scenario presented earlier, but focused primarily on the data set. Criteria include representativeness, completeness (defined as coverage of originating camera models), coverage of special cases and likely sources of error (defined as the presence of specific distribution of features and metadata tags). Generating files artificially for this case seems infeasible. Instead, we focused on collecting and potentially crowd-sourcing strategies for compiling a robust data set. Several practitioners know of potential data in their organizations and will reach out to identify opportunities for sharing information.

Next steps include building a list of camera models to be covered with the first benchmark; developing a set of properties associated with boundary test cases and likely errors; and use content profiling tools for purposeful sampling. Data quality may be low initially and improve over time, so it is of paramount importance to develop a structured method of measuring data quality.

Several specific roles and contributions were identified:

The Open Preservation Foundation plans to host an open test platform that facilitates automated execution of comparison tests and provides continuity for hosting the data sets. BenchmarkDP and Proforma are developing a Memorandum of Understanding for close collaboration between the projects to facilitate robust, objective comparison of tools. The Preforma project team will continue to collaborate with BenchmarkDP on the specification of benchmarks, which will be published and distributed widely. BenchmarkDP will coordinate efforts to specify and share, in standardized structured form, draft benchmarks under development and completed benchmarks.

2.6.7. Tool grid Digital Preservation Solutions

This tool grid is the product of researching digital preservation tools by Digital POWRR (Preserving digital Objects with Restricted Resources) Mita (2016).

The evaluation of the preservation tools is based on aspects: Ingest, Processing, Access, Storage, Maintenance and other.

The following tables Table 2.4,Table 2.5, Table 2.6 and Table 2.7 mention several solutions and scores them according to the criteria postulated.

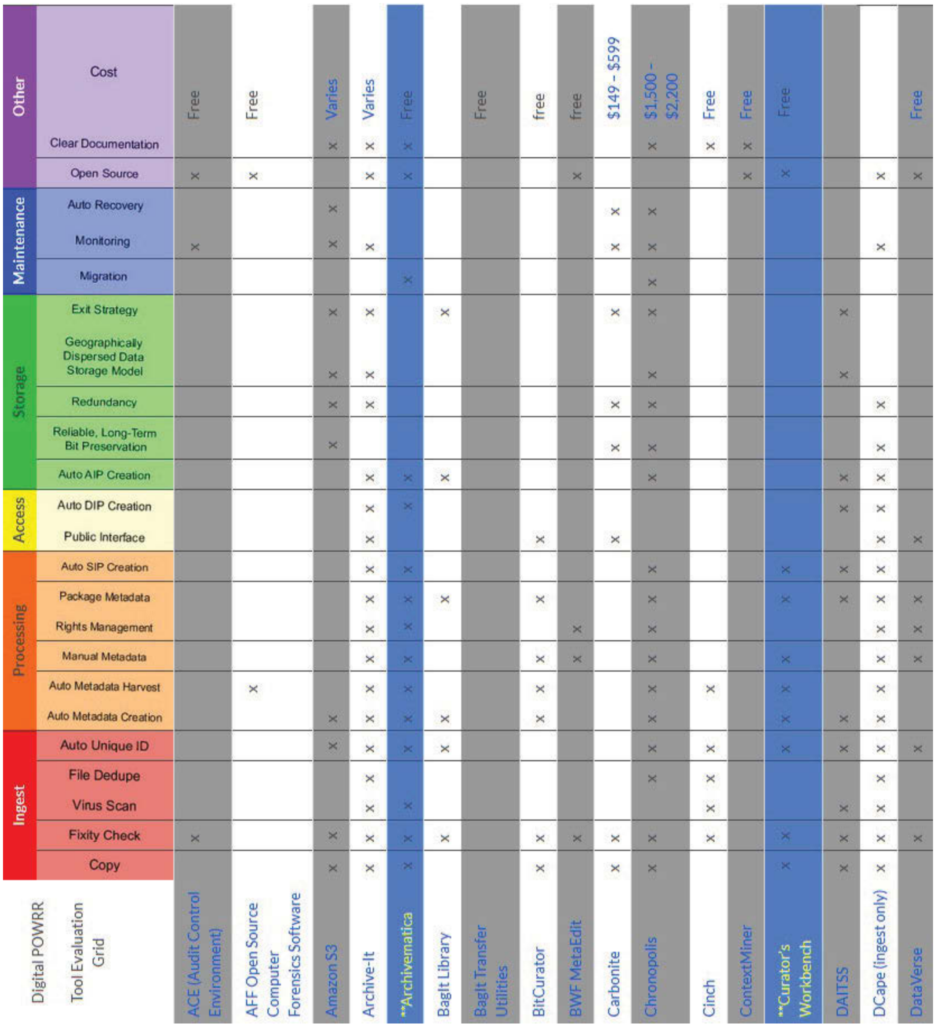
Going on depth with every field in the tables below, there are six areas of study. In every field there is subcategories to consider relevant in the research. The analysis explain which of them are important and in which stage.

- Ingest
  - Auto Unique ID and Fixity Check: The identification and the check of the integrity of data is one of the priorities in the study.
- Processing
  - Package Metadata, Right Management, Manual Metadata, Auto Metadata Creation: Metadata as a definition is one of the pillar of the research. Management in terms of processing information is relevant, too.
- Access
  - Public Interface: The accessibility to the information in open way is one of the considerations for the best tool evaluation..
- Storage

- Storage Model, Redundancy, Reliable, Long-Term, Bit Preservation: The preservation of data and to keep it in reliable manner is the one of the storage considerations.

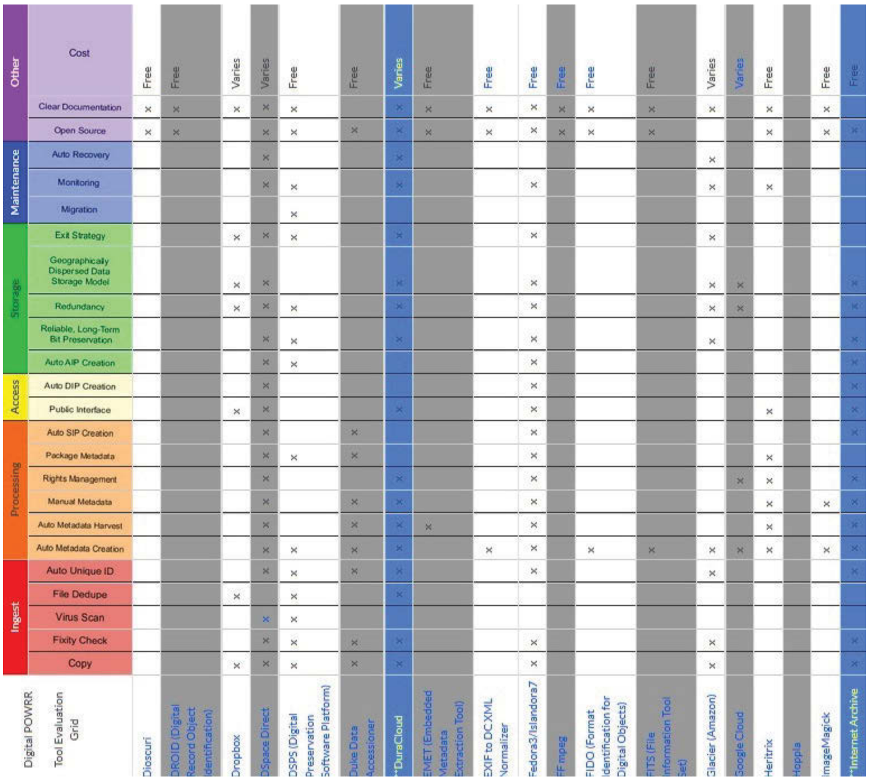
- Maintenance
    - Auto recovery, Monitoring, Migration: The maintenance of data considering the way of monitoring and the migration of it, are strong definitions for the research,
- Other
    - Cost, Clear Documentation, Open Source: How to documentate the real information and if the tool is open source, define the correct tool as a DHP solution.

Table 2.4.: Tool grid Solutions Comparison 1. Source: (Mita, 2016)



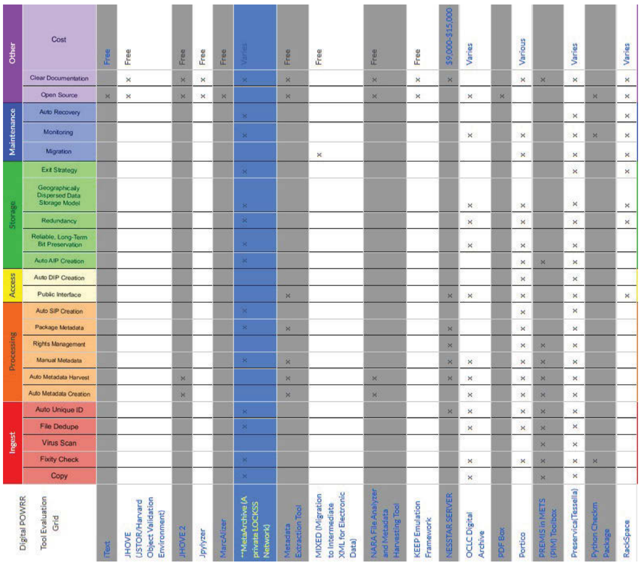
Digital POWRR Tool Evaluation Grid	Ingest					Processing						Access		Storage					Maintenance			Other		
	Copy	Fixity Check	Virus Scan	File Dedupe	Auto Unique ID	Auto Metadata Creation	Auto Metadata Harvest	Manual Metadata	Rights Management	Package Metadata	Auto SIP Creation	Public Interface	Auto DIP Creation	Auto AIP Creation	Reliable, Long-Term Bit Preservation	Redundancy	Geographically Dispersed Data Storage Model	Exit Strategy	Migration	Monitoring	Auto Recovery	Open Source	Clear Documentation	Cost
ACE (Audit Control Environment)		×																		×		×		Free
AFF Open Source Computer Forensics Software							×															×		Free
Amazon S3	×	×			×	×									×	×	×	×		×	×		×	Varies
Archive-It	×	×	×	×	×	×	×	×	×	×	×	×	×	×		×	×	×		×		×	×	Varies
**Archivematica	×	×	×		×	×	×	×	×	×	×		×	×					×			×	×	Free
BagIt Library		×			×	×				×				×				×						
BagIt Transfer Utilities																								Free
BitCurator	×	×				×	×	×		×		×												free
BWF MetaEdit		×						×	×													×		free
Carbonite	×	×										×			×	×		×		×	×			$149 - $599
Chronopolis	×	×		×	×	×	×	×	×	×	×			×	×	×	×	×	×	×	×		×	$1,500 - $2,200
Cinch		×	×	×	×		×																×	Free
ContextMiner																						×	×	Free
**Curator's Workbench	×	×			×	×	×	×		×	×											×		Free
DAITSS	×	×	×		×	×				×	×		×	×			×	×						
DCape (ingest only)	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×				×		×		
DataVerse		×			×			×	×	×		×										×		Free

Table 2.5.: Tool grid Solutions Comparison 2. Source: (Mita, 2016)



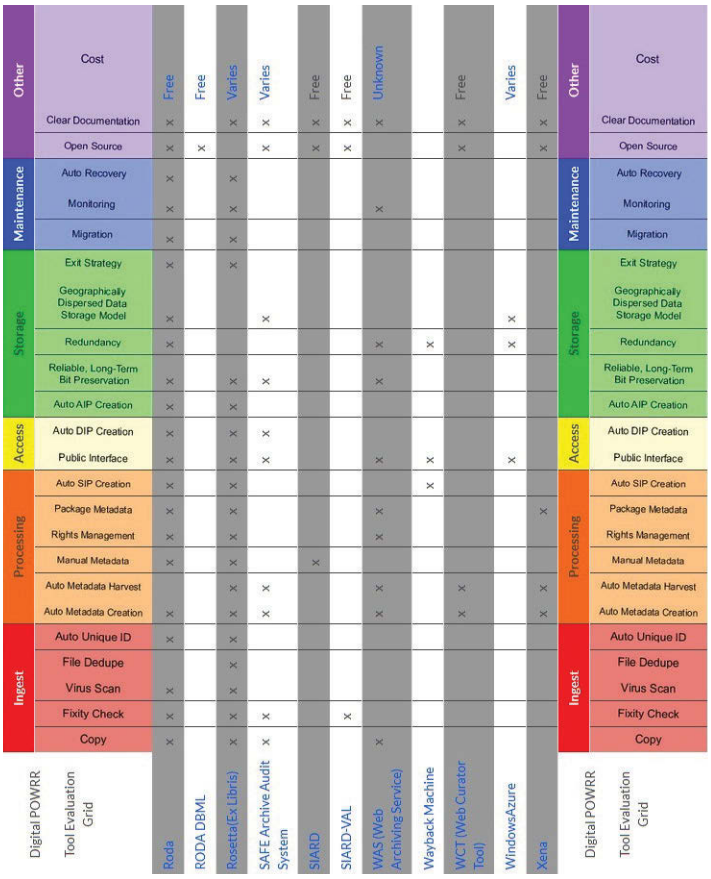
Digital POWRR Tool Evaluation Grid	Other			Maintenance			Storage					Access		Processing						Ingest				
	Cost	Clear Documentation	Open Source	Auto Recovery	Monitoring	Migration	Exit Strategy	Geographically Dispersed Data Storage Model	Redundancy	Reliable, Long-Term Bit Preservation	Auto AIP Creation	Auto DIP Creation	Public Interface	Auto SIP Creation	Package Metadata	Rights Management	Manual Metadata	Auto Metadata Harvest	Auto Metadata Creation	Auto Unique ID	File Dedupe	Virus Scan	Fixity Check	Copy
Dioscuri	Free	×	×																					
DROID (Digital Record Object Identification)	Free	×	×																					
Dropbox	Varies	×					×	×	×				×								×			×
DSpace Direct	Varies	×	×	×	×		×	×	×	×	×	×	×	×	×	×	×	×	×	×		×	×	×
DSPS (Digital Preservation Software Platform)	Free	×	×		×	×	×		×	×	×				×				×	×	×	×	×	×
Duke Data Accessioner	Free		×											×	×		×	×	×	×			×	×
**DuraCloud	Varies	×	×	×	×		×	×	×	×			×			×	×	×	×	×	×		×	×
EMET (Embedded Metadata Extraction Tool)	Free	×	×															×						
EXIF to DC XML Normalizer	Free	×	×																×					
Fedora3/Islandora7	Free	×	×		×		×	×	×	×	×	×	×	×	×	×	×	×	×	×			×	×
FF mpeg	Free	×	×																					
FIDO (Format Identification for Digital Objects)	Free	×	×																×					
FITS (File Information Tool Set)	Free	×	×																×					
Glacier (Amazon)	Varies	×		×	×		×	×	×	×									×	×			×	×
Google Cloud	Varies							×	×							×			×					
Heritrix	Free	×	×		×								×		×	×	×	×	×					
Hoppla																								
ImageMagick	Free	×	×														×		×					
**Internet Archive	Free		×					×	×	×	×	×	×	×		×	×	×	×	×			×	×

Table 2.6.: Tool grid Solutions Comparison 3. Source (Mita, 2016)



Digital POWRR Tool Evaluation Grid	Other			Maintenance			Storage					Access		Processing						Ingest				
	Cost	Clear Documentation	Open Source	Auto Recovery	Monitoring	Migration	Exit Strategy	Geographically Dispersed Data Storage Model	Redundancy	Reliable, Long-Term Bit Preservation	Auto AIP Creation	Auto DIP Creation	Public Interface	Auto SIP Creation	Package Metadata	Rights Management	Manual Metadata	Auto Metadata Harvest	Auto Metadata Creation	Auto Unique ID	File Dedupe	Virus Scan	Fixity Check	Copy
iText	Free		×																					
JHOVE (JSTOR/Harvard Object Validation Environment)	Free	×	×																					
JHOVE 2	Free	×	×															×	×					
Jpylyzer	Free	×	×																					
MarcAlizer	Free		×																					
**MetaArchive (A private LOCKSS Network)	Varies	×		×	×		×	×	×	×	×			×	×		×			×	×		×	×
Metadata Extraction Tool	Free	×	×										×		×		×	×	×					
MIXED (Migration to Intermediate XML for Electronic Data)	Free					×																		
NARA File Analyzer and Metadata Harvesting Tool	Free	×	×															×	×					
KEEP Emulation Framework	Free	×	×																					
NESSTAR SERVER	$9,000-$15,000	×											×		×	×	×	×		×				
OCLC Digital Archive	Varies		×		×			×	×	×			×				×	×	×	×	×		×	×
PDF Box			×																					
Portico	Various	×			×	×		×	×	×	×	×	×	×	×	×	×	×	×	×	×		×	
PREMIS in METS (PIM) Toolbox		×									×					×	×	×	×	×	×	×	×	×
Preservica(Tessella)	Varies	×		×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
Python Checkm Package			×		×																		×	
RackSpace	Varies	×	×	×	×	×	×	×	×				×											

Table 2.7.: Tool grid Solutions Comparison 4. Source: (Mita, 2016)



Digital POWRR	Tool Evaluation Grid	Roda	RODA DBML	Rosetta(Ex Libris)	SAFE Archive Audit System	SIARD	SIARD-VAL	WAS (Web Archiving Service)	Wayback Machine	WCT (Web Curator Tool)	WindowsAzure	Xena	Digital POWRR	Tool Evaluation Grid
Other	Cost	Free	Free	Varies	Varies	Free	Free	Unknown		Free	Varies	Free	Other	Cost
	Clear Documentation	×		×	×	×	×	×		×		×		Clear Documentation
	Open Source	×	×		×	×	×			×		×		Open Source
Maintenance	Auto Recovery	×		×									Maintenance	Auto Recovery
	Monitoring	×		×				×						Monitoring
	Migration	×		×										Migration
Storage	Exit Strategy	×		×									Storage	Exit Strategy
	Geographically Dispersed Data Storage Model	×			×						×			Geographically Dispersed Data Storage Model
	Redundancy	×						×	×		×			Redundancy
	Reliable, Long-Term Bit Preservation	×		×	×			×						Reliable, Long-Term Bit Preservation
	Auto AIP Creation	×		×										Auto AIP Creation
Access	Auto DIP Creation	×		×	×								Access	Auto DIP Creation
	Public Interface	×		×	×			×	×		×			Public Interface
Processing	Auto SIP Creation	×		×					×				Processing	Auto SIP Creation
	Package Metadata	×		×				×				×		Package Metadata
	Rights Management	×		×				×						Rights Management
	Manual Metadata	×		×		×								Manual Metadata
	Auto Metadata Harvest			×	×			×		×		×		Auto Metadata Harvest
	Auto Metadata Creation	×		×	×			×		×		×		Auto Metadata Creation
Ingest	Auto Unique ID	×		×									Ingest	Auto Unique ID
	File Dedupe			×										File Dedupe
	Virus Scan	×		×										Virus Scan
	Fixity Check	×		×	×		×							Fixity Check
	Copy	×		×	×			×						Copy

2.7. Knowledge Engineering

The knowledge engineering is the domain that covers components as: Knowledge Acquisition that is capturing and organizing of implicit (tacit) and explicit knowledge and the acquisition of (expert) knowledge is the bottleneck of KE. Other consideration is the formalization. There are two steps, computer-processable modeling and representation of knowledge within a knowledgebase.

On the other hand the Knowledge Processing for problem solving e.g. with inference (reasoners) or expert systems and the knowledge Representation (visualization).

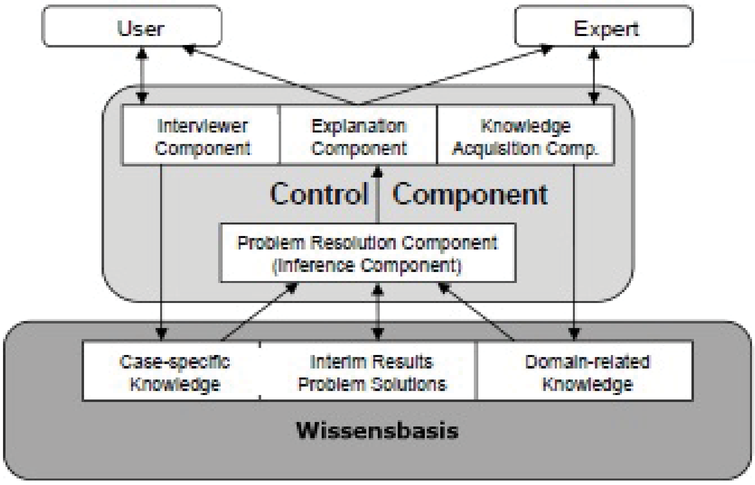


Figure 2.26.: Knowledge Engineering

2.8. Cost Model DP

Based on the publication stated Gordon et al. (2015) there are cost factors which affect Digital Preservation Model.
Cost equations

$$Costperactivity = (Time \times Wage) + Purchase \tag{2.1}$$

$$c(a) = \sum_{i-0}^{N1} t_i * \sum_{i-0}^{N1} W_i + P \tag{2.2}$$

Costing Preservation Planning and Digital Migrations While the goal is to model the whole lifecycle of digital preservation, the first version of the model only deals with the cost of Preservation Planning and digital migrations. The factor is defined by the time it takes to identify and understand a format's specifications and any other relevant documentation(Kejser et al. 2011*b*). This depends on the amount of documentation (number of pages); the complexity of the documentation (low, medium, high); and on the quality of the documentation (low, medium, high), reflecting how flawed and inadequate it is:

$$FI = NoP \times TPP \times Cx \times Q \tag{2.3}$$

FI: Format Interpretation

NoP: Number of Pages

TPP: Time per Page

Cx: Complexity

Q: Quality

2.9. Sensitivity: Personal Information Factor

Data sets that do not identify particular individuals may be used to create personally identifiable information if other data sets are accessed which enable identification of the individuals to whom the shared data sets relate. This other information might be available either:

- internally - for example, by looking up another data set and cross-matching transaction data sorted by transactor key or device identifier

- externally, such as re-identification of individuals through matching of data sets through use of searchable databases such as ASIC records, Land Titles Office property records or through search engines.

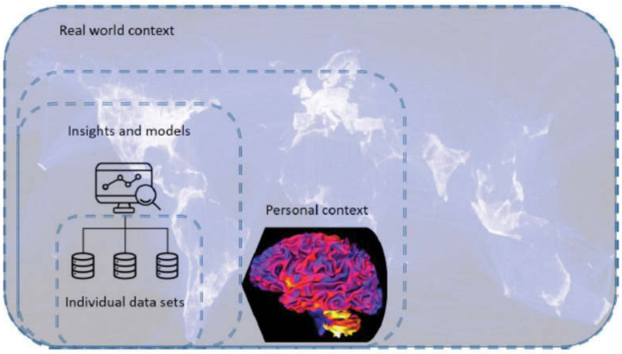
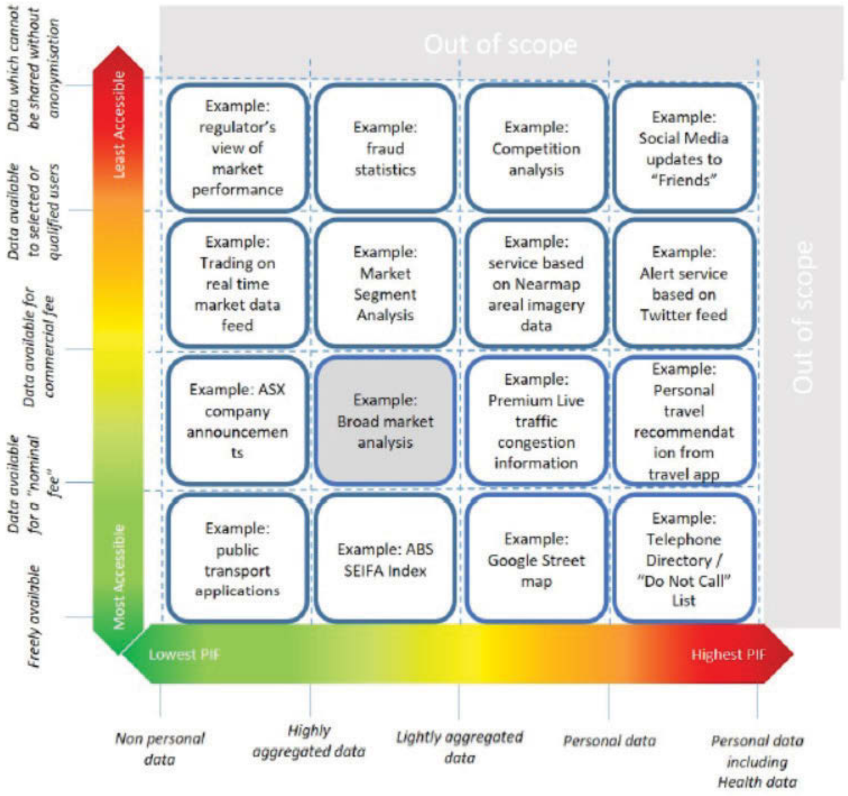


Figure 2.27.: Real world context (Oppermann, 2017)

In the following Figure 2.28 there is the Service Types described according to Personal Information Factor and Access. There are examples related with the concepts described.

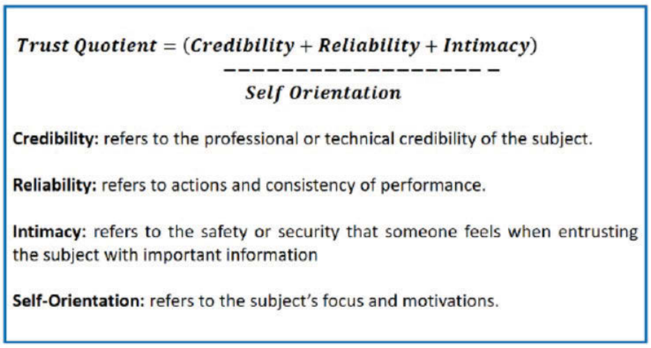

Data which cannot be shared without anonymisation
Data available to selected or qualified users
Data available for commercial fee
Data available for a "nominal fee"
Freely available
Least Accessible
Most Accessible
Out of scope
Out of scope
Lowest PIF
Highest PIF
Non personal data
Highly aggregated data
Lightly aggregated data
Personal data
Personal data including Health data

Example: regulator's view of market performance	Example: fraud statistics	Example: Competition analysis	Example: Social Media updates to "Friends"
Example: Trading on real time market data feed	Example: Market Segment Analysis	Example: service based on Nearmap areal imagery data	Example: Alert service based on Twitter feed
Example: ASX company announcemen ts	Example: Broad market analysis	Example: Premium Live traffic congestion information	Example: Personal travel recommendat ion from travel app
Example: public transport applications	Example: ABS SEIFA Index	Example: Google Street map	Example: Telephone Directory / "Do Not Call" List

Figure 2.28.: Services Types (Oppermann, 2017)

Personal Information Factor and access control generates the two dimensions model. The 'value' is a many factored issue associated with business impact, cost, and consequences of loss of exclusivity. The 'accessible' axis uses cost as a proxy for value with highly accessible data being considered low value, data which would be supplied for a nominal or commercial fee reflecting higher commercial value, and data which can only be shared under restricted conditions being highest value for which monetary compensation is insufficient consideration.

A heuristic model of trust was developed to describe the major components of trust and how the challenges of developing a trusted relationship could be addressed. The Trust equation described uses four objective variables to measure trustworthiness best described as: Credibility, Reliability, Intimacy, and Self-Orientation. The Trust equation provides a framework for potential interventions to improve the effectiveness of an engagement between individuals, or between an individual and an organisation.



$$\mathbf{\textit{Trust Quotient}} = \frac{(\mathbf{\textit{Credibility}} + \mathbf{\textit{Reliability}} + \mathbf{\textit{Intimacy}})}{\mathbf{\textit{Self Orientation}}}$$

Credibility: refers to the professional or technical credibility of the subject.

Reliability: refers to actions and consistency of performance.

Intimacy: refers to the safety or security that someone feels when entrusting the subject with important information

Self-Orientation: refers to the subject's focus and motivations.

Figure 2.29.: The trust equation (Oppermann, 2017)

2.10. Digital Preservation Initiatives

A comprehensive review of several important Digital Preservation initiatives was provided by Lee et al. (2002). These initiatives involve representative projects, define their specific preservation strategies and explore use of digital preservation technologies.

The initial phases of listed below projects, involved developing, evaluating and implementing digital preservation strategies in various european countries, the United States and Australia.

Australian Projects

Australia has been examining digital preservation issues since 1994. Several projects are aiming to preserve various digital materials such as electronic records, online publications, digital audio resources, theses and cartographic materials. The Victorian Electronic Records Strategy (VERS) project produced a standard for management and preservation of electronic records stated by Quenault (2004). The standard proposed by the VERS project recommends encapsulating the documents and their context in a single object based on XML. The Preserving and Accessing Networked Documentary Resources of Australia (PANDORA) project by the National Library of Australia has led the way in archiving the Web (Phillips 1999). The primary objective of PANDORA is to capture, archive, and provide long-term access to significant online publications. The project aims at addressing both archiving and preservation processes. The PANDORA archiving processes refer to the collection and provision of immediate access to the publications while preservation processes involve managing the materials and applying appropriate strategies (e.g., migration) to ensure long-term access.

The archiving processes have been developed while techniques for managing long-term access to these digital resources is still being developed. The project has also embarked on migration experiments with some HTML pages. The format of these pages is not yet obsolete, but the HTML specification has declared many mark-up tags as dead and not to be supported in this or future versions. The aim of these trials is to make changes in the HTML source code to remove tags declared as dead and replace them with current tags, effectively migrating the source code to a different version to reduce problems of future compatibility with Web browsers. Additionally, the National Archives of Australia is undertaking a project to develop advice for commonwealth agencies on using migration as a preservation strategy for electronic records.

CAMiLEON

The CAMiLEON project is funded by the Joint Information Systems Committee (JISC) in the UK and the National Science Foundation (NSF) in the USA. As reported by Higgins (2011), the project presents guidelines for use of emulation and argues that emulation is a valid method for both complex digital resources that include executable files and resources for which the documentation is not available in electronic form.

CEDARS

The Consortium of University Research Libraries, which represents both university and national libraries across the UK and Ireland, leads the CEDARS project. The project is based on the OAIS model described earlier by Russell (2000). Essentially it is an archival model for an archiving system but does not explicitly include a preservation module. They suggest that both migration and emulation strategies are viable for different types of digital materials. They also believe that no strategy is a panacea, and the strategy adopted for providing access to preserved resources will very much depend on the nature of the resource itself and the reason for its preservation. More work is planned to investigate the information loss associated with each strategy.

Canadian Projects

E-preservation was developed through a cooperative effort between the National Library of Canada and the Canadian Initiative on Digital Libraries (CIDL)(Hodges & Lunau 1999). E-preservation is intended: to provide Canadians with easy access to policies and to perform research on the creation, use, and preservation of digital collections. The project includes guidelines about various aspects including acquiring digital materials, formats, and metadata.

InterPARES

The International Research on Permanent Authentic Records Electronic Systems (InterPARES) project is a multinational research initiative, in which archival scholars, computer engineering scholars, national archival institutions, and private industry representatives are collaborating to develop the theoretical knowledge and methodology required for the permanent preservation of authentic records created using electronic systems (Duranti & Thibodeau 2006). Specifically, the research areas are divided into four complementary domains: authenticity, appraisal, preservation, and strategies. In terms of authenticity, the purpose is to identify the specific elements of electronic records that must be preserved, over time and across technologies, in order to verify the record's authenticity. As a first step, a template to guide the analysis of electronic records was developed and is being evaluated. The project is based in the School of Library, Archival and Information Studies at the University of British Columbia in Canada.

Kulturarw Heritage

As reported by Muir (2004), the Kulturarw Heritage Project of the Royal Library in Sweden explores methods of collecting, archiving, and providing access to Swedish electronic documents. Web crawlers or robots are used in order to collect all the Swedish Web pages automatically. Although the project currently does not focus on preservation, it is growing into a broader Nordic initiative that may explore the long-term preservation of this archive.

Library of Congress

The Computer Science and Telecommunications Board (CSTB) of the National Academies convened the Committee on the Information Technology Strategy for the Library of Congress for advice on digital preservation. The Library of Congress' pilot project, working with the Internet Archive, has worked through all aspects of archiving the Web in the area of political Web sites. Saracevic & Dalbello (2001) report that this project uses the Digital Library SunSITE Collection and Preservation Policy from the University of California, Berkeley, which provides several digital collecting levels, as guidance.

NARA: Persistent Archives and Electronic Records Management

The NARA (National Archives and Records Administration) project, which is led by the San Diego Supercomputer Center and funded by NARA, aims to develop a persistent archive to support ingestion, archival storage, information discovery, and preservation of digital collections (Rajasekar et al. 2006). One of its premises is the importance of preserving the organization of digital collections simultaneously with the digital objects that comprise the collection. The goal is to preserve not only the original data, but also the context that permits the data to be interpreted. The project emphasizes the synergy that is achieved through the identification of the unique capabilities provided by each environment, and the construction of interoperability mechanisms for integrating these environments. According to this project, collection-based persistent archives are now feasible and can manage massive amounts of information based on XML.

NEDLIB

The NEDLIB project was initiated by a permanent standing committee of the Conference of European National Libraries (CENL) in 1998, with funding from the European Commission's Telematics Application Program. The National Library of the Netherlands leads the project. Beagrie (2003)states that the project aims to develop a common architectural framework and basic tools for building deposit systems for electronic publications. The project has also adopted the OAIS migration model described earlier. The project has taken a first step to test the technicalities of the preservation mechanisms by starting an emulation experiment mentioned. The fundamental idea of the work is to test whether emulating obsolete computer hardware on future systems could be used to ensure long-term access to digital publications. Rothenberg has performed the first phase of this experimental work. It involved developing a prototype experimental environment for trying out emulation-based preservation and using commercial emulation tools to provide an initial proof-of-concept. The experimental results indicate that emulation should work in principle, assuming that suitable emulators for obsolete computing platforms can be hosted on future platforms.

Preservation Projects at the National Institute of Standards and Technology (NIST)

The earliest work on data preservation at NIST can be traced back to the 1980s when Podio (1992)performed research on the lifetime measurement of compact discs. His work provided a basis for a standard methodology for the lifetime measurement of optical discs. With the increasing usage of digital storage in libraries and the archiving of the government agencies, the great importance of digital preservation became clear to NIST's Information Technology Laboratory and accordingly new projects on the study of digital data preservation have started in the following aspects:

- Longevity testing. This project initially consisted of an examination of the effects of heat and humidity on the lifetime of optical discs and was later extended to include the effects of light exposure. The focus is not only on the lifetime itself, but also on the deterioration process. The results may be useful both for new disc production and for the classification of existing recorded discs.

- Testing of interchangeability and interoperability of optical discs for use in high-density storage systems such as optical disc "Jukeboxes." Combined with the application and further development of XML, a new preservation strategy may be developed. This program is being conducted in collaboration with the High Density Storage Association (HDSA). An open testing laboratory is being developed that will include interoperability and interchangeability testing as well as testing of the suitability of various types of high capacity storage systems for different applications including preservation.

- Development of the Turbo coding system. Unlike the traditional digital preservation techniques that aim to keep the information readable in the long term, this new technique aims to develop a method for finding and recovering useful information from failed discs.

PRISM

The Preservation, Reliability, Interoperability, Security, Metadata (PRISM) project of the Cornell University is a 4 year project funded by the Digital Library Initiative to investigate and develop policies and mechanisms needed for information integrity in digital libraries (Park 2001). The project focuses on long-term survivability of digital information, reliability of information resources and services, interoperability, security, and metadata. The current direction of the project is toward developing techniques for monitoring the integrity of distributed Web-based information resources and enforcing preservation policies set by the owners and users of collections. Monitoring resources will involve both the automated capture of information using a specialized Web crawler and the manual gathering of data on the organizational status of particular resources and collections. The ultimate objective is to develop a cost-effective and event-based metadata scheme that will enable users to define preservation policies and enforce them automatically.

3. Methodological Perspective Data Preservation

3.1. Research Methodologies

The following depicts the two major philosophical traditions, their respective assumptions, and the terminology associated with them. The first assumption listed in Figure 2, ontology, relates to the nature of reality, that is, what things, if any, have existence or whether reality is 'the product of one's mind' (Burrell and Morgan, 1979).

As explained later, the researcher's view of reality is the corner stone to all other assumptions, that is, what is assumed here predicates the researcher's other assumptions. The second assumption, epistemology, concerns the study of the nature of knowledge, that is, 'How is it possible, if it is, for us to gain knowledge of the world?' (Hughes and Sharrock, 1997). It is concerned with 'the nature, validity, and limits of inquiry' (Rosenau, 1992).

Much of the research that has been completed in organisational science has assumed that reality is objective and out there 'waiting to be discovered and that this knowledge can be identified and communicated to others. The third assumption, concerning human nature, involves if the researcher perceives man as the controller or as the controlled (Burrell and Morgan, 1979), and the final assumption, methodology, is the researcher's tool-kit – it represents all the means available to social scientists to investigate phenomena (Holden and Lynch, 2004).

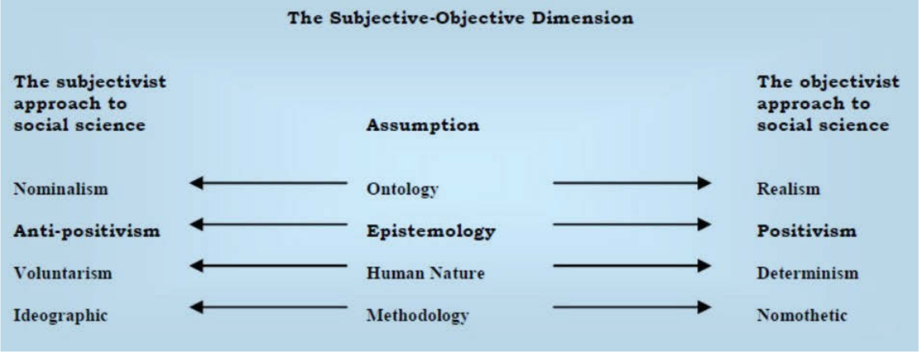
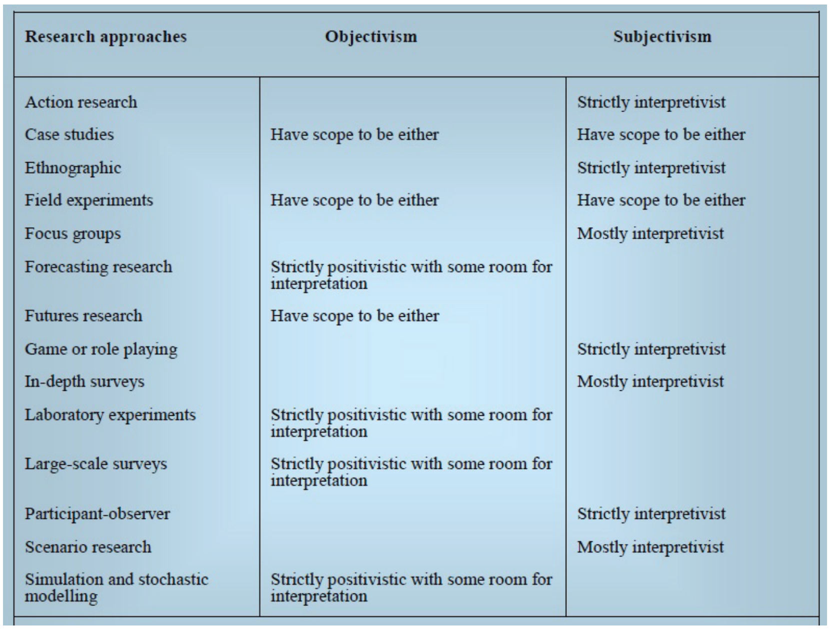


Figure 3.1.: Scheme for Analysing Assumptions. Source: Burrell and Morgan (1979)



Research approaches	Objectivism	Subjectivism
Action research		Strictly interpretivist
Case studies	Have scope to be either	Have scope to be either
Ethnographic		Strictly interpretivist
Field experiments	Have scope to be either	Have scope to be either
Focus groups		Mostly interpretivist
Forecasting research	Strictly positivistic with some room for interpretation	
Futures research	Have scope to be either	
Game or role playing		Strictly interpretivist
In-depth surveys		Mostly interpretivist
Laboratory experiments	Strictly positivistic with some room for interpretation	
Large-scale surveys	Strictly positivistic with some room for interpretation	
Participant-observer		Strictly interpretivist
Scenario research		Mostly interpretivist
Simulation and stochastic modelling	Strictly positivistic with some room for interpretation	

Figure 3.2.: Research Tactics and Philosophical Bases. Source: Remenyi et al. (1998)

The Digital Curation Centre is committed to advancing knowledge in digital preservation. As digital preservation is a young discipline many of its methods are untested. As reported by Jones et al. (2009) the testbed methodology developed by DCC researchers will allow practitioners to validate preservation approaches, thereby ensuring digital assets remain usable well into the future.

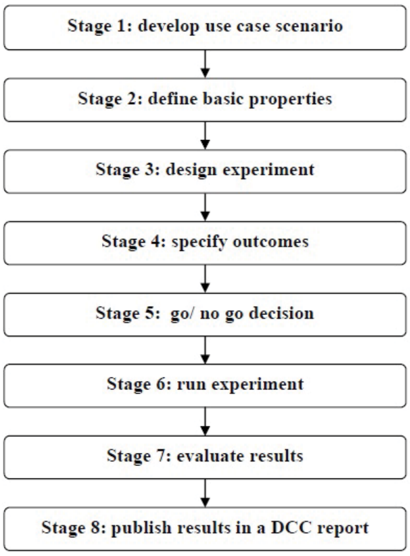

Stage 1: develop use case scenario
Stage 2: define basic properties
Stage 3: design experiment
Stage 4: specify outcomes
Stage 5: go/ no go decision
Stage 6: run experiment
Stage 7: evaluate results
Stage 8: publish results in a DCC report

Figure 3.3.: The eight stage DCC testbed methodology (Jones, 2009)

Stage 1: develop use case scenario

The use case will be used to validate the success of the preservation action in stage 7. As such it will need to identify how the resource is currently used to ensure user requirements, e.g. the ability to perform full text searches, are still met. A use case will minimally consist of:

1. what digital resource is being used (what)
2. for what purpose is it being used (why)
3. in what way is it being used (how)
4. by whom (e.g. researcher, student) is it being used (who)
5. within which designated community is it being used (where)
6. when (e.g. daily, regularly, one time) is it being used (when)

Stage 2: define basic properties

The premise of this stage is to provide an overview of the proposed experiment. Basic details should be recorded, such as an experiment name, description, purpose and focus. The scope of the experiment should be set by noting any key considerations or research questions and recording parameters defined by the use case. An experiment for the text mining community, for example, may restrict the test to focus on textual records and discount any image formats. Links to relevant research or contextual literature can also be recorded at this stage.

Stage 3: design experiment

The experiment design provides the framework for the experiment and method for running it. A decision will be made as to what type of experiment to run (e.g. characterisation / format identification, migration etc), and the number and type of digital objects being input for testing. The emphasis at this stage will be on establishing the practical issues involved in running the experiment. Contextual details are set in related stages: stage 2 addresses the scope and purpose of the experiment, while stage 4 defines the expected outcome and criteria for evaluation.

Stage 4: specify outcomes

This stage will determine the success criteria for the experiment. These could be based on the use case, organisational objectives, collective knowledge of the curation community, or other such factors. The criteria noted will act as a key input during evaluation. A number of quality levels and characteristics are already provided within the Planets testbed for various types of digital object, enabling the experimenter to specify exactly which aspects are most crucial to maintain in a given context. For example, the bit depth and appropriate resolution may be paramount for an image migration from png to jpeg. Metrics to evaluate these criteria will be developed.

Stage 5: go/ no go decision

In this stage the Planets testbed will automatically consider the experiment design and parameters to determine if it is feasible to proceed. The result should be recorded along with an explanation or record of changes required if the experiment could not go ahead or was postponed.

Patterns for software development are: prototype, spiral and waterfall. The option considered to analyse DHP is the top-down methodology, because it is for dynamic systems.

Stage 6: run experiment

The Planets testbed will run the experiment according to the inputs and parameters identified in stages 2 & 3. The experiment will test one or more aspects of applying a preservation approach to a defined set of objects. Running the experiment will produce preserved digital objects and an assessment of how they differ technically from the input. This can then be evaluated in the next stage.

Stage 7: evaluate results

The results of the experiment will be evaluated to determine how successfully the requirements were met. Validation is achieved by comparing the data submitted with the preserved object, the output after the preservation action was performed. There will be two main stages to the evaluation:

1. a technical assessment of how well the preservation action was performed based on criteria recorded within the Planets testbed environment;

2. a qualitative assessment achieved by implementing the use case.

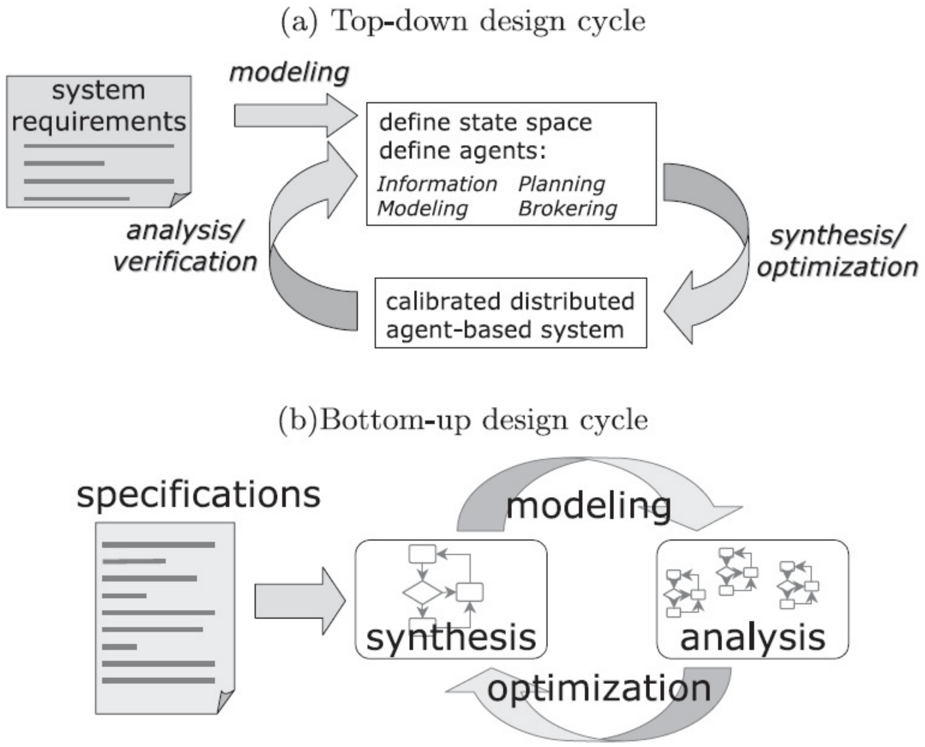
In this secondary evaluation checks will be made to ensure the preserved object can continue to perform its function as stated by the use case. In the case of an online journal, for example, the required functions may be indexing and full text searches to ensure retrieval. As such, significant character corruption would render the document void. Both forms of evaluation will be done with help of metrics developed in stage 4 when expected outcomes were specified.

Stage 8: publish the results in a DCC report

By publishing results with the DCC we allow others to re-run our experiments and learn from our experiences. Building up a body of knowledge in this way is crucial to advancing knowledge in the field of digital preservation.

Methodology Bottom-up and Top-down: Top-down and bottom-up approaches have similarities in the interpretation in context of Preservation. These statements around the methodologies relate to the following Figure 3.4.

Top Down: This methodology has been developed in control theory. The steps include: model, synthesize, analyses and optimisation Sabatier (1986).



(a) Top-down design cycle

(b)Bottom-up design cycle

Figure 3.4.: Conceptual representation Multi agent

3.2. The Open Group Architecture Framework (TOGAF)

Using the Open Group Architectural Framework (TOGAF) it is found to be more suitable as a result of dissemination of the data. Meanwhile, it is associated with an Ontology and Knowledge Management terms to be more specific and with a deep sense of definition. The experimentation explained in the next chapters will cover concepts, for instance:

BP Engineering Tools , OLTP , BPM Tools, GISDB, 3D Web as a tool, Methods, Embedded Visualization, SOM, DNA Computing, SOTA , DHP

3.3. Approach and Methodology

The methodology has diverse levels and different dimensions. The principal classification is divided in: General Investigation, Data Analysis, Contribution and Evaluation.

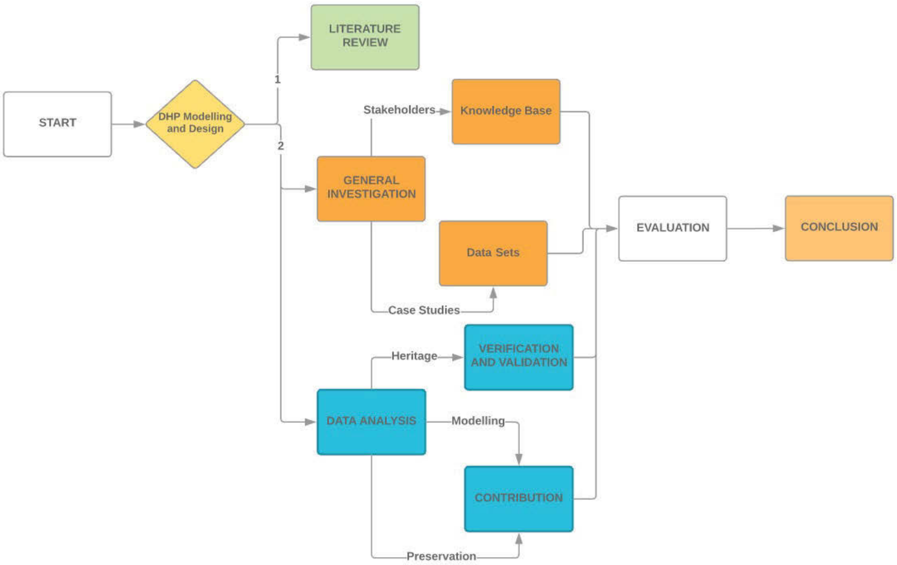


Figure 3.5.: Methodology Flowchart

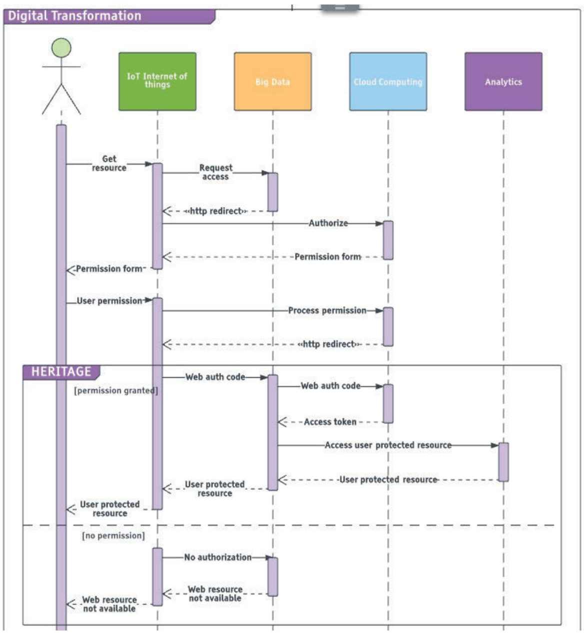


Figure 3.6.: Digital Transformation

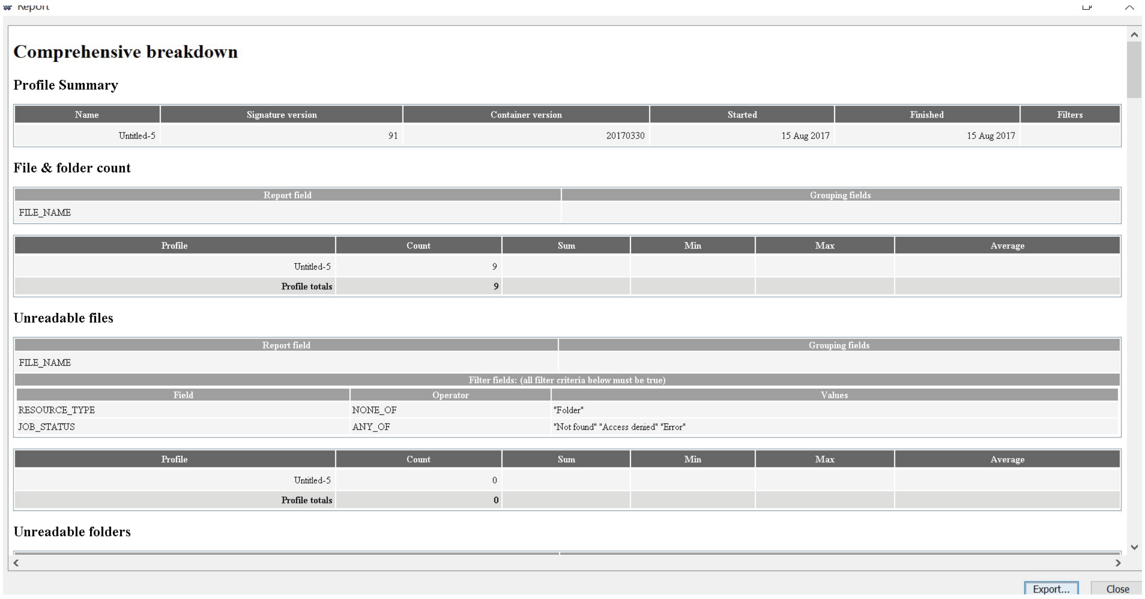
Methodology based on Internet of Things: The methodology mentioned in Figure 3.5 of DHP is well connected with the definitions of IoT. The Principal Stakeholders are related to technology and the conceptualization around this environment. There are four fields related mentioned in the following Figure 3.6. These are: IoT, Big Data, Cloud Computing and Analytics.

Pointed out by Hegarty et al. (2014) the IoT defines a place where ordinary devices are identifiable with unique ID, to address and to contact online. Practical Description: The first approach with the experimentation and practical description is with technology of DROID (Digital Record and Object Identification).

3.4. Characterisation Software

According to Brown et al. (2009), there is importance in the approach of characterisation software to support digital objects. Technical metadata will assist in the preservation of digital objects for electronic records.

PRONOM / DROID: This software accomplishes with the characteristics according to the requirements of Digital Preservation, Brown et al. (2009). DROID stands for Digital Record Object Identification.



Comprehensive breakdown

Profile Summary

Name	Signature version	Container version	Started	Finished	Filters
Untitled-5	91	20170330	15 Aug 2017	15 Aug 2017	

File & folder count

Report field	Grouping fields
FILE_NAME	

Profile	Count	Sum	Min	Max	Average
Untitled-5	9				
Profile totals	**9**				

Unreadable files

Report field	Grouping fields
FILE_NAME	

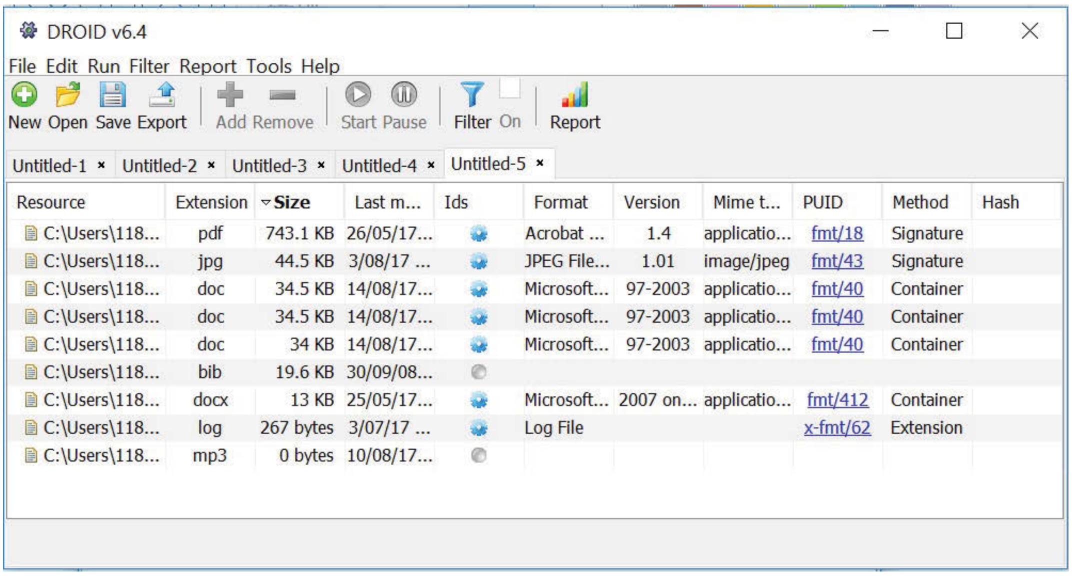
Filter fields: (all filter criteria below must be true)

Field	Operator	Values
RESOURCE_TYPE	NONE_OF	"Folder"
JOB_STATUS	ANY_OF	"Not found" "Access denied" "Error"

Profile	Count	Sum	Min	Max	Average
Untitled-5	0				
Profile totals	**0**				

Unreadable folders

Figure 3.7.: Droid Summary Record



Resource	Extension	Size	Last m...	Ids	Format	Version	Mime t...	PUID	Method	Hash
C:\Users\118...	pdf	743.1 KB	26/05/17...	⚙	Acrobat ...	1.4	applicatio...	fmt/18	Signature	
C:\Users\118...	jpg	44.5 KB	3/08/17 ...	⚙	JPEG File...	1.01	image/jpeg	fmt/43	Signature	
C:\Users\118...	doc	34.5 KB	14/08/17...	⚙	Microsoft...	97-2003	applicatio...	fmt/40	Container	
C:\Users\118...	doc	34.5 KB	14/08/17...	⚙	Microsoft...	97-2003	applicatio...	fmt/40	Container	
C:\Users\118...	doc	34 KB	14/08/17...	⚙	Microsoft...	97-2003	applicatio...	fmt/40	Container	
C:\Users\118...	bib	19.6 KB	30/09/08...	●						
C:\Users\118...	docx	13 KB	25/05/17...	⚙	Microsoft...	2007 on...	applicatio...	fmt/412	Container	
C:\Users\118...	log	267 bytes	3/07/17 ...	⚙	Log File			x-fmt/62	Extension	
C:\Users\118...	mp3	0 bytes	10/08/17...	●						

Figure 3.8.: Droid Screen

3.5. Modelling Theory

The areas of knowledge can be represented through modeling techniques. There are functionalities for processing. Stated in Fill (2016) for instance: Enterprise Architecture Management, Business Process, Performance Management, Workflow Management.

Methodology Approach

SeMFIS Meta Models are different from the meta models used for the knowledge area to deal with machine processable semantics for models of the knowledge area in order to allow for the machine processing of their semantics Bork et al. (2017)Frank (2011). The framework aims to eliminate a shortcoming of other approaches in the context of semantic enrichment of conceptual models: As the approach is also based on meta models, a coherent approach for dealing with meta models on the different levels of information systems as well as the machine processable semantics can be realized. Concerning the later implementation of the meta models, this allows to re-use the same functionality and mechanisms for conceptual and SeMFIS models, in regard to the visualization of the models, for the provision of analysis and search functionalities or for the storage and change management of the models.

Furthermore, the users of the framework are not required to switch between different environments and methodologies for modeling and the handling of semantics.

SeMFIS Tools and Services: SeMFIS Support Tools and Services shall be used to integrate specific tools and services for dealing with semantics into the SeMFIS framework. This concerns in particular tools for creating, manipulating, processing, and storing ontologies. By providing linkages to the other components of the SeMFIS framework bi-directional interaction of the SeMFIS components and the support tools and services shall be established. Thereby the available knowledge in the form of these tools and services can be integrated in the framework as well as the knowledge contained in the SeMFIS framework be made available to third parties. Furthermore, the re-use of existing tools and services allows to achieve results in a shorter time.

SeMFIS Mechanisms and Algorithms: SeMFIS Mechanisms and Algorithms shall be designed for the application to SeMFIS Meta Models and meta models of the knowledge area for the purpose of semantic information processing. As the goal of the framework is to provide answers in regard to the machine processing of semantic information in conceptual models, this component is of critical importance. It shall allow to analyze end evaluate the semantic information in conceptual models and provide functionality that is specific to the SeMFIS framework and is not available using support tools or services.

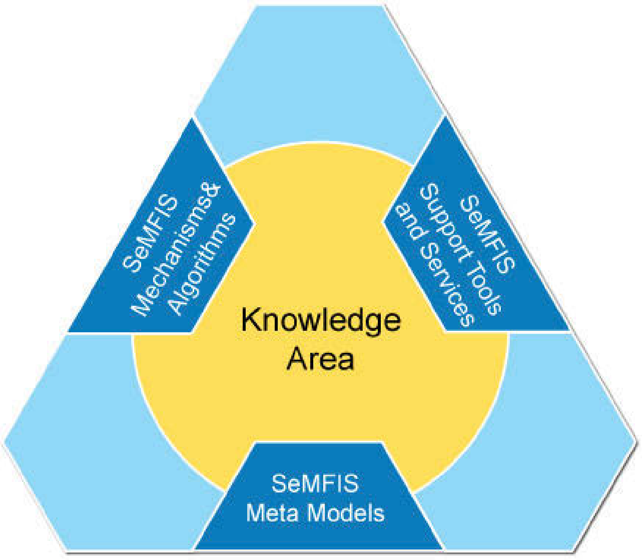

SeMFIS Mechanisms& Algorithms
SeMFIS Support Tools and Services
Knowledge Area
SeMFIS Meta Models

Figure 3.9.: Conception of SeMFIS Framework. Source: (Fill, 2016)

3.6. Self Organising Maps (SOM)

Self-organising neural networks are used to cluster input patterns into groups of similar patterns. They're called 'maps' because they assume a topological structure among their cluster units; effectively mapping weights to input data. The Kohonen network is probably the best example, because it's simple, yet introduces the concepts of self-organization and unsupervised learning easily Kohonen (1989).

Each weight is representative of a certain input. Input patterns are shown to all neurons simultaneously.

Self-organising networks can be either supervised or unsupervised. Unsupervised learning is a means of modifying the weights of a neural network without specifying the desired output for any input patterns. The advantage is that it allows the network to find its own solution, making it more efficient with pattern association. The disadvantage is that other programs or users have to figure out how to interpret the output Kohonen (1989).

The structure of a self-organising map involves m cluster units, arranged in either a one- or two-dimensional array, with vectors of n input signals Fausett (1994).

The Figure 3.10 below expresses the Example self-organising network with five cluster units, Yi, and seven input units, Xi. The five cluster units are arranged in a linear array

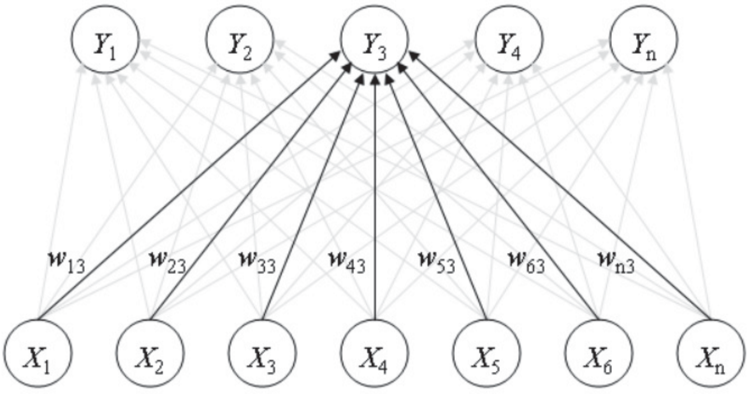


Figure 3.10.: Example self-organising network. Source: (Kohonen 1989)

The weight vectors define each cluster. Input patterns are compared to each cluster, and associated with the cluster it best matches. The comparison is usually based on the square of the minimum Euclidean distance. When a best match is found, the associated cluster gets its weights and its neighboring units updated.

Weight vectors are arranged into lines or various grid structures. Some neighborhoods closer to the ends or edges will have fewer weights, because the algorithm doesn't wrap around. The Figure 3.11 below states Neighborhoods (R) for a linear array of cluster units: R = 0 in black brackets, R = 1 in red, and R = 2 in blue. Also, the Neighborhoods (R) for a rectangular matrix of cluster units: R = 0 in black brackets, R = 1 in red, and R = 2 in blue expresses in the Figure 3.12 below. Other way is Neighborhoods (R) for a hexagonal matrix of cluster units: R = 0 in black brackets, R = 1 in red, and R = 2 in blue in the Figure 3.13.

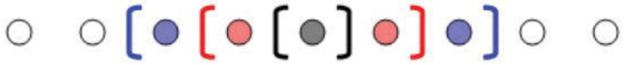


Figure 3.11.: Linear array of cluster. Source: (Kohonen 1989)

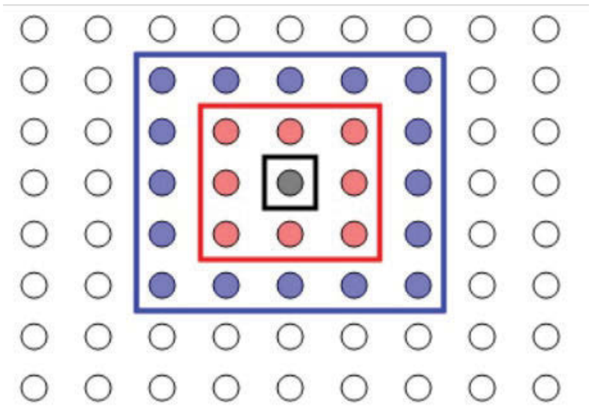


Figure 3.12.: Rectangular matrix of cluster. Source: (Kohonen 1989)

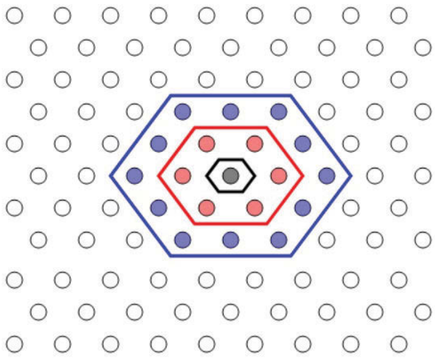


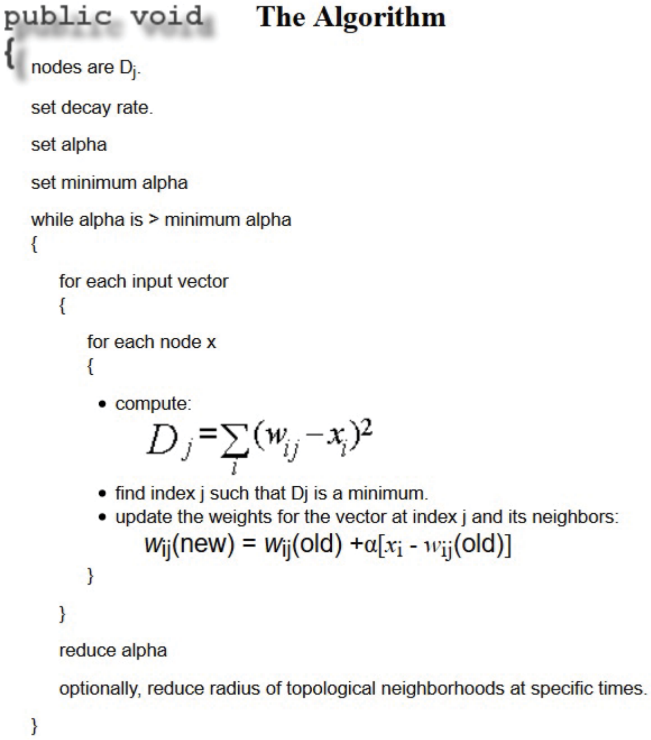
Figure 3.13.: Hexagonal matrix of cluster. Source: (Kohonen 1989)

The learning rate α alpha is a slowly decreased with each epoch. The size or radius of the neighborhood around a cluster unit can also decrease during the later epochs.

The formation of a map occurs in two stages:

1. The initial formation of the correct order
2. The final convergence

The second stage takes much longer, and usually occurs when the learning rate gets smaller. The initial weights can be random values.



```
public void        The Algorithm
{
 nodes are $D_j$.
 set decay rate.
 set alpha
 set minimum alpha
 while alpha is > minimum alpha
 {
     for each input vector
     {
         for each node x
         {
           • compute:
                 $D_j = \sum_i (w_{ij} - x_i)^2$
           • find index j such that Dj is a minimum.
           • update the weights for the vector at index j and its neighbors:
                 $w_{ij}(new) = w_{ij}(old) + \alpha[x_i - w_{ij}(old)]$
         }
     }
     reduce alpha
     optionally, reduce radius of topological neighborhoods at specific times.
 }
```

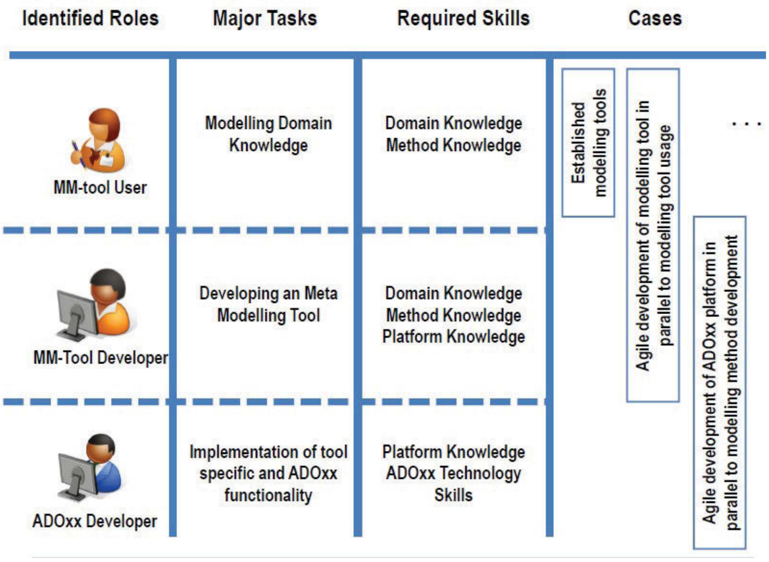
Figure 3.14.: SOM Algorithm

With the WEBSOM method a textual document collection may be organized onto a graphical map display that provides an overview of the collection and facilitates interactive browsing. Interesting documents can be located on the map using a content-directed search. Each document is encoded as a histogram of word categories which are formed by the self-organising map (SOM) algorithm based on the similarities in the contexts of the words. The encoded documents are organised on another self-organising map, a document map, on which nearby locations contain similar documents. Special consideration is given to the computation of very large document maps which is possible with general-purpose computers if the dimensionality of the word category histograms is first reduced with a random mapping method and if computationally efficient algorithms are used in computing the SOMs.

3.7. ADOxx Modelling Method

The Open Models (OMi) Laboratory is a dedicated research and experimentation space for modelling method engineering. OMi Laboratory makes an open use of the ADOxx meta-modelling platform.

A growing number of groups develop individual modelling-methods, in addition to existing standard ones for a variety of application domains. To support domain-specific modelling, dedicated tool-functionality like query, simulation or transformation is provided in order to process models in a domain-specific manner. Today's meta-modelling platforms provide capabilities for developing modelling tools based on domain specific modelling languages (DSML).



Identified Roles	Major Tasks	Required Skills
MM-tool User	Modelling Domain Knowledge	Domain Knowledge Method Knowledge
MM-Tool Developer	Developing an Meta Modelling Tool	Domain Knowledge Method Knowledge Platform Knowledge
ADOxx Developer	Implementation of tool specific and ADOxx functionality	Platform Knowledge ADOxx Technology Skills

Cases
Established modelling tools
Agile development of modelling tool in parallel to modelling tool usage
Agile development of ADOxx platform in parallel to modelling method development
. . .

Figure 3.15.: ADOxx platform. Source: Fill (2016)

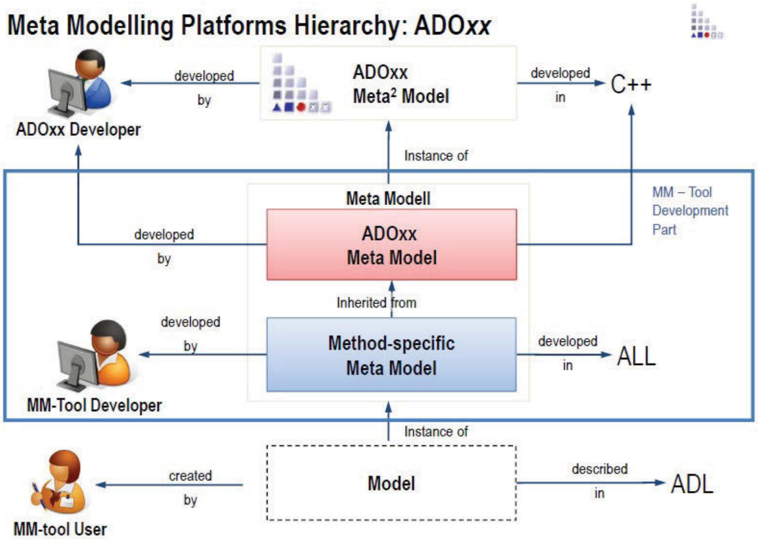
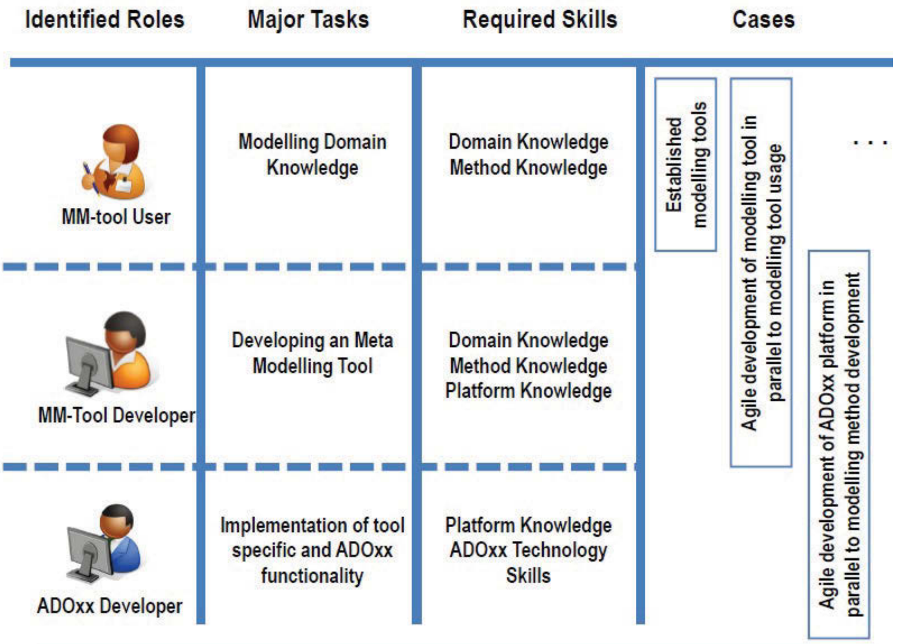


Figure 3.16.: ADOxx Modelling. Source: Fill (2016)



Identified Roles	Major Tasks	Required Skills	Cases
MM-tool User	Modelling Domain Knowledge	Domain Knowledge Method Knowledge	Established modelling tools Agile development of modelling tool in parallel to modelling tool usage . . .
MM-Tool Developer	Developing an Meta Modelling Tool	Domain Knowledge Method Knowledge Platform Knowledge	Agile development of ADOxx platform in parallel to modelling method development
ADOxx Developer	Implementation of tool specific and ADOxx functionality	Platform Knowledge ADOxx Technology Skills	

Figure 3.17.: ADOxx platform Source: Fill (2016)

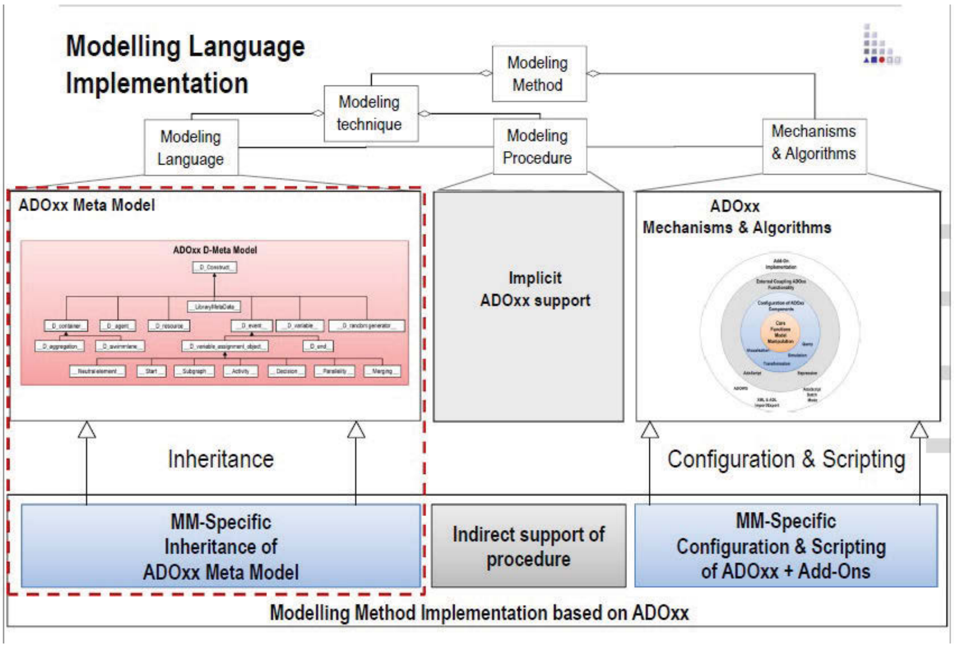

Modelling Language Implementation
Modeling Method
Modeling technique
Modeling Language
Modeling Procedure
Mechanisms & Algorithms
ADOxx Meta Model
ADOxx D-Meta Model
Implicit ADOxx support
ADOxx Mechanisms & Algorithms
Inheritance
Configuration & Scripting
MM-Specific Inheritance of ADOxx Meta Model
Indirect support of procedure
MM-Specific Configuration & Scripting of ADOxx + Add-Ons
Modelling Method Implementation based on ADOxx

Figure 3.18.: ADOxx Implementation. Source: Fill (2016)

The reason why we chose the ADOxx meta modelling approach and the corresponding software platform is, that it has been successfully developed, used, and tested for over more than fifteen years in a large number of research and industrial projects that included some of the largest German and Austrian companies as customers. It is therefore an industry-proven approach that goes far beyond a research prototype in terms of functionalities, scalability, and reliability. In this way we will be able to derive insights into the conceptualisation of modelling methods that are also relevant from an industry perspective Fill & Karagiannis (2013).

To clarify our understanding of the terms and elements of modelling methods and thus provide a solid basis for the further discussion, we revert to a framework that has originally been proposed by Karagiannis and Kühn in 2002 see Figure 3.18. In this framework a modelling method is composed of a modelling technique and mechanisms and algorithms. Thereby the modelling technique is further divided into a modelling language and a modelling procedure. The modelling procedure consists of steps for defining the application of the modelling language and delivers results. For this, it reverts to mechanisms and algorithms. The modelling language has a syntax that defines the grammar and semantics that defines the meaning of the elements of the syntax. This is achieved by a semantic mapping that connects the syntactical constructs with their meaning defined in a semantic schema. The semantic schema may either be formally defined or may come in the form of (informal) textual descriptions as it is used, e.g., in the definition of the UML or BPMN (Object Management Group OMG 2007, 2011a) Lorie (2001).

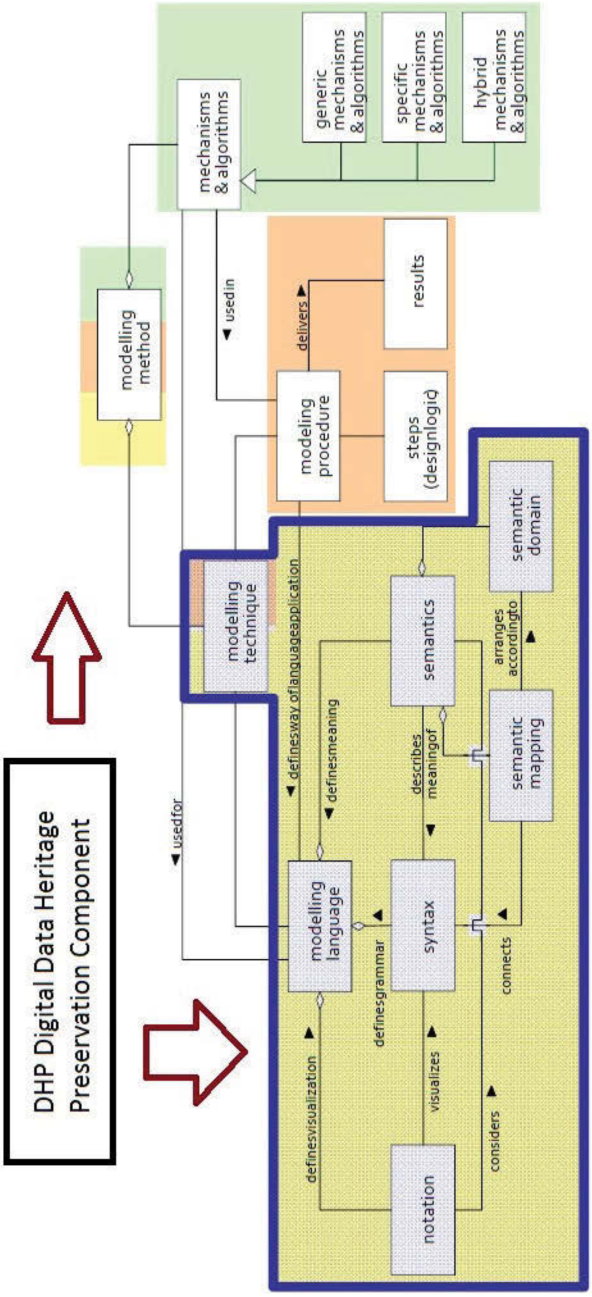


Figure 3.19.: ADOxx Framework. Adapted from: Fill (2016)

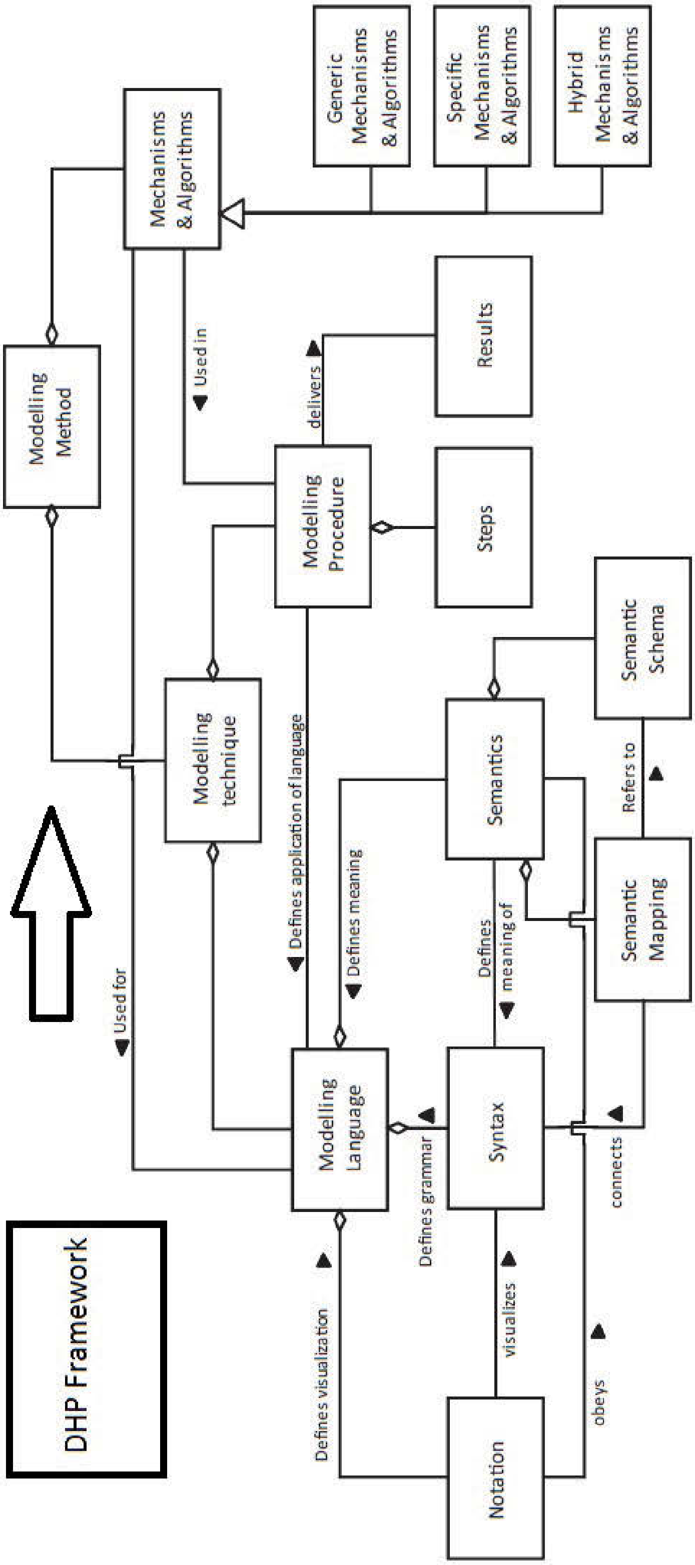


Figure 3.20.: ADOxx Modelling Methods. Source: Fill (2016)

Part II.

Experimental Work of Digital Data Heritage Preservation (DHP)

4. DHP Modelling and Design

Data modeling is the process of documenting a complex software system design as an easily understood diagram, using text and symbols to represent the way data needs to flow. The diagram can be used as a blueprint for the construction of new software or for re-engineering a legacy application.

The process of designing a candidate system aims to solve the defined gap between Digital Data and Heritage Preservation. It considers as a plan for the designing of the DHP solution. Designing the candidate system involves taking the output of the literature review which identify the purpose and the specification of the research approach. The candidate system aims to develop a plan for the selected solution. The design includes low-level component and approaches, implementation encounters and designing architectural model. A new digital preservation framework has been developed to achieve the aims. Preservation strategies and the use of software tools for migration and emulation should be chosen according to business requirements. Applicability of both strategies are context dependent. It is related with the complexity of the analysis. Once the alternative and the preservation paths are specified, the experiments could be proved. Then the analysis of the utility has to be integrated in inhomogeneous criteria, like cases of study for evaluating different strategies.

The accessibility and the categorisation of filtering information is related with the state of evaluation of data set is based from three components: Maturity Matrix, Activity level and Stakeholder, suggested by Weaver et al. (2008). It seems to be a good application of clustering and how to classify. A further motivation is also to provide the convenient way to search the information.

There are two considerations in this part:

- Physical (access)
- Cognitive (contextualization).

Other issue that we have to consider is the management of the artefacts in terms of preservation. Nowadays the requirements of useful storage should be enough in terms of technology, but the specific classification of the information with criteria without losing the original knowledge is the real challenge in the intelligent and efficient preservation. The objective for understanding the perspective throughout the development process has been for building the basement concepts and knowledge.

User requirements includes reliable and authentic data to generate knowledge information. Framework, Methodology and Patterns are important steps to follow to generate DHP metamodel.

Basically, Case Studies and Applications are to understand throughout the development process the basic concepts for the implementation.

The users and custodians require integrating two basic concepts in terms of preservation: authenticity and reliability.

For the designing of the model there are previous steps to develop. For the organisation we will use SIPOC as it is explained in the Figure 4.1.

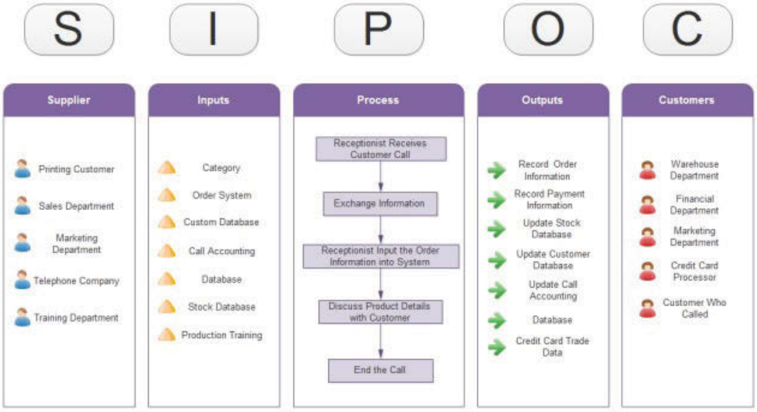


Figure 4.1.: SIPOC (Mishra, 2014)

4.1. Modelling DHP

The basic steps used for model-building are the same across all modeling methods. The details vary somewhat from method to method, but an understanding of the common steps, combined with the typical underlying assumptions needed for the analysis, provides a framework in which the results from almost any method can be interpreted and understood. The basic steps of the model-building process are: model selection, model fitting, and model validation. These three basic steps are used iteratively until an appropriate model for the data has been developed. In the model selection step, plots of the data, process knowledge and assumptions about the process are used to determine the form of the model to be fit to the data. Then, using the selected model and possibly information about the data, an appropriate model-fitting method is used to estimate the unknown parameters in the model. When the parameter estimates have been made, the model is then carefully assessed to see if the underlying assumptions of the analysis appear plausible. If the assumptions seem valid, the model can be used to answer the scientific or engineering questions that prompted the modeling effort. If the model validation identifies problems with the current model, however, then the modeling process is repeated using information from the model validation step to select and/or fit an improved model.

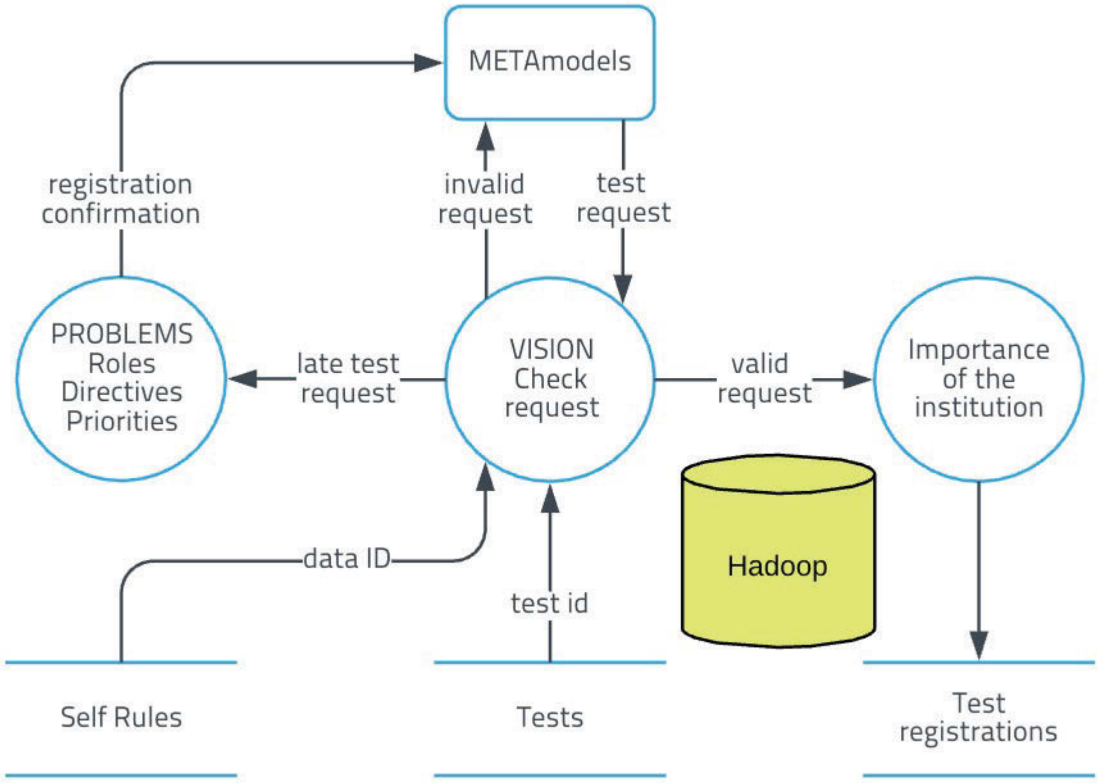


Figure 4.2.: DHP Vision

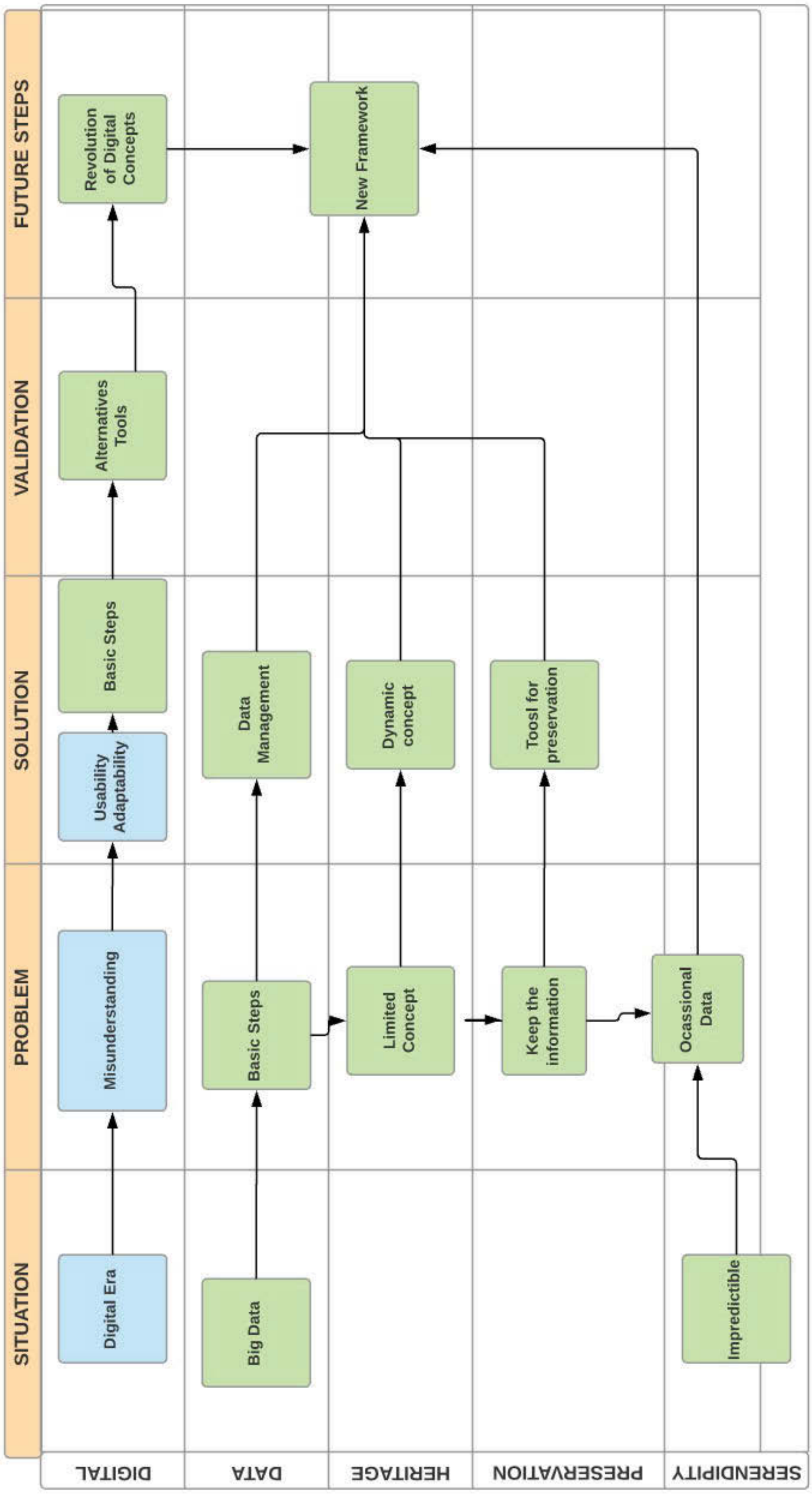


Figure 4.3.: DHP Flow Solution

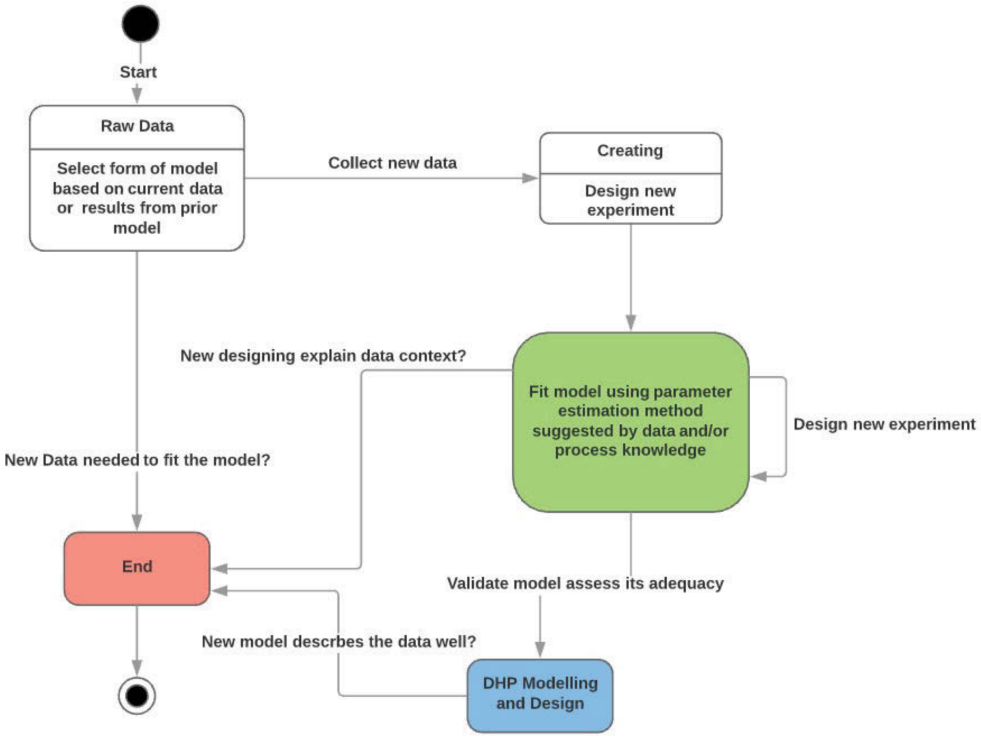

Start
Raw Data
Select form of model based on current data or results from prior model
Collect new data
Creating
Design new experiment
New designing explain data context?
Fit model using parameter estimation method suggested by data and/or process knowledge
Design new experiment
New Data needed to fit the model?
End
Validate model assess its adequacy
New model descrbes the data well?
DHP Modelling and Design

Figure 4.4.: Flowchart Modelling

4.1.1. Data Collection for Process Modeling

Collecting Good Data

It lays out some general principles for collecting data for construction of process models. Using well-planned data collection procedures is often the difference between successful and unsuccessful experiments. In addition, well-designed experiments are often less expensive than those that are less well thought-out, regardless of overall success or failure. Specifically, this section will answer the question: What can the analyst do even prior to collecting the data (that is, at the experimental design stage) that would allow the analyst to do an optimal job of modeling the process?

It deals with the following five questions:

1. What is design of experiments (DOE)?

2. Why is experimental design important for process modeling?

3. What are some general design principles for process modeling?

4. I've heard some people refer to 'optimal' designs, shouldn't I use those?

5. How can I tell if a particular experimental design is good for my application?

What is design of experiments (DOE)?

4.1.2. Systematic Approach to Data Collection

Design of experiments (DOE) is a systematic, rigorous approach to engineering problem-solving that applies principles and techniques at the data collection stage so as to ensure the generation of valid, defensible, and supportable engineering conclusions. In addition, all of this is carried out under the constraint of a minimal expenditure of engineering runs, time, and money.

DOE Problem Areas

There are four general engineering problem areas in which DOE may be applied:

1. Comparative
2. Screening/Characterizing
3. Modeling
4. Optimizing

Comparative

In the first case, the engineer is interested in assessing whether a change in a single factor has in fact resulted in a change/improvement to the process as a whole.

Screening Characterization

In the second case, the engineer is interested in 'understanding' the process as a whole in the sense that he/she wishes (after design and analysis) to have in hand a ranked list of important through unimportant factors (most important to least important) that affect the process.

Modeling

In the third case, the engineer is interested in functionally modeling the process with the output being a good-fitting (= high predictive power) mathematical function, and to have good (= maximal accuracy) estimates of the coefficients in that function.

Optimizing

In the fourth case, the engineer is interested in determining optimal settings of the process factors; that is, to determine for each factor the level of the factor that optimizes the process response. In this section, we focus on case 3: modeling.

4.1.3. Underlying Assumptions for Process Modeling

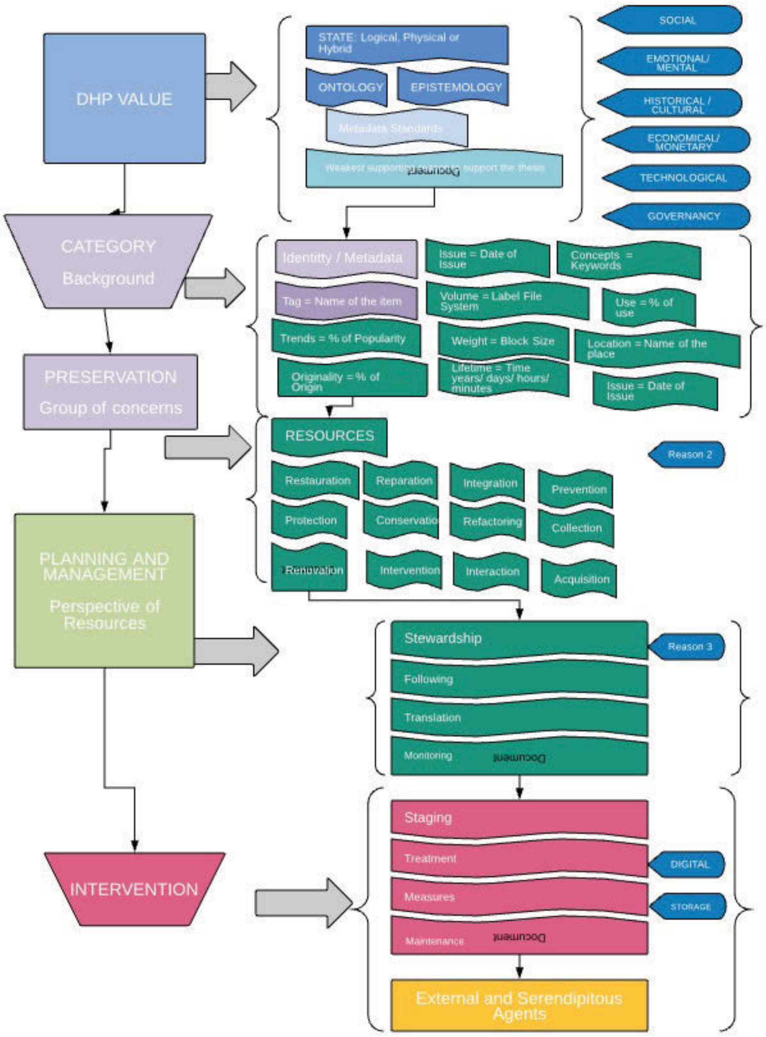


Figure 4.5.: Modelling Steps

Most, if not all, thoughtful actions that people take are based on ideas, or assumptions, about how those actions will affect the goals they want to achieve. The actual assumptions used to decide on a particular course of action are rarely laid out explicitly, however. Instead, they are only implied by the nature of the action itself. Implicit assumptions are inherent to process modeling actions, just as they are to most other types of action. It is important to understand what the implicit assumptions are for any process modeling method because the validity of these assumptions affect whether or not the goals of the analysis will be met.

If the implicit assumptions that underlie a particular action are not true, then that action is not likely to meet expectations either. Sometimes it is abundantly clear when a goal has been met, but unfortunately that is not always the case. In particular, it is usually not possible to obtain immediate feedback on the attainment of goals in most process modeling applications. The goals of process modeling, such as answering a scientific or engineering question, depend on the correctness of a process model, which can often only be directly and absolutely determined over time. In lieu of immediate, direct feedback, however, indirect information on the effectiveness of a process modeling analysis can be obtained by checking the validity of the underlying assumptions. Confirming that the underlying assumptions are valid helps ensure that the methods of analysis were appropriate and that the results will be consistent with the goals.

The proposal for the modelling is multidimensional. It is represented in 5+2 model. There are 5 dimensions called as following: Privacy, Rights, Integrity, Security, Trust. Recognised as PRIST model Challa et al. (2005), Chaczko et al. (2014)

To complete the proposal, the additional contribution, include two more parts (CP) Cognitive (Authenticity) and Physical (Access) dimensions. The following formulas (4.1 and 4.2) expressed the conjunction between Model A and Model B as a 7 dimensional model. The new model formulas can be expressed as follows:

$$M_A + M_B = 7PC \tag{4.1}$$

$$PRIST + CP = DIADM - (DIGITAL - AGE - DATA - MODEL) \tag{4.2}$$

Cognitive Dimension: Identifying as knowledge, meaning, situation, understandability, context, interpretability, cataloguing, **Physical Dimension:** Access control, Historical Retrieval, Digital Rights Management, extraction, interface, storage,

Processes: Techniques, Approach.

When representing data in Privacy, Rights, Integrity, Security and Trust dimensions, the proposed model needs to contextualize the interpretation, conservation and preservation of the natural and cultural heritage of the real world artifacts. Thus, the model needs to incorporate physical and cognitive dimensions to form a unified model we can be called the Digital Age of Data Model or simply DiADM.

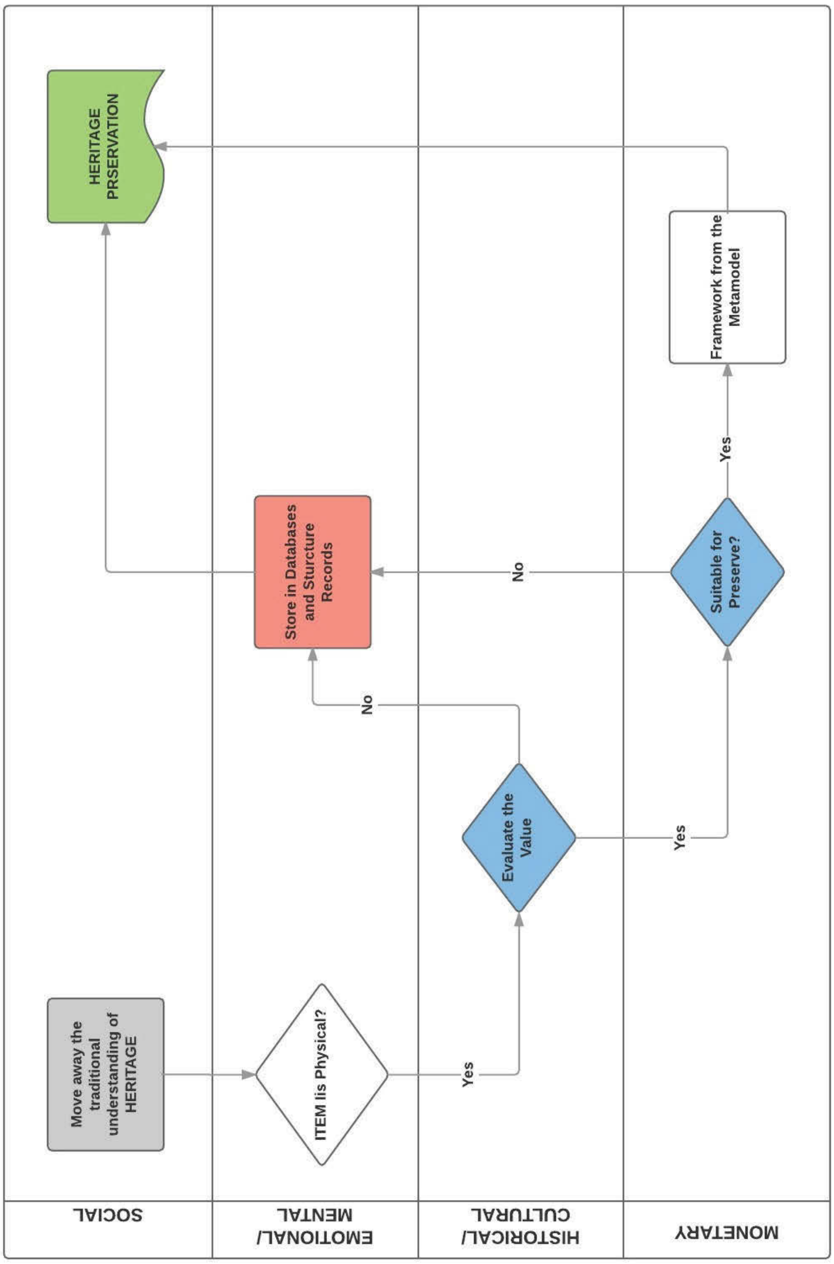


Figure 4.6.: MetaModel Steps

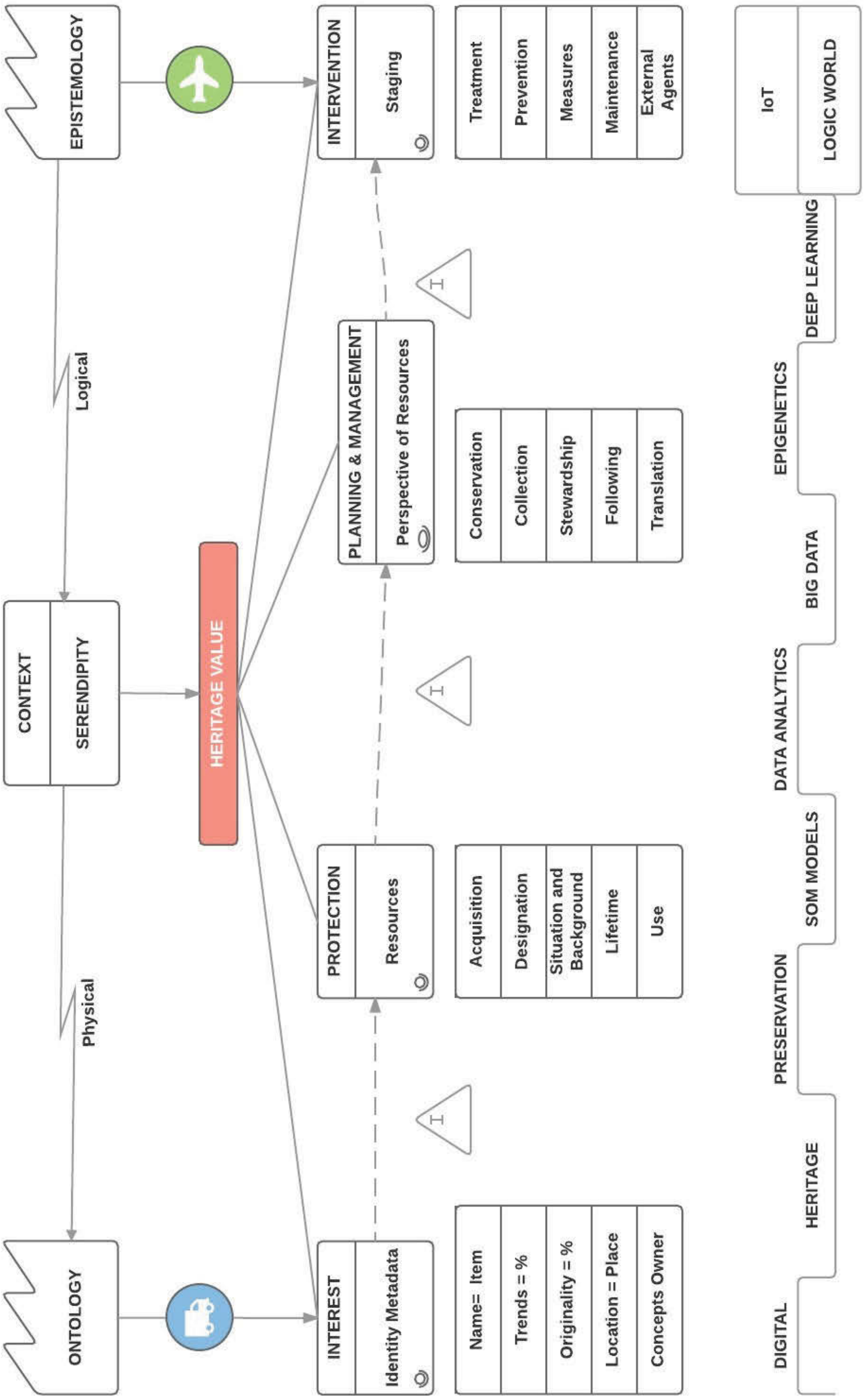


Figure 4.7.: MetaModel Steps

DHP Modelling Metaheuristic: There are many strategies to keep these solutions. Assisted Data Heritage Preservation is one of the proposals. Based on different aspects, one of the options is to use concepts as Domain Language (Karagiannis, Buchmann, Burzynski, Reimer & Walch 2016).

A meta-model should specify the following: All concepts (modelling classes) of a modelling language, all relationships between modelling classes, Cardinalities of the relationships, Semantic names for the relationships like 'interact_with' etc., Maybe inheritance relationships between concepts, Maybe important attributes of the concepts (take care of ensuring readibility as the meta models can get quiet complex with attributes).

The reference to the paper mentioned on designing domain-specific modeling languages: Frank (2011). 'Optimization knowledge is extended by learning about the domain' Wagner et al. (2010). HEURISTIC LAB most of the times includes representation of the code Wagner & Kronberger (2012). The Process Constraints Stemming from Heterogeneous Sources Imposed on Current PAIS mentioned in the Figure 4.8.

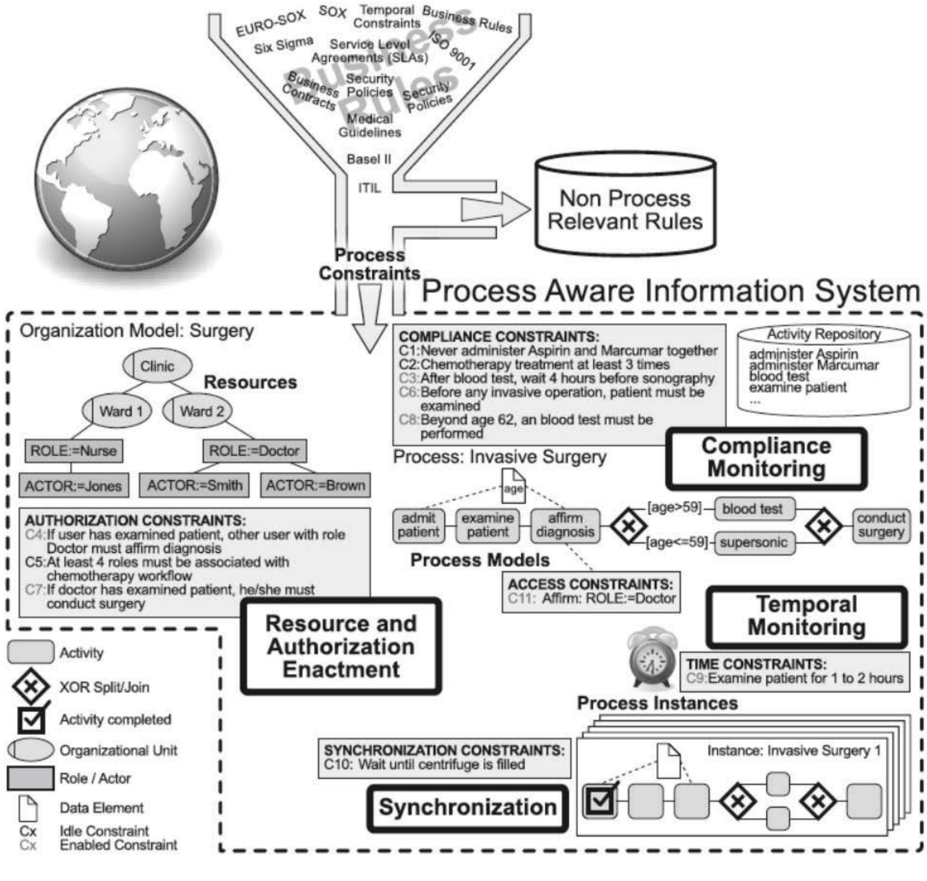

EURO-SOX
SOX
Temporal Constraints
Business Rules
Six Sigma
Service Level Agreements (SLAs)
ISO 9001
Business Rules
Business Contracts
Security Policies
Security Policies
Medical Guidelines
Basel II
ITIL
Non Process Relevant Rules
Process Constraints
Process Aware Information System
Organization Model: Surgery
Clinic
Resources
Ward 1
Ward 2
ROLE:=Nurse
ROLE:=Doctor
ACTOR:=Jones
ACTOR:=Smith
ACTOR:=Brown
AUTHORIZATION CONSTRAINTS:
C4:If user has examined patient, other user with role Doctor must affirm diagnosis
C5:At least 4 roles must be associated with chemotherapy workflow
C7:If doctor has examined patient, he/she must conduct surgery
COMPLIANCE CONSTRAINTS:
C1:Never administer Aspirin and Marcumar together
C2:Chemotherapy treatment at least 3 times
C3:After blood test, wait 4 hours before sonography
C6:Before any invasive operation, patient must be examined
C8:Beyond age 62, an blood test must be performed
Activity Repository
administer Aspirin
administer Marcumar
blood test
examine patient
...
Compliance Monitoring
Process: Invasive Surgery
age
admit patient
examine patient
affirm diagnosis
[age>59]
blood test
[age<=59]
supersonic
conduct surgery
Process Models
ACCESS CONSTRAINTS:
C11: Affirm: ROLE:=Doctor
Temporal Monitoring
Resource and Authorization Enactment
TIME CONSTRAINTS:
C9:Examine patient for 1 to 2 hours
Process Instances
SYNCHRONIZATION CONSTRAINTS:
C10: Wait until centrifuge is filled
Instance: Invasive Surgery 1
Synchronization
Activity
XOR Split/Join
Activity completed
Organizational Unit
Role / Actor
Data Element
Cx Idle Constraint
Cx Enabled Constraint

Figure 4.8.: Relationship in Digital Information Model. Source (Frank, 2011)

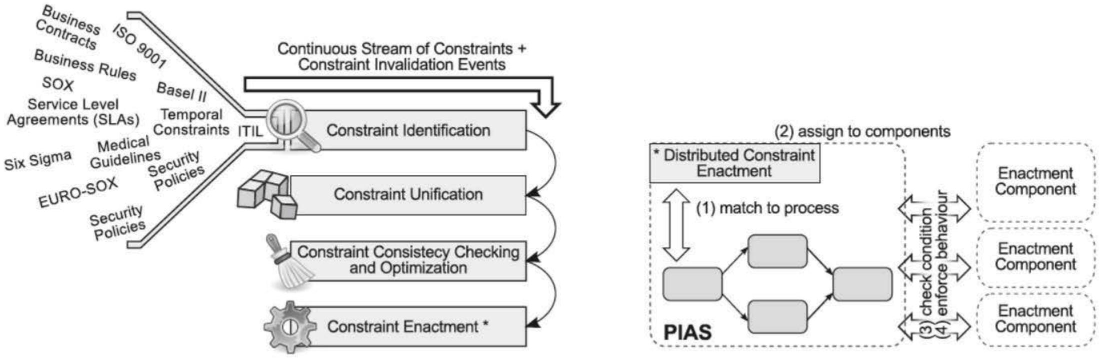
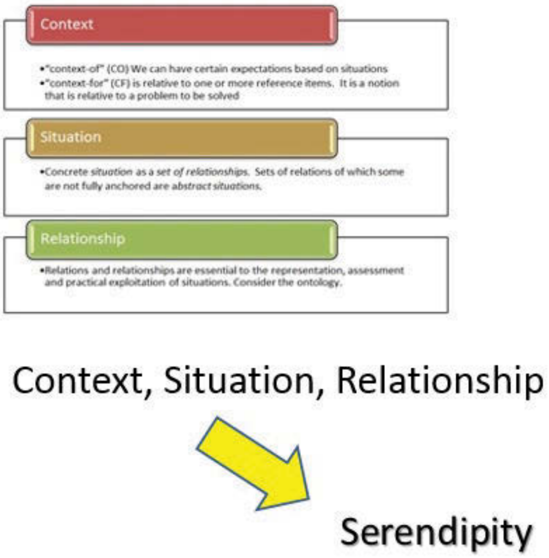


Figure 4.9.: Unified Constraint Optimization and Checking. Source (Frank, 2011)

4.2. Patterns for DHP



Context

- "context-of" (CO) We can have certain expectations based on situations
- "context-for" (CF) is relative to one or more reference items. It is a notion that is relative to a problem to be solved

Situation

- Concrete *situation* as a *set of relationships*. Sets of relations of which some are not fully anchored are *abstract situations*.

Relationship

- Relations and relationships are essential to the representation, assessment and practical exploitation of situations. Consider the ontology.

Context, Situation, Relationship

Serendipity

Figure 4.10.: Vision of Context

One of the primarily concerns about the explosive amount of information and the complexity of the classification, is how to keep the principal characteristics data. A need to move away the traditional understanding of Heritage reflects the real meaning of the data. More artifacts and everyday life tendency is to have less physical representation in the World of Logic. The representation of the items refers to the tendency of more things nonphysical and how through the Heritage it passes the attributes.

4.2.1. Ontology

The consideration mentioned in the publication Gordon et al. (2015) ontology is defined as the unique result for a digital age of the information. Ontology and Knowledge Management are used to analyse data. Suggested use of the interface and metadata generation, are the principal considerations to optimise data preservation Gordon & Chaczko (2017*a*).

Base for data and knowledge sharing, Ontologies are a common vocabulary to ensure that all have the same understanding of the terms used. There is Automation (Reasoning Tasks): Deriving implicit knowledge, Identification of inconsistencies, Determining concept hierarchies.

In computer science ontologies are formally well-ordered representations of a set of concepts and the relationships between them in a particular subject area. They are used to interchange "knowledge" digitally and formally between application programs and services. Ontologies include: the description and definition of terms (concepts) their properties, the relationships between them, limitations under which the relationships are valid, inference and integrity rules for conclusions and to ensure their validity and ultimately concrete individuals.

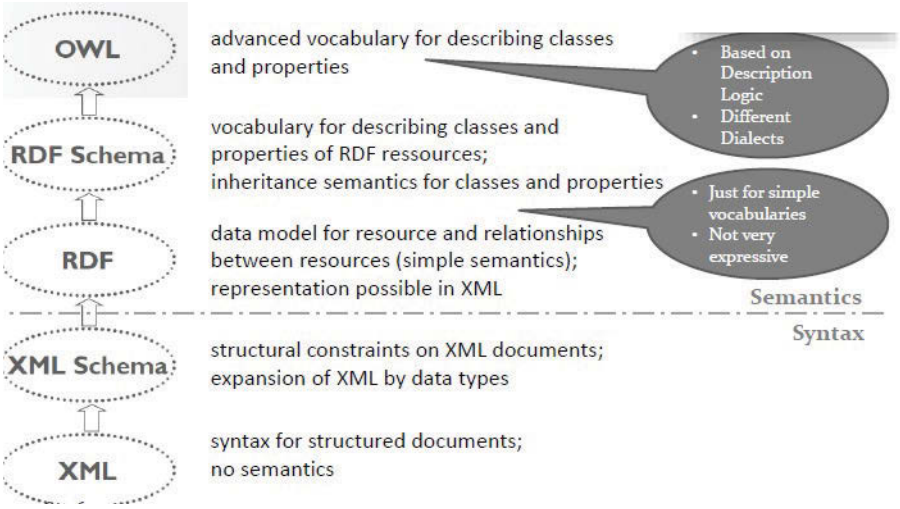

OWL
advanced vocabulary for describing classes and properties
• Based on Description Logic
• Different Dialects
RDF Schema
vocabulary for describing classes and properties of RDF ressources; inheritance semantics for classes and properties
• Just for simple vocabularies
• Not very expressive
RDF
data model for resource and relationships between resources (simple semantics); representation possible in XML
Semantics
Syntax
XML Schema
structural constraints on XML documents; expansion of XML by data types
XML
syntax for structured documents; no semantics

Figure 4.11.: Syntax and Semantics

Ontologies as distributed, machine-processable vocabularies for a specific domain.

Describe a detail of the world: What things are there? What is the relation of these things? What characteristics do these things and the relationships with each other have?

Ontology description = Syntax, Ontology rules = Semantics, Open-World vs. Closed World Assumption, Open World

Assumption: A reasoner does not assume that something does not exist, as long as it is not explicitly defined that it does not exist. It is assumed that the knowledge has not been added to the knowledge base yet.

This allows uncertainty and ambiguity. Not to know that a statement is true, does not mean that this is wrong. Be careful with restrictive conditions when reasoning (e.g.: max, exactly, only). In databases, typically the Closed World Assumption is applied.

4.2.2. Architectural Framework

The Architecture is one of the principal bases to build the Framework. Considering an approach for the designing of DHP, the proposed framework includes: Identification, Value , Context , Situation , Serendipity , Storage and Stakeholders.

Through detailed examination and focusing in the priorities of data treatment, the proposal of this component was based on critical stages and studies abour digital preservation in general. The final proposal addresses these steps.

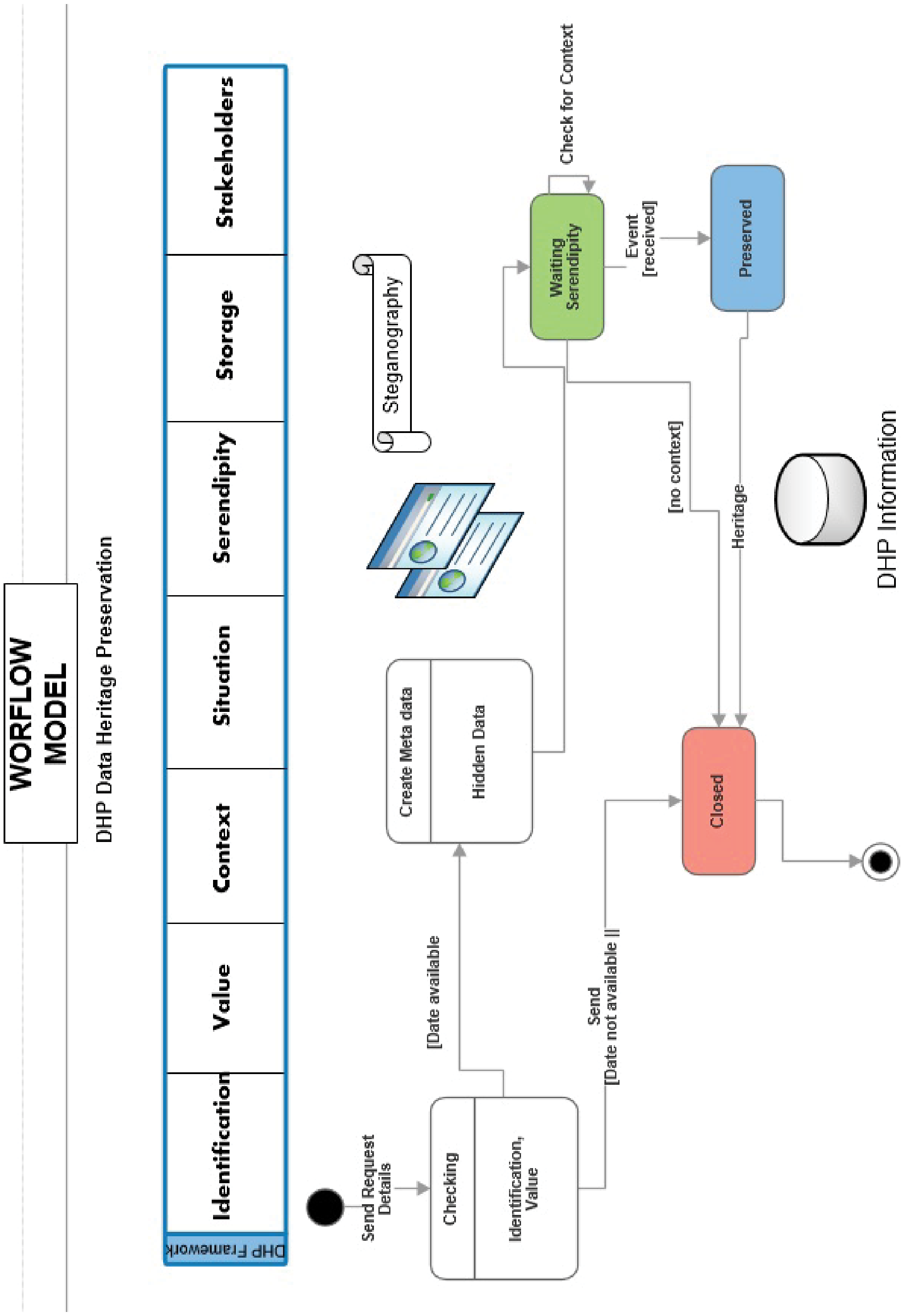


Figure 4.12.: DHP Workflow Model

4.2.3. Middleware of DHP Data Heritage Preservation

Digital preservation as strategy is crucial for the organisations. Digital preservation focuses on developing a strategy for long-term continuity of important digital items for example documents, images, websites, emails, and outputs from research data.

There are two tendencies around the understanding of the management of the ideas. The ontology and the Epistemology of this study, centralized the future use and the Serendipity tendency of the item. However, in the perspective of the nonphysical items there is a World of Physical and Logical and how the Preservation need to look items and how will be the manifestation. Digital Preservation has evolved into a specialised, interdisciplinary research. Through the time the challenge in to jointly develop solutions.. As the patterns and alternative solutions there are Information Retrieval and, Machine Learning or Software Engineering. The real fact of Digital Preservation shows us the reality of the understanding of the World about the facility to have digital expressions rather than just physical.

The Heritage of the collected information define the quality of the Data. At this stage, the definition of Heritage involved the presence not only the content. It is the express by itself the real meaning of the data. The perception of the importance and relevance of the information is measured through the definitions and metamodel that is proposed.

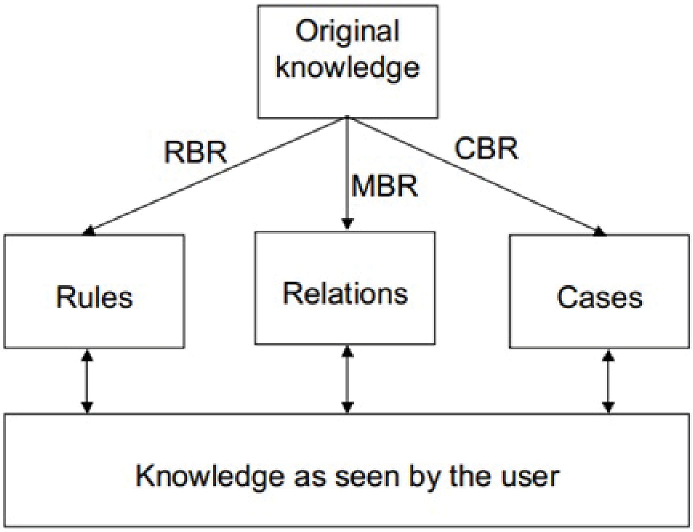


Figure 4.13.: General Approaches

4.3. DHP Metaclass

The concepts around DHP set the definitions on the relationships. Metaclasses are the way to define the development of the classes. Dividing in the three different phases, there are classes based on metada. In the following Figure 4.14 there is the explanation and the details of the proceses based on metadata.

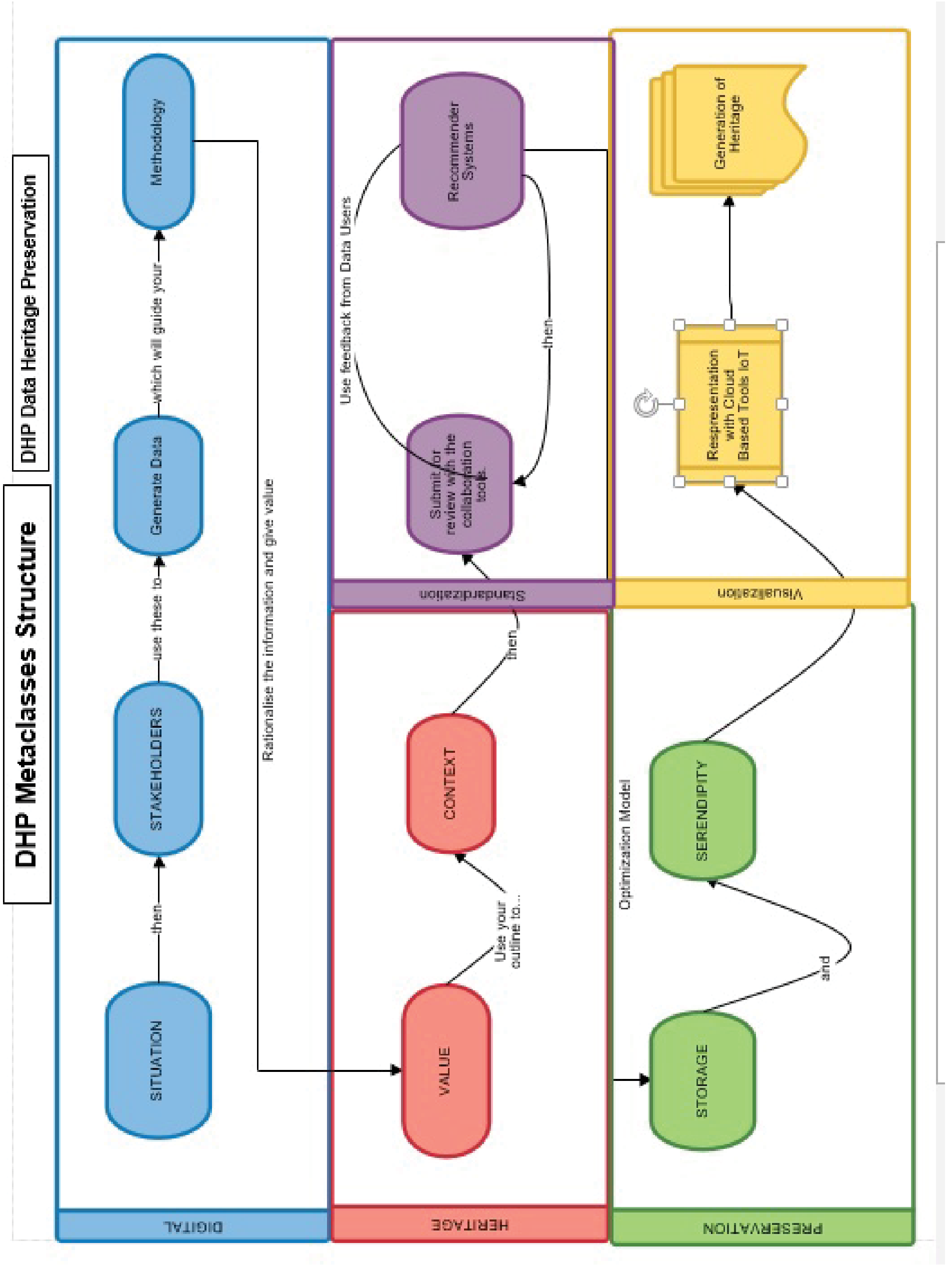


Figure 4.14.: Active Structures in DHP

The metamodel in the following Figure 4.15 explain the relationshops with the concepts and stakeholders. The three component and most important are: Software Patterns, Domain Analysis and Knowledge Maps. These concepts are considerig the methods and the techniques for the definition of DHP.

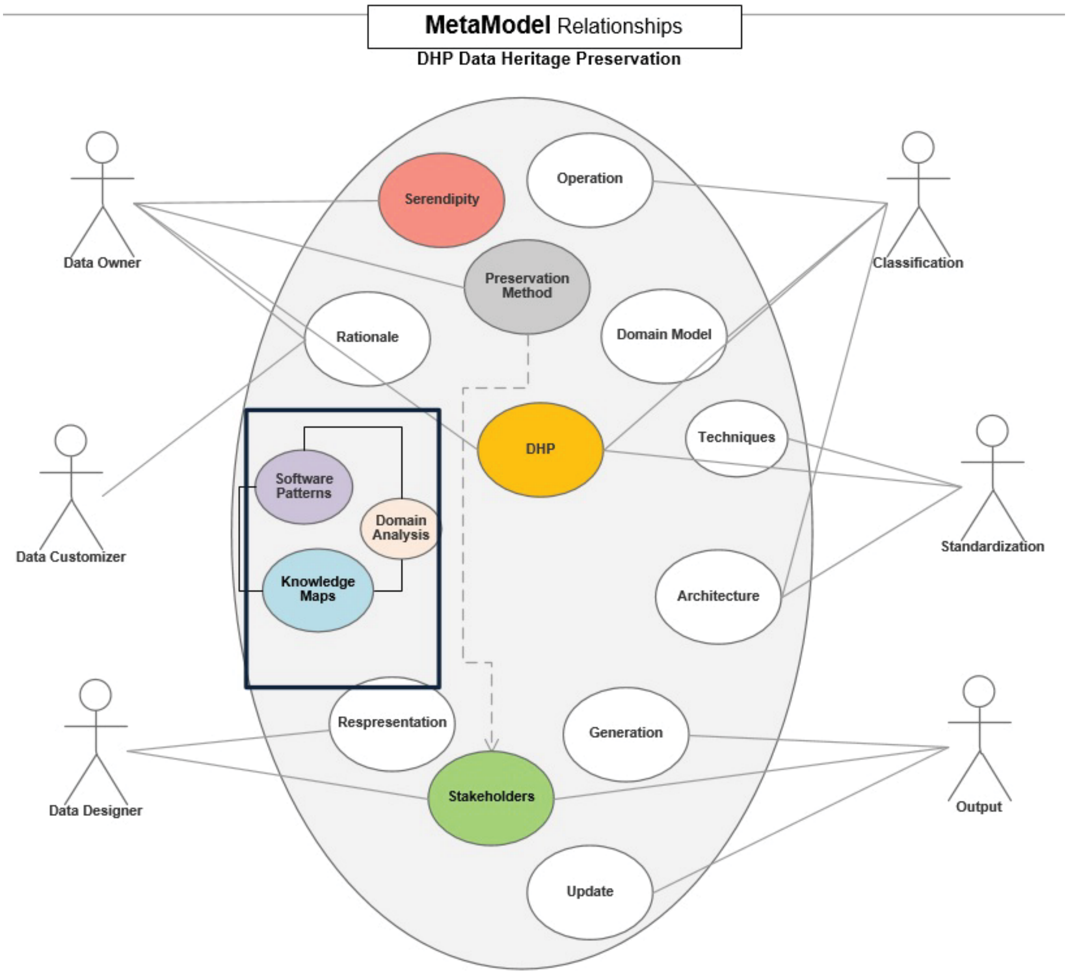


Figure 4.15.: Relationship in Digital Information Model

4.4. DHP Mapping and Processing

For the process, the Metamodel has to include the seven stages for the framework. The relationship between behavior and data treatment. The preservation of the information appropriates resources in digital context. The following Figure 4.16shows the steps, stages and level of the model. The processes are the most important in the definition of the fonal DHP framework.

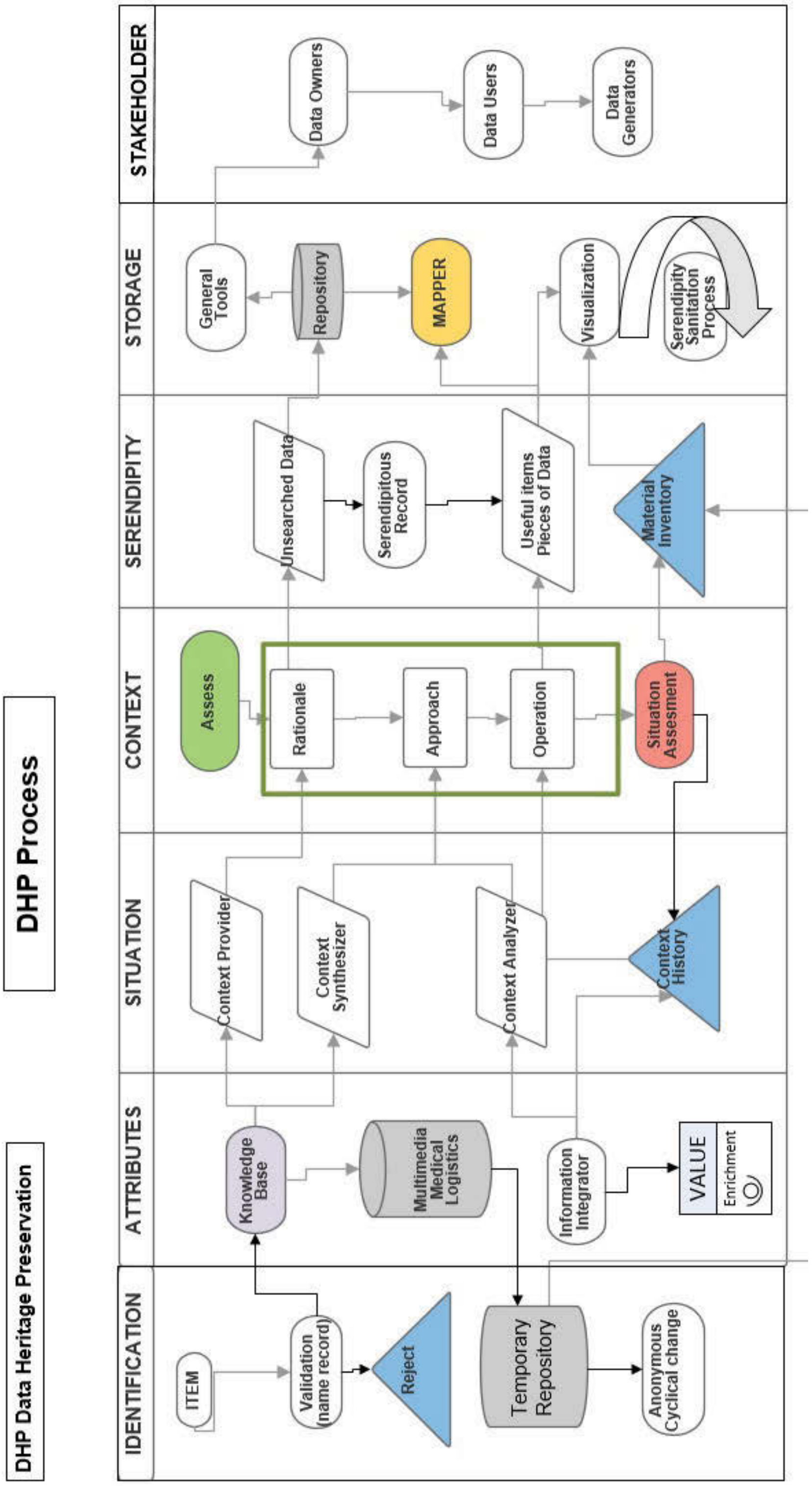


Figure 4.16.: DHP Process

4.5. DHP Metamodel

Explore Metadata Models for Dynamic Data Representation. Data Aging is one of the key for the definition of metamodel and shows the new relations and mechanism. For the experimentation, there are three different tools were explored and adopted. SOM (Self-Organising Maps): Representation and maps. It is an efficient tool for visualization of multidimensional numerical data. In this paper, an overview and categorization of both old and new methods for the visualization of SOM is presented. The purpose is to give an idea of what kind of information can be acquired from different presentations and how the SOM can best be utilized in exploratory data visualization. Most of the presented methods can also be applied in the more general case of first making a vector quantization (e.g. k-means) and then a vector projection (e.g. Sammon's mapping). SOM Toolbox also features other data analysis methods related to VQ, clustering, dimension reduction, and proximity preserving projections, e.g.,data preprocessing tools, K-means, K-nearest neighbor classifier and LVQ (learning vector quantizer), agglomerative hierarchical clustering and dendrograms, principal component analysis (PCA), Sammon's projection, and Curvilinear Component Analysis (CCA).

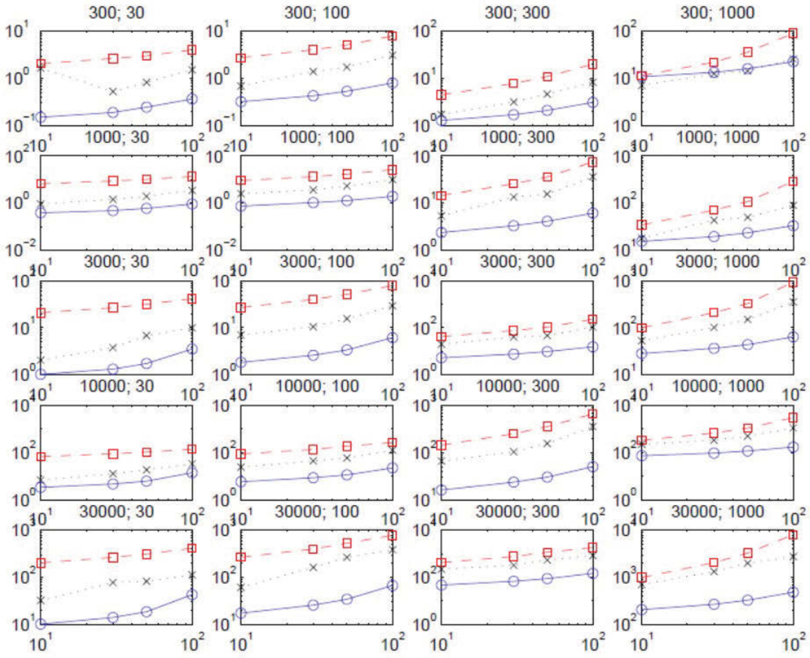


Figure 4.17.: SOM Toolbox (Vesanto, 2000)

Heuristic LAB: Through this topic there are specific concepts related with Heritage

Preservation. Digital Data and Historical Analytics. The relation is the IoT mainly cloud, because the Heritage Preservation as Data Analytics.

It works as input a table with all the columns are attributes and different features or object to classify. There are two features. The structure is they are connected the closest, the most influence. The steps are Select Random input, Compute Winner neuron, Update Neurons, Repeat for all input data and Classify input data.

These guaranteed along the iterations try to fit the model behind the data. It is related with Neural Networks. It is fast and powerful technique to solve many problems. It 'learn' from experience. It can deal with incomplete information or noisy data. It is useful in complex situation where it is not possible to define the rules or steps that lead to the solutions of the problems. We talk about: Learning Methods, Topology and Applications. There are correct approaches related with these definitions (Karagiannis & Kühn 2002).

4.5.1. Adaptation Blockchain Model

DHP model adapts the Blockchain models as a certification method of Digital Objects. Blockchain based solutions have recently appear as a new way to solve many of qualification process. Now, blockchain storage is generally considered expensive for the use of storing large data, adding an item to a portfolio will incur sub-cent fees (Jason Gavriel, 2018). Blockchain layer addresses the following requirements separated in two layers: Record Activity and Retrieve Portfolio. The first layer includes the identification component and second includes certification component.

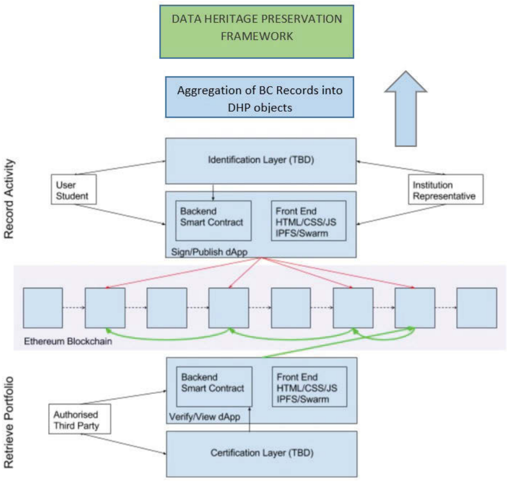

DATA HERITAGE PRESERVATION FRAMEWORK
Aggregation of BC Records into DHP objects
Record Activity
Identification Layer (TBD)
User Student
Institution Representative
Backend Smart Contract
Front End HTML/CSS/JS IPFS/Swarm
Sign/Publish dApp
Ethereum Blockchain
Retrieve Portfolio
Backend Smart Contract
Front End HTML/CSS/JS IPFS/Swarm
Verify/View dApp
Authorised Third Party
Certification Layer (TBD)

Figure 4.18.: Certification through Blockchain. Adapted (Jason Gavriel, 2017)

4.5.2. Blockchain System Architecture

The blockchain technology contains inside the concept of the usability of the keys. The principal concern is about Key Management. The definitions of the security, certification and qualification are inherent inside if the proposal.

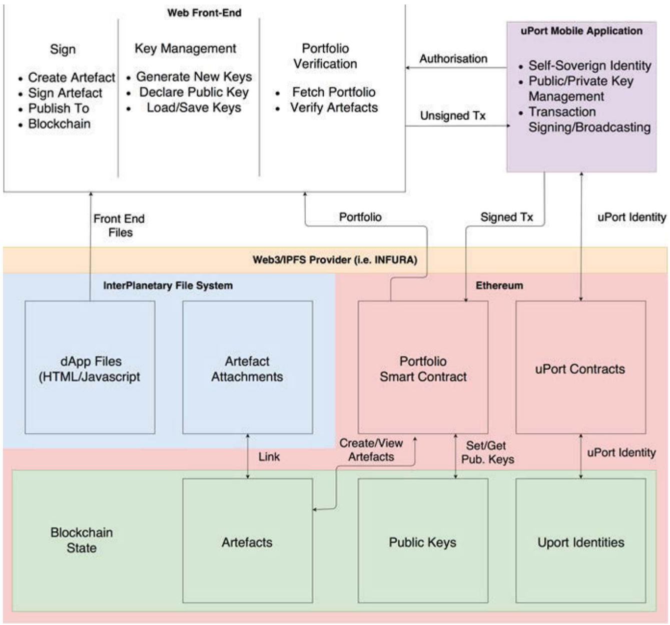


Figure 4.19.: Blockchain System Architecture Adapted (Jason Gavriel, 2017)

5. DHP Implementation and Simulation

For the DHP Implementation the DROID is a perfect environment because it covers the two principles for the DHP framework: D2HP.

The National Archives of UK, develops DROID as a software tool to perform 'automated batch identification of file formats' Brody et al. (2008). It was developed by the Digital Preservation department as part of its digital preservation activities. DROID is designed to meet the fundamental requirements of any digital repository, in order to be able to identify the precise format of all stored digital objects and to link that identification to a central registry of technical information about the format and its dependencies. DROID uses internal signatures to identify and report on the specific file format and version of digital files. These signatures are stored in an XML signature file, generated from information recorded in the PRONOM technical registry. New and updated signatures are regularly added to PRONOM and DROID can be configured to automatically download updated signature files. DROID is a free and open source software made available under the New BSD License. The source code can be downloaded from our GitHub repository.

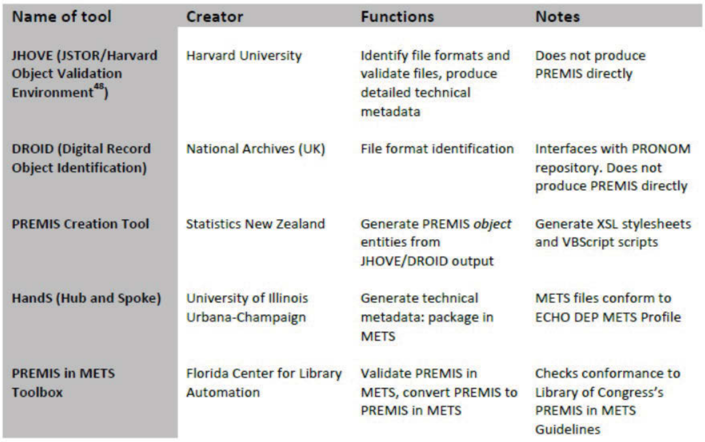
The DHP components extend the traditional application of the DROID framework to identify and preserve digital objects. The addition of computer-analysable software components allow for a significant improvement in the usability and manageability of the digital objects in the DROID environment.

The advantages of the original open-source DROID framework and the DHP application are combined, while avoiding individual drawbacks of each of the solutions in separation. Additional Java programs can execute without relying on the use of a browser to provide a run-time environment.

Techniques for improving the packaging of Java components, including run-time environments and extensions as well as applications, are defined. Dependencies are specified in a manner which enables them to be dynamically located and installed, and enables the sharing of dependent modules (including run-time environments) among applications. The dependency specification technique ensures that all dependent codes will be automatically available at run-time, without requiring a user to perform manual installation. The run-time environment required for an application is specified, and a technique is provided for dynamically changing the run-time that will be used (including the ability to change run-times on a per-program basis), without requiring user intervention.

DROID (Digital Record Object Identification), a batch-processing file format identification package which interfaces directly with PRONOM, a continuously updated registry of file-format-related technical information maintained by the UK National Archives. DROID can output directly into XML. The XML file includes all formats for which identifications can be made with entries in PRONOM and an indication of the quality of the identification. DROID can not produce PREMIS metadata directly, and so their outputs must be processed into PREMIS XML format using one of a number of other tools. It takes DROID output and produces PREMIS Object records, slightly modified from the original PREMIS schema to allow information on the software package used to generate each element to be recorded.

Table 5.1.: Five key PREMIS tools. Adapted from (Gartner & Lavoie (2013)



Name of tool	Creator	Functions	Notes
JHOVE (JSTOR/Harvard Object Validation Environment[48])	Harvard University	Identify file formats and validate files, produce detailed technical metadata	Does not produce PREMIS directly
DROID (Digital Record Object Identification)	National Archives (UK)	File format identification	Interfaces with PRONOM repository. Does not produce PREMIS directly
PREMIS Creation Tool	Statistics New Zealand	Generate PREMIS *object* entities from JHOVE/DROID output	Generate XSL stylesheets and VBScript scripts
HandS (Hub and Spoke)	University of Illinois Urbana-Champaign	Generate technical metadata: package in METS	METS files conform to ECHO DEP METS Profile
PREMIS in METS Toolbox	Florida Center for Library Automation	Validate PREMIS in METS, convert PREMIS to PREMIS in METS	Checks conformance to Library of Congress's PREMIS in METS Guidelines

5.1. Simulation

The identification of the technological tools is an important consideration, for the simulation of the application for the final solution.

- MATLAB: The following steps include the use of Digital Preservation Toolbox. The generation of the algorithms play an important role to have the examples for the using of the code. The main part is to understand the description, the purpose of the mode, key steps, the definition of the data set and the interpretation. The idea is to have, as a result, the INPUTS and the OUTPUTS. The final result will proof the proficiency.

5.1.1. DHP Framework

The methodology includes following activities:

- Mapping and Analysing as a process
- Reengineering preparation
- Process of Designing
- Implement reengineered process
- Improve continuously

Information Management Standard: Well-managed business information is a valuable asset that contributes to good government through:

- Supporting efficient business
- Informing decision-making
- Protecting rights and entitlements.
- Demonstrating government accountability and transparency
- Adding economic value
- Assisting to mitigate risks

The Standard: The Information Management Standard assists Australian Government agencies to create and manage business information effectively by outlining:

- Principles for well-managed information within the Australian Government jurisdiction

- The National Archives of Australia's expectations for the management of business information to enable agencies to meet business, government and community needs and expectations.

The Standard is consistent with the key concepts and principles of International Standard ISO 15489 (2016) Records Management.

Principles: The Standard is based on the following eight Principles that provide the foundation for well-managed business information:

Principle 1: Business information is systematically governed

Principle 2: Necessary business information is created

Principle 3: Business information is adequately described

Principle 4: Business information is suitably stored and preserved

Principle 5: How long business information should be kept is known

Principle 6: Business information is accountably destroyed or transferred

Principle 7: Business information is saved in systems where it can be appropriately managed

Principle 8: Business information is available for use and reuse

The National Archives will review how agencies are performing against the Standard as part of its regular survey and evaluation of the Australian Government information management environment.

Implementation

This Standard is part of a framework that includes implementation guidelines linking the recommended actions to additional detail and technical guidance. The Archives will progressively release guidance for each of the Principles. Additional advice will be created to address any gaps. The Standard does not prescribe how should meet the Principles, but guide to implement the actions to meet the needs of business and information environments.

DROID based Digital Heritage Preservation (D2HP)

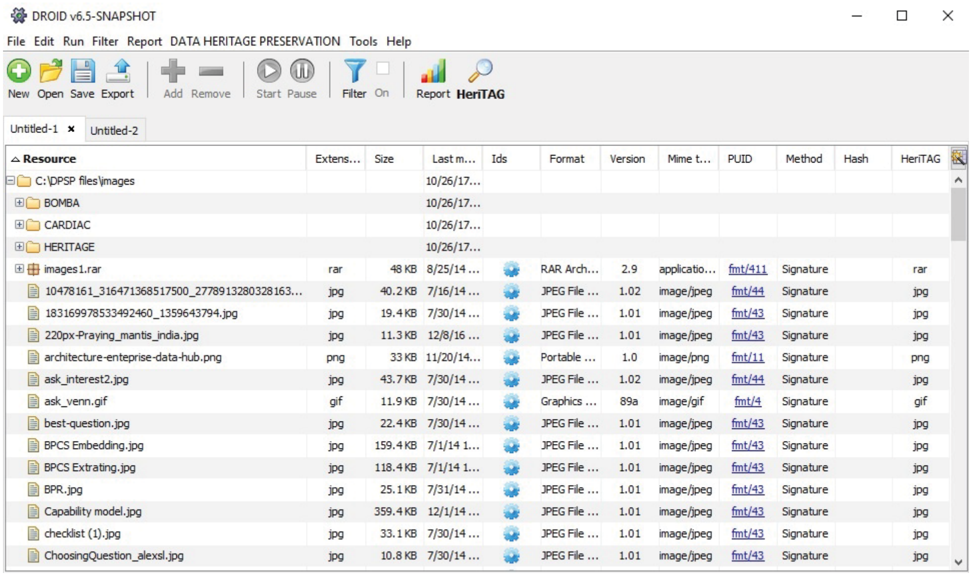
Through the comparison of the Digital Preservation proposals, there is one alternative as a solution for the identification of the objects. To do a successful experimentation, it is primordial explain the principal DHP definitions. Based on the explanation of this solution Ferreira et al. (2006), and taking advantage that it is open source, the implementation process allows us to define DHP through this tool.

DHP through DROID implement some definitions: HeriTAG and DHP Process.

The implementation model in context of the DHP, allows us to evaluate the model proposed. There are three basic steps involved in the implementation: Dataset Definition,Evaluation of the information (through HeriTAG and options) and Validation of results.

5.1.2. Datasets

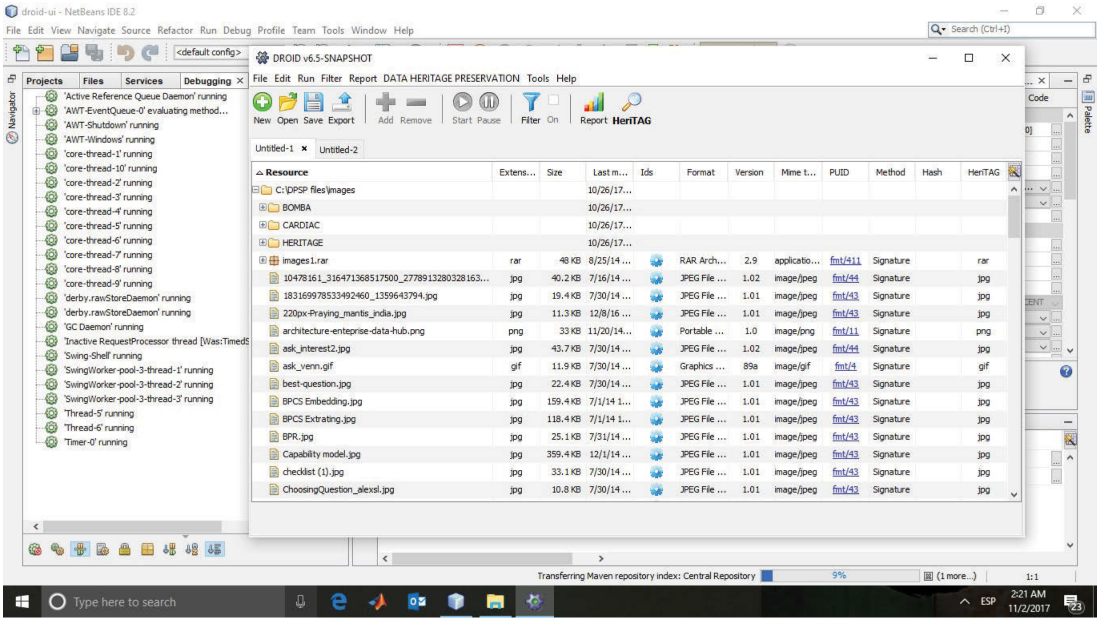
In this implementation the dataset contains many types of data. Images (JPEG, GIF)), Compressed files (RAR). The principal advantage is to differentiate the formats. DROID is perfect in this stage, because it give us a real idea of the management of the Dataset.



Resource	Extens...	Size	Last m...	Ids	Format	Version	Mime t...	PUID	Method	Hash	HeriTAG
C:\DPSP files\images			10/26/17...								
BOMBA			10/26/17...								
CARDIAC			10/26/17...								
HERITAGE			10/26/17...								
images1.rar	rar	48 KB	8/25/14 ...	⚙	RAR Arch...	2.9	applicatio...	fmt/411	Signature		rar
10478161_316471368517500_2778913280328163...	jpg	40.2 KB	7/16/14 ...	⚙	JPEG File ...	1.02	image/jpeg	fmt/44	Signature		jpg
1831699785334924 60_1359643794.jpg	jpg	19.4 KB	7/30/14 ...	⚙	JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
220px-Praying_mantis_india.jpg	jpg	11.3 KB	12/8/16 ...	⚙	JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
architecture-enteprise-data-hub.png	png	33 KB	11/20/14...	⚙	Portable ...	1.0	image/png	fmt/11	Signature		png
ask_interest2.jpg	jpg	43.7 KB	7/30/14 ...	⚙	JPEG File ...	1.02	image/jpeg	fmt/44	Signature		jpg
ask_venn.gif	gif	11.9 KB	7/30/14 ...	⚙	Graphics ...	89a	image/gif	fmt/4	Signature		gif
best-question.jpg	jpg	22.4 KB	7/30/14 ...	⚙	JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
BPCS Embedding.jpg	jpg	159.4 KB	7/1/14 1...	⚙	JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
BPCS Extrating.jpg	jpg	118.4 KB	7/1/14 1...	⚙	JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
BPR.jpg	jpg	25.1 KB	7/31/14 ...	⚙	JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
Capability model.jpg	jpg	359.4 KB	12/1/14 ...	⚙	JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
checklist (1).jpg	jpg	33.1 KB	7/30/14 ...	⚙	JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
ChoosingQuestion_alexsl.jpg	jpg	10.8 KB	7/30/14 ...	⚙	JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg

Figure 5.1.: DHP based on DROID

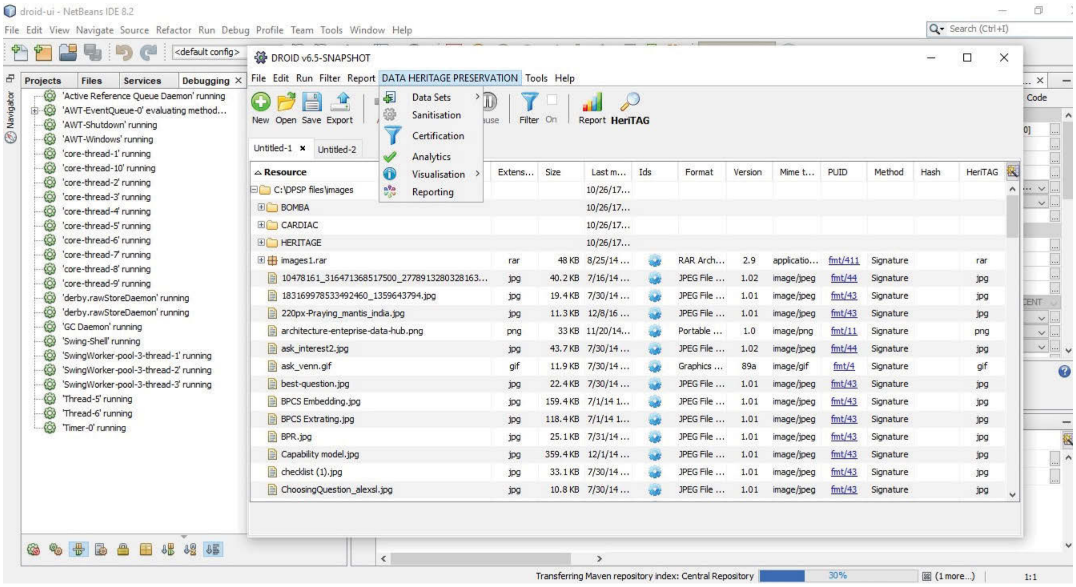
It is the stage, where starts the principal implementation of DHP model, HeriTAG. It is a label to stamp in the data that is validate through the model. The qualification and certification of the data doing the beginning of the implementation of the model.


droid-ui - NetBeans IDE 8.2
File Edit View Navigate Source Refactor Run Debug Profile Team Tools Window Help
DROID v6.5-SNAPSHOT
File Edit Run Filter Report DATA HERITAGE PRESERVATION Tools Help
New Open Save Export Add Remove Start Pause Filter On Report HeriTAG
Untitled-1 Untitled-2
Transferring Maven repository index: Central Repository 9%
Type here to search
2:21 AM 11/2/2017

Resource	Extens...	Size	Last m...	Ids	Format	Version	Mime t...	PUID	Method	Hash	HeriTAG
C:\DPSP files\images			10/26/17...								
BOMBA			10/26/17...								
CARDIAC			10/26/17...								
HERITAGE			10/26/17...								
images1.rar	rar	48 KB	8/25/14 ...		RAR Arch...	2.9	applicatio...	fmt/411	Signature		rar
10478161_316471368517500_2778913280328163...	jpg	40.2 KB	7/16/14 ...		JPEG File ...	1.02	image/jpeg	fmt/44	Signature		jpg
183169978533492460_1359643794.jpg	jpg	19.4 KB	7/30/14 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
220px-Praying_mantis_india.jpg	jpg	11.3 KB	12/8/16 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
architecture-enteprise-data-hub.png	png	33 KB	11/20/14...		Portable ...	1.0	image/png	fmt/11	Signature		png
ask_interest2.jpg	jpg	43.7 KB	7/30/14 ...		JPEG File ...	1.02	image/jpeg	fmt/44	Signature		jpg
ask_venn.gif	gif	11.9 KB	7/30/14 ...		Graphics ...	89a	image/gif	fmt/4	Signature		gif
best-question.jpg	jpg	22.4 KB	7/30/14 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
BPCS Embedding.jpg	jpg	159.4 KB	7/1/14 1...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
BPCS Extrating.jpg	jpg	118.4 KB	7/1/14 1...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
BPR.jpg	jpg	25.1 KB	7/31/14 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
Capability model.jpg	jpg	359.4 KB	12/1/14 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
checklist (1).jpg	jpg	33.1 KB	7/30/14 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
ChoosingQuestion_alexsl.jpg	jpg	10.8 KB	7/30/14 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg

Figure 5.2.: Field HeriTAG DHP based on DROID

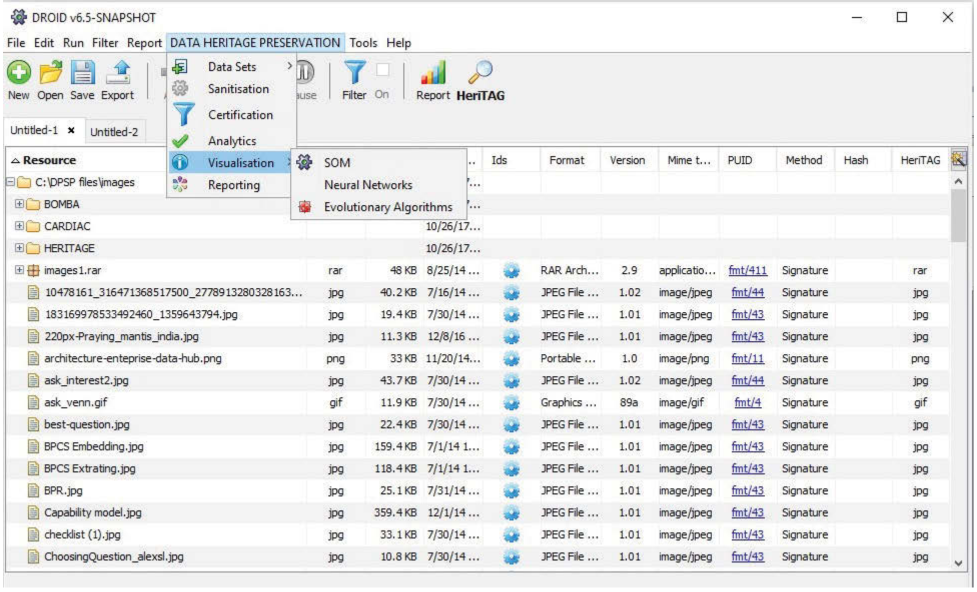
One of the modifications are to add the option of add label DATA HERITAGE PRESERVATION, the name of the DHP proposal. Inside it, as it is explained in the modelling chapter, there are more steps. The D2HP main menu covers: Data Sets, Sanitation, Certification, Analytics, Visualisation, Reporting.


droid-ui - NetBeans IDE 8.2
File Edit View Navigate Source Refactor Run Debug Profile Team Tools Window Help
DROID v6.5-SNAPSHOT
File Edit Run Filter Report DATA HERITAGE PRESERVATION Tools Help
Data Sets
Sanitisation
Certification
Analytics
Visualisation
Reporting
New Open Save Export Filter On Report HeriTAG
Untitled-1 Untitled-2
Transferring Maven repository index: Central Repository 30%

Resource	Extens...	Size	Last m...	Ids	Format	Version	Mime t...	PUID	Method	Hash	HeriTAG
C:\DPSP files\images			10/26/17...								
BOMBA			10/26/17...								
CARDIAC			10/26/17...								
HERITAGE			10/26/17...								
images1.rar	rar	48 KB	8/25/14 ...		RAR Arch...	2.9	applicatio...	fmt/411	Signature		rar
10478161_316471368517500_2778913280328163...	jpg	40.2 KB	7/16/14 ...		JPEG File ...	1.02	image/jpeg	fmt/44	Signature		jpg
183169978533492460_1359643794.jpg	jpg	19.4 KB	7/30/14 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
220px-Praying_mantis_india.jpg	jpg	11.3 KB	12/8/16 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
architecture-enteprise-data-hub.png	png	33 KB	11/20/14...		Portable ...	1.0	image/png	fmt/11	Signature		png
ask_interest2.jpg	jpg	43.7 KB	7/30/14 ...		JPEG File ...	1.02	image/jpeg	fmt/44	Signature		jpg
ask_venn.gif	gif	11.9 KB	7/30/14 ...		Graphics ...	89a	image/gif	fmt/4	Signature		gif
best-question.jpg	jpg	22.4 KB	7/30/14 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
BPCS Embedding.jpg	jpg	159.4 KB	7/1/14 1...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
BPCS Extrating.jpg	jpg	118.4 KB	7/1/14 1...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
BPR.jpg	jpg	25.1 KB	7/31/14 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
Capability model.jpg	jpg	359.4 KB	12/1/14 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
checklist (1).jpg	jpg	33.1 KB	7/30/14 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
ChoosingQuestion_alexsl.jpg	jpg	10.8 KB	7/30/14 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg

Figure 5.3.: Menu for DHP on DROID

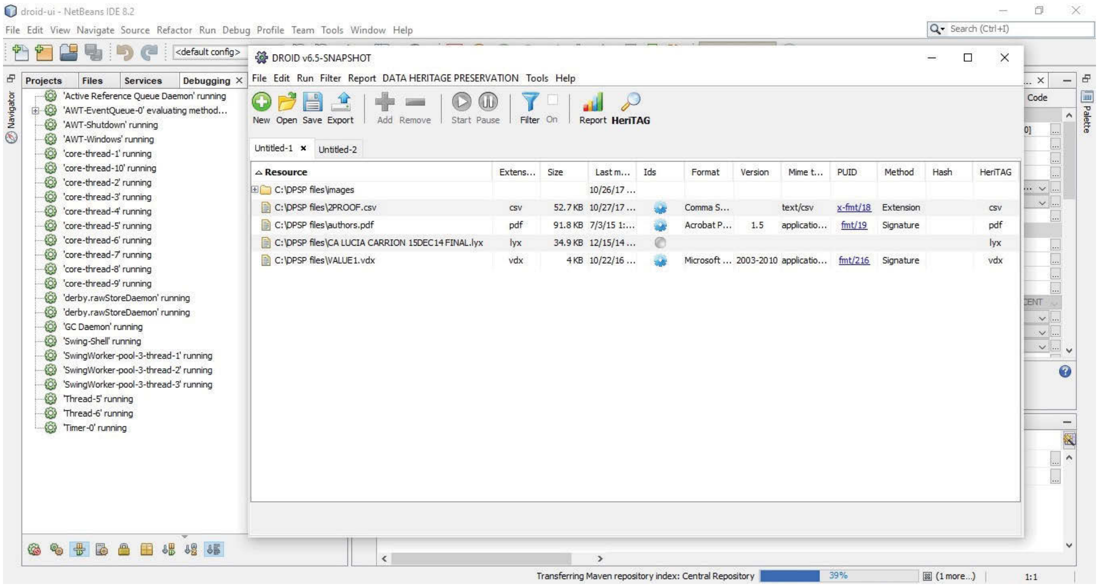
At this stage the options for Visualisation are: SOM, Neural Networks and Evolutionary Algorithms. Each of these options need and extended development in terms of the complexity inside of the concepts around preservation. However, the consideration of these alternatives is very important in order to have complete extraction of the information. The next step is to develop the visualisation of data. In the Figure 5.4 there are the details.



Resource			..	Ids	Format	Version	Mime t...	PUID	Method	Hash	HeriTAG
C:\DPSP files\images			...								
BOMBA			...								
CARDIAC			10/26/17...								
HERITAGE			10/26/17...								
images1.rar	rar	48 KB	8/25/14 ...		RAR Arch...	2.9	applicatio...	fmt/411	Signature		rar
10478161_316471368517500_2778913280328163...	jpg	40.2 KB	7/16/14 ...		JPEG File ...	1.02	image/jpeg	fmt/44	Signature		jpg
183169978533492460_1359643794.jpg	jpg	19.4 KB	7/30/14 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
220px-Praying_mantis_india.jpg	jpg	11.3 KB	12/8/16 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
architecture-enteprise-data-hub.png	png	33 KB	11/20/14...		Portable ...	1.0	image/png	fmt/11	Signature		png
ask_interest2.jpg	jpg	43.7 KB	7/30/14 ...		JPEG File ...	1.02	image/jpeg	fmt/44	Signature		jpg
ask_venn.gif	gif	11.9 KB	7/30/14 ...		Graphics ...	89a	image/gif	fmt/4	Signature		gif
best-question.jpg	jpg	22.4 KB	7/30/14 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
BPCS Embedding.jpg	jpg	159.4 KB	7/1/14 1...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
BPCS Extrating.jpg	jpg	118.4 KB	7/1/14 1...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
BPR.jpg	jpg	25.1 KB	7/31/14 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
Capability model.jpg	jpg	359.4 KB	12/1/14 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
checklist (1).jpg	jpg	33.1 KB	7/30/14 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
ChoosingQuestion_alexsl.jpg	jpg	10.8 KB	7/30/14 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg

Figure 5.4.: Menu for Visualisation on DROID

For the Evaluation of Data the attributes, the lists that manage DROID


droid-ui - NetBeans IDE 8.2
File Edit View Navigate Source Refactor Run Debug Profile Team Tools Window Help
Search (Ctrl+I)
<default config>
Projects Files Services Debugging
Navigator
Palette
Code
DROID v6.5-SNAPSHOT
File Edit Run Filter Report DATA HERITAGE PRESERVATION Tools Help
New Open Save Export
Add Remove
Start Pause
Filter On
Report HeriTAG
Untitled-1
Untitled-2

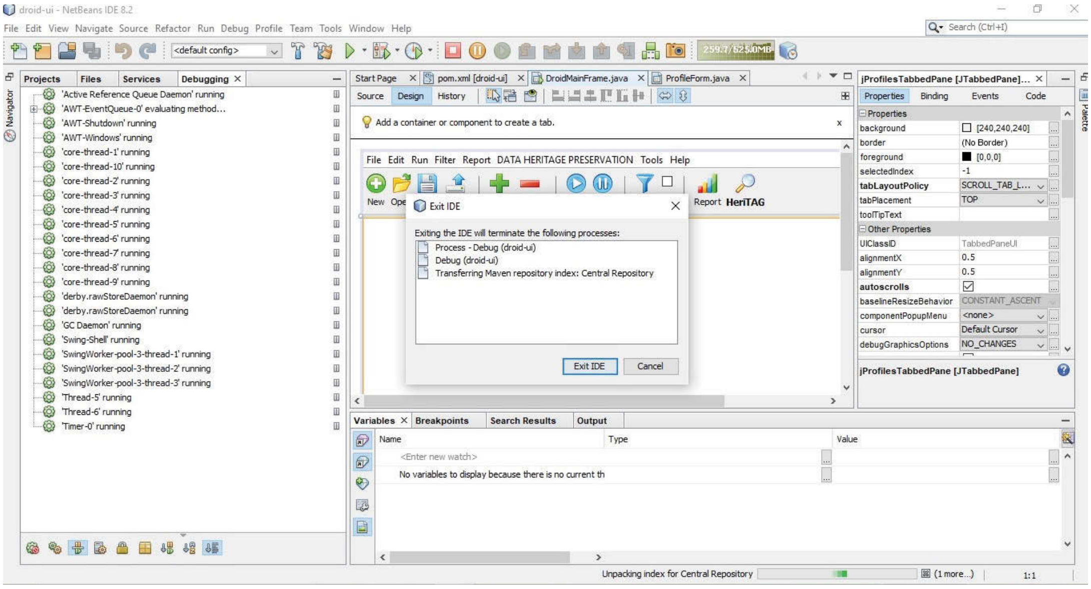
'Active Reference Queue Daemon' running
'AWT-EventQueue-0' evaluating method...
'AWT-Shutdown' running
'AWT-Windows' running
'core-thread-1' running
'core-thread-10' running
'core-thread-2' running
'core-thread-3' running
'core-thread-4' running
'core-thread-5' running
'core-thread-6' running
'core-thread-7' running
'core-thread-8' running
'core-thread-9' running
'derby.rawStoreDaemon' running
'derby.rawStoreDaemon' running
'GC Daemon' running
'Swing-Shell' running
'SwingWorker-pool-3-thread-1' running
'SwingWorker-pool-3-thread-2' running
'SwingWorker-pool-3-thread-3' running
'Thread-5' running
'Thread-6' running
'Timer-0' running

Resource	Extens...	Size	Last m...	Ids	Format	Version	Mime t...	PUID	Method	Hash	HeriTAG
C:\DPSP files\images			10/26/17 ...								
C:\DPSP files\2PROOF.csv	csv	52.7 KB	10/27/17 ...		Comma S...		text/csv	x-fmt/18	Extension		csv
C:\DPSP files\authors.pdf	pdf	91.8 KB	7/3/15 1:...		Acrobat P...	1.5	applicatio...	fmt/19	Signature		pdf
C:\DPSP files\CA LUCIA CARRION 15DEC14 FINAL.lyx	lyx	34.9 KB	12/15/14 ...								lyx
C:\DPSP files\VALUE1.vdx	vdx	4 KB	10/22/16 ...		Microsoft ...	2003-2010	applicatio...	fmt/216	Signature		vdx

Transferring Maven repository index: Central Repository 39% (1 more...) 1:1

Figure 5.5.: Dataset for DHP

The next Figure 5.6shows the implementation inside the code of DROID. The definition of the repository and the correct place to define the label and the tags. These design stage, define the integration of DHP in DROID. As a solution, DHP comes to extend the functionality of the original opens source code.


droid-ui - NetBeans IDE 8.2
File Edit View Navigate Source Refactor Run Debug Profile Team Tools Window Help
Search (Ctrl+I)
<default config>
Projects Files Services Debugging
Start Page pom.xml [droid-ui] DroidMainFrame.java ProfileForm.java
Source Design History
File Edit Run Filter Report DATA HERITAGE PRESERVATION Tools Help
New Ope Report HeriTAG
Exit IDE Cancel
Variables Breakpoints Search Results Output
Name Type Value
jProfilesTabbedPane [JTabbedPane] Properties Binding Events Code
Navigator
Palette

'Active Reference Queue Daemon' running
'AWT-EventQueue-0' evaluating method...
'AWT-Shutdown' running
'AWT-Windows' running
'core-thread-1' running
'core-thread-10' running
'core-thread-2' running
'core-thread-3' running
'core-thread-4' running
'core-thread-5' running
'core-thread-6' running
'core-thread-7' running
'core-thread-8' running
'core-thread-9' running
'derby.rawStoreDaemon' running
'derby.rawStoreDaemon' running
'GC Daemon' running
'Swing-Shell' running
'SwingWorker-pool-3-thread-1' running
'SwingWorker-pool-3-thread-2' running
'SwingWorker-pool-3-thread-3' running
'Thread-5' running
'Thread-6' running
'Timer-0' running

Add a container or component to create a tab.

Exit IDE

Exiting the IDE will terminate the following processes:

- Process - Debug (droid-ui)
- Debug (droid-ui)
- Transferring Maven repository index: Central Repository

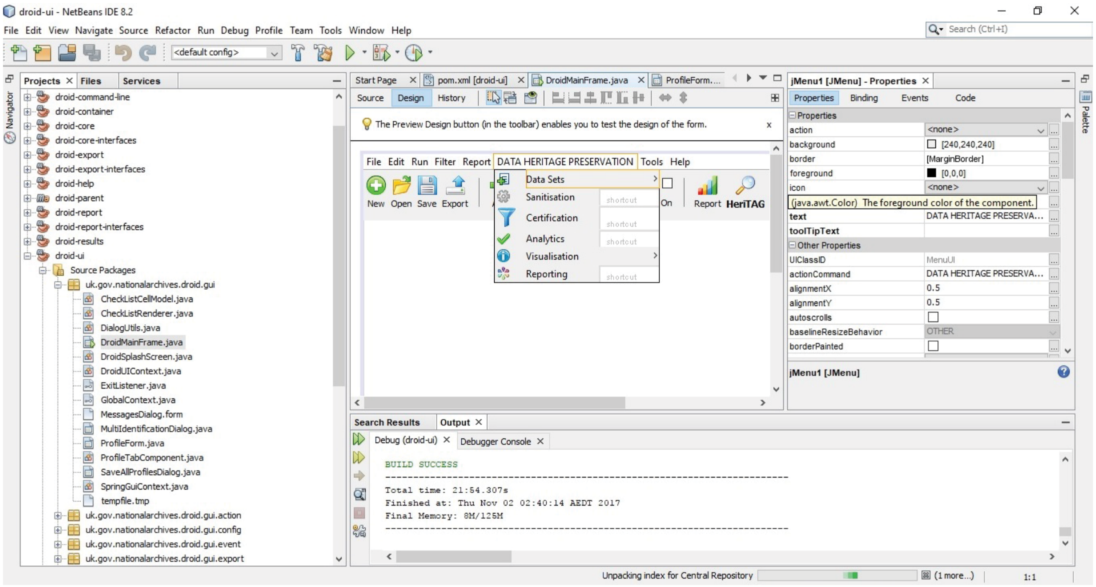
Properties	
background	□ [240,240,240]
border	(No Border)
foreground	■ [0,0,0]
selectedIndex	-1
tabLayoutPolicy	SCROLL_TAB_L...
tabPlacement	TOP
toolTipText	
Other Properties	
UIClassID	TabbedPaneUI
alignmentX	0.5
alignmentY	0.5
autoscrolls	☑
baselineResizeBehavior	CONSTANT_ASCENT
componentPopupMenu	<none>
cursor	Default Cursor
debugGraphicsOptions	NO_CHANGES

<Enter new watch>
No variables to display because there is no current th

Unpacking index for Central Repository (1 more...) 1:1

Figure 5.6.: Designing the Menu

For the Main Menu in DHP, inside the 'DroidMainFrame.java', there are ways to set the appearance for the different characteristics.


droid-ui - NetBeans IDE 8.2
File Edit View Navigate Source Refactor Run Debug Profile Team Tools Window Help
Search (Ctrl+I)
<default config>
Projects Files Services
Start Page pom.xml [droid-ui] DroidMainFrame.java ProfileForm....
Source Design History
File Edit Run Filter Report DATA HERITAGE PRESERVATION Tools Help
New Open Save Export On Report HeriTAG
Data Sets
Sanitisation
Certification
Analytics
Visualisation
Reporting
shortcut
jMenu1 [JMenu] - Properties Properties Binding Events Code
Search Results Output
Debug (droid-ui) Debugger Console
Navigator
Palette

droid-command-line
droid-container
droid-core
droid-core-interfaces
droid-export
droid-export-interfaces
droid-help
droid-parent
droid-report
droid-report-interfaces
droid-results
droid-ui
Source Packages
uk.gov.nationalarchives.droid.gui
CheckListCellModel.java
CheckListRenderer.java
DialogUtils.java
DroidMainFrame.java
DroidSplashScreen.java
DroidUIContext.java
ExitListener.java
GlobalContext.java
MessagesDialog.form
MultiIdentificationDialog.java
ProfileForm.java
ProfileTabComponent.java
SaveAllProfilesDialog.java
SpringGuiContext.java
tempfile.tmp
uk.gov.nationalarchives.droid.gui.action
uk.gov.nationalarchives.droid.gui.config
uk.gov.nationalarchives.droid.gui.event
uk.gov.nationalarchives.droid.gui.export

The Preview Design button (in the toolbar) enables you to test the design of the form.

Properties	
action	<none>
background	□ [240,240,240]
border	[MarginBorder]
foreground	■ [0,0,0]
icon	<none>
(java.awt.Color) The foreground color of the component.	
text	DATA HERITAGE PRESERVA...
toolTipText	
Other Properties	
UIClassID	MenuUI
actionCommand	DATA HERITAGE PRESERVA...
alignmentX	0.5
alignmentY	0.5
autoscrolls	□
baselineResizeBehavior	OTHER
borderPainted	□

jMenu1 [JMenu]

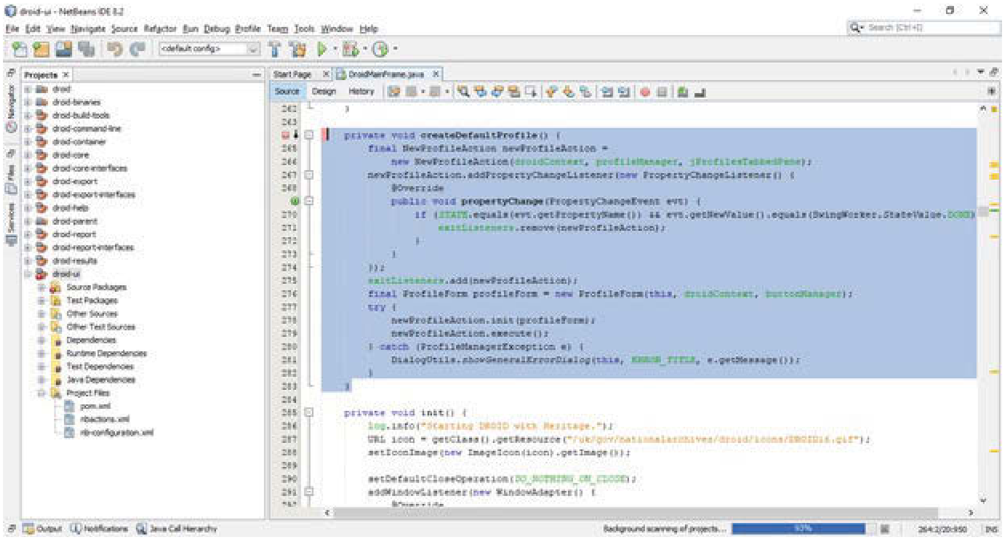
```
BUILD SUCCESS
------------------------------------------------------------------------
Total time: 21:54.307s
Finished at: Thu Nov 02 02:40:14 AEDT 2017
Final Memory: 8M/125M
------------------------------------------------------------------------
```

Unpacking index for Central Repository (1 more...) 1:1

Figure 5.7.: DHP Debug on DROID

5.1.3. Coding for the Implementation

As an example, the code for the implementation of the layers of the program are defined in programming language Java.



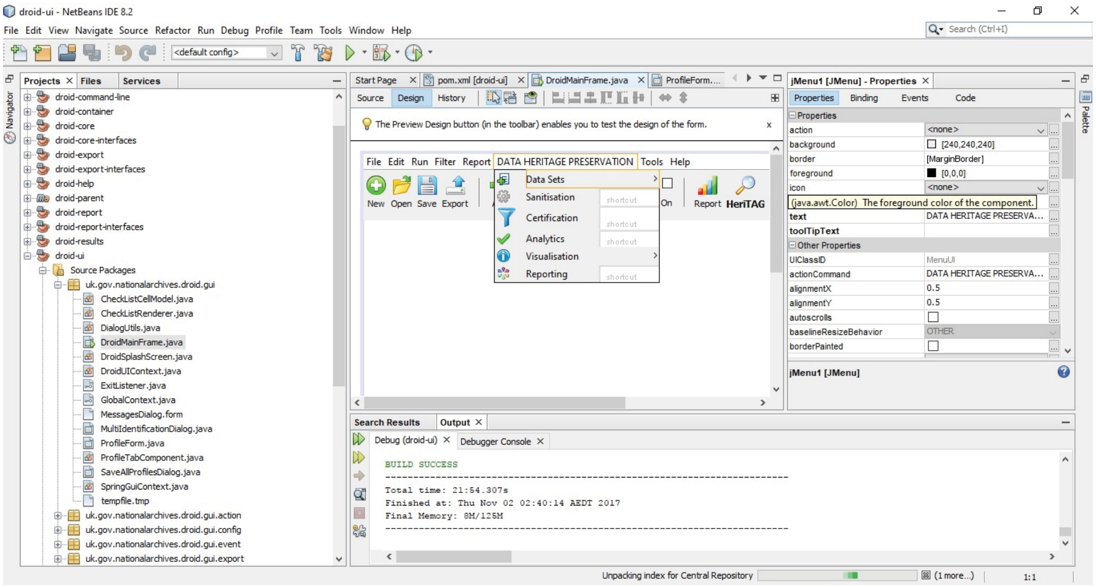
```
    }

    private void createDefaultProfile() {
        final NewProfileAction newProfileAction =
            new NewProfileAction(droidContext, profileManager, jProfilesTabbedPane);
        newProfileAction.addPropertyChangeListener(new PropertyChangeListener() {
            @Override
            public void propertyChange(PropertyChangeEvent evt) {
                if (STATE.equals(evt.getPropertyName()) && evt.getNewValue().equals(SwingWorker.StateValue.DONE)
                    exitListeners.remove(newProfileAction);
                }
            }
        });
        exitListeners.add(newProfileAction);
        final ProfileForm profileForm = new ProfileForm(this, droidContext, buttonManager);
        try {
            newProfileAction.init(profileForm);
            newProfileAction.execute();
        } catch (ProfileManagerException e) {
            DialogUtils.showGeneralErrorDialog(this, ERROR_TITLE, e.getMessage());
        }
    }

    private void init() {
        log.info("Starting DROID with Heritage.");
        URL icon = getClass().getResource("/uk/gov/nationalarchives/droid/icons/DROID16.gif");
        setIconImage(new ImageIcon(icon).getImage());

        setDefaultCloseOperation(DO_NOTHING_ON_CLOSE);
        addWindowListener(new WindowAdapter() {
            @Override
```

Figure 5.8.: Create Default Profile

Code in DROID

```
private void createDefaultProfile() {
final NewProfileAction newProfileAction =
new NewProfileAction(droidContext, profileManager, jProfilesTabbedPane);
newProfileAction.addPropertyChangeListener(new PropertyChangeListener() {
@Override
public void propertyChange(PropertyChangeEvent evt) {
if (STATE.equals(evt.getPropertyName())
&& evt.getNewValue().equals(SwingWorker.StateValue.DONE)) {
exitListeners.remove(newProfileAction);
}
}
});
exitListeners.add(newProfileAction);
final ProfileForm profileForm = new ProfileForm(this, droidContext, buttonManager);
try {
newProfileAction.init(profileForm);
newProfileAction.execute();
} catch (ProfileManagerException e) {
```

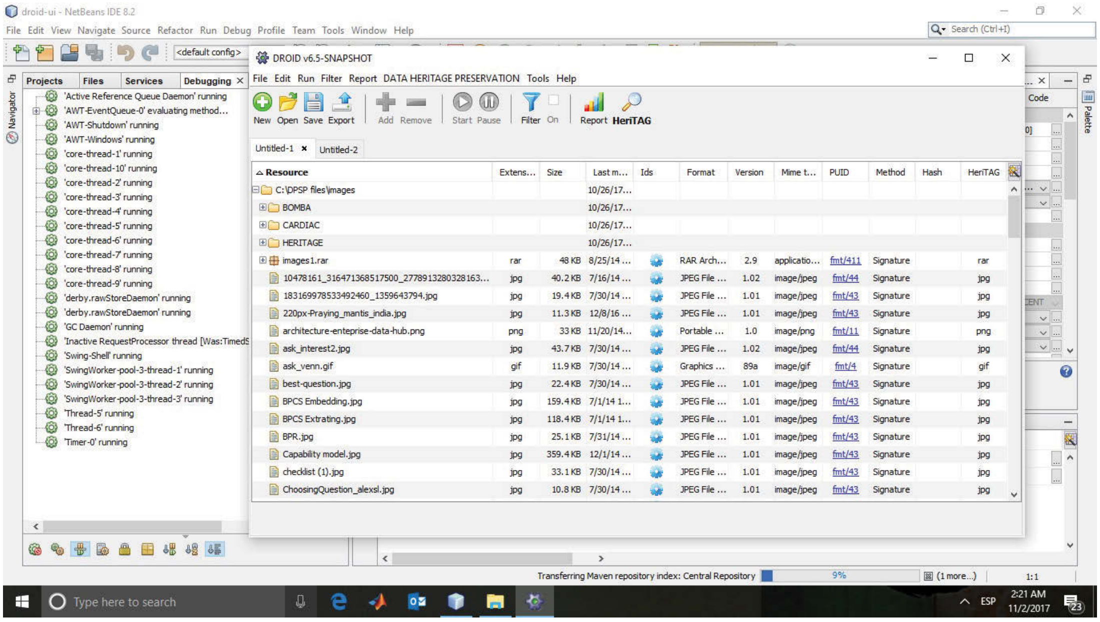
```
DialogUtils.showGeneralErrorDialog(this, ERROR_TITLE, e.getMessage());

}

}
```



droid-ui - NetBeans IDE 8.2
File Edit View Navigate Source Refactor Run Debug Profile Team Tools Window Help
DATA HERITAGE PRESERVATION
Data Sets
Sanitisation
Certification
Analytics
Visualisation
Reporting
Report
HeriTAG
jMenu1 [JMenu] - Properties

```
BUILD SUCCESS
------------------------------------------------------------------------
Total time: 21:54.307s
Finished at: Thu Nov 02 02:40:14 AEDT 2017
Final Memory: 8M/125M
------------------------------------------------------------------------
```

Figure 5.9.: Build Success on DROID



DROID v6.5-SNAPSHOT
File Edit Run Filter Report DATA HERITAGE PRESERVATION Tools Help
New Open Save Export Add Remove Start Pause Filter On Report HeriTAG
Untitled-1 Untitled-2
Transferring Maven repository index: Central Repository 9%

Resource	Extens...	Size	Last m...	Ids	Format	Version	Mime t...	PUID	Method	Hash	HeriTAG
C:\DPSP files\images			10/26/17...								
BOMBA			10/26/17...								
CARDIAC			10/26/17...								
HERITAGE			10/26/17...								
images1.rar	rar	48 KB	8/25/14 ...		RAR Arch...	2.9	applicatio...	fmt/411	Signature		rar
10478161_316471368517500_2778913280328163...	jpg	40.2 KB	7/16/14 ...		JPEG File ...	1.02	image/jpeg	fmt/44	Signature		jpg
183169978533492460_1359643794.jpg	jpg	19.4 KB	7/30/14 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
220px-Praying_mantis_india.jpg	jpg	11.3 KB	12/8/16 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
architecture-enteprise-data-hub.png	png	33 KB	11/20/14...		Portable ...	1.0	image/png	fmt/11	Signature		png
ask_interest2.jpg	jpg	43.7 KB	7/30/14 ...		JPEG File ...	1.02	image/jpeg	fmt/44	Signature		jpg
ask_venn.gif	gif	11.9 KB	7/30/14 ...		Graphics ...	89a	image/gif	fmt/4	Signature		gif
best-question.jpg	jpg	22.4 KB	7/30/14 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
BPCS Embedding.jpg	jpg	159.4 KB	7/1/14 1...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
BPCS Extrating.jpg	jpg	118.4 KB	7/1/14 1...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
BPR.jpg	jpg	25.1 KB	7/31/14 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
Capability model.jpg	jpg	359.4 KB	12/1/14 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
checklist (1).jpg	jpg	33.1 KB	7/30/14 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg
ChoosingQuestion_alexsl.jpg	jpg	10.8 KB	7/30/14 ...		JPEG File ...	1.01	image/jpeg	fmt/43	Signature		jpg

Figure 5.10.: Field HeriTAG DHP based on DROID

5.1.4. Experimentation Results

After the definition of the experimental environment, and application in the DHP Framework, there are some results based on the number of files vs. time that they take.

Both Performance and the Accessibility are included. In the hypothesis, the consideration is centralized on checkup the results based on manageability and dependability.

The Functional requirements and the environment as below descriptionGartner & Lavoie (2013):

- The system must conform to the current version of the TNA Application Development Guidelines.
- The server-side elements of the system must be compatible with Windows 2003 server and SQL Server 2000.
- The client-side elements of the system must be compatible with the following browsers: Internet Explorer 5 and later, Netscape Navigator 4 and later, Safari 1.0 and later, Firefox 0.9 and later, and Mozilla 1.0 and later.
- The system must be year 2000 compliant.
- The system must comply with e-GIF version 6 and the Guidelines for UK Government Websites.

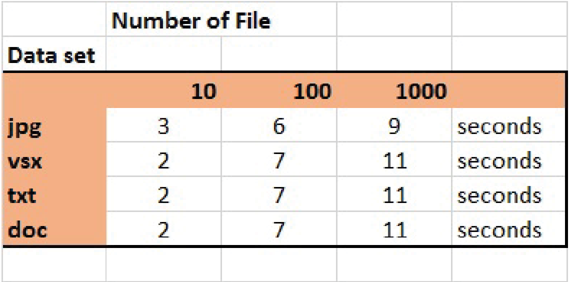
The description of the identification tool detailed below based on Gartner & Lavoie (2013):

- A standalone Java (JRE 1.4 or later) application will be developed to perform automatic file format identification, using signatures recorded in PRONOM.
- The tool will allow a user to browse any file system accessible from the computer on which the application is installed, and select single or multiple files for identification, including the contents of entire folders, to create a list of files to be identified. The user must also be able to subsequently remove files from the list.
- The tool will display the resultant file list in a text box. The list will display the full filename, with the option to toggle display of the fully-qualified pathname on or off.

- The tool will perform automatic file format identification of all files on the selected list in a single operation, using an XML signature file generated from PRONOM. TNA will work with the developer to provide an appropriate algorithm.
- The result of the identification will be displayed against each file in the list (e.g. "Identified", "Tentative" or "Not Identified")
- Full details of each identification (PUID, Format, Version) will be displayed in a separate text box, dynamically linked to the currently highlighted file in the file list text box. The tool will support multiple tentative identifications for a file, where the precise format version cannot be ascertained.
- The tool will allow the results of the identification to be saved as an XML, CSV or printer-friendly report.
- The tool will automatically check the PRONOM website at user-defined intervals for updated signature files and download them.
- The update mechanism will be user-configurable in terms of timing, proxy server settings and authentication.
- TNA will generate updated XML signature files as required, using the maintenance system, and make them available on the PRONOM website for download.
- The tool will be deployed as a single JAR package, to be downloaded from the PRONOM website.
- A separate functional specification will be required for this tool.

5.1.4.1. First Scenario: Performance

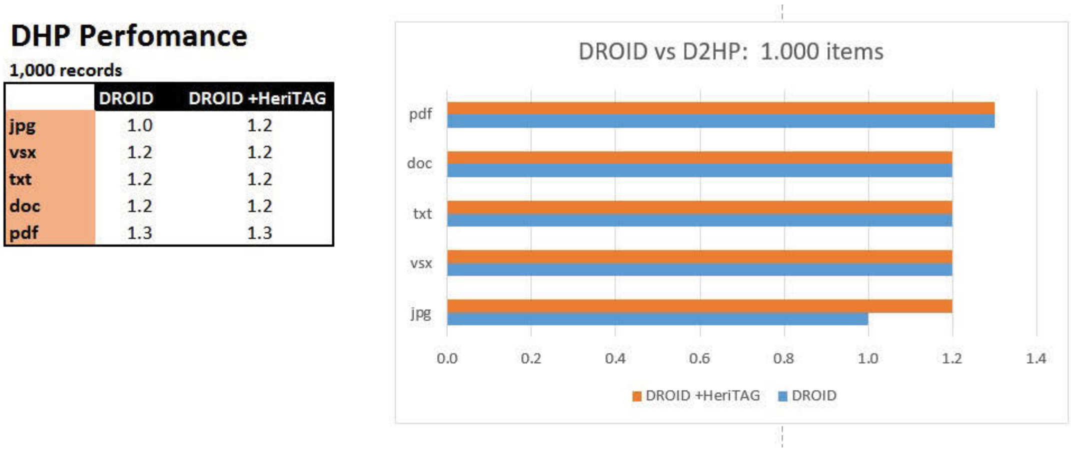
Performance DROID environment vs. DROID + HeriTAG



	Number of File			
Data set				
	10	100	1000	
jpg	3	6	9	seconds
vsx	2	7	11	seconds
txt	2	7	11	seconds
doc	2	7	11	seconds

Figure 5.11.: Data set and Number of Files

The delay handling Heritage component.1.000, 5000 10.000. The experimentation is Big Data Application and deals with amounts of data.

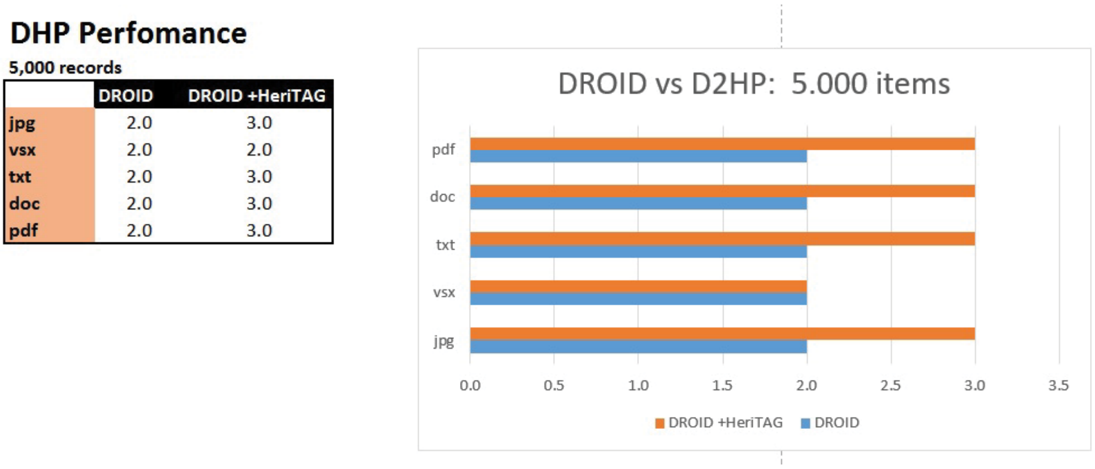


DHP Perfomance

1,000 records

	DROID	DROID +HeriTAG
jpg	1.0	1.2
vsx	1.2	1.2
txt	1.2	1.2
doc	1.2	1.2
pdf	1.3	1.3

Figure 5.12.: Data set 1.000

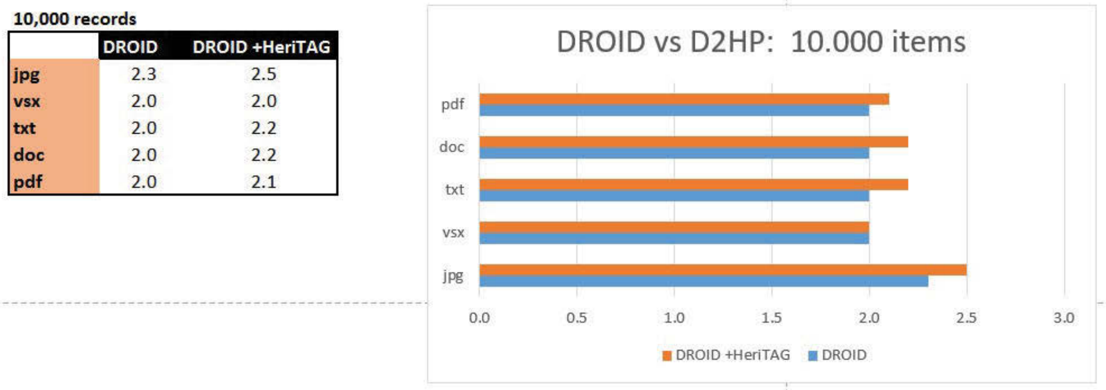


DHP Perfomance

5,000 records

	DROID	DROID +HeriTAG
jpg	2.0	3.0
vsx	2.0	2.0
txt	2.0	3.0
doc	2.0	3.0
pdf	2.0	3.0

Figure 5.13.: Data set 5.000



10,000 records

	DROID	DROID +HeriTAG
jpg	2.3	2.5
vsx	2.0	2.0
txt	2.0	2.2
doc	2.0	2.2
pdf	2.0	2.1

Figure 5.14.: Data set 10.000

5.1.4.2. Summary of Identification of Usage of D2HP

Based on the above analysis the identification could be managed by the first menu. Below we describe the experimental work and the empirical evaluation component of the research. We measured with DHP model the identification of the electronic objects to confirm that the experimentation was acceptably. Methods used on DHP framework based on DROID identify the dataset through generalization.

The base Experiment is on function of time. There will be provided a brief overview of how D2HP works and its parameters, and then explain how we modified these parameters for the identification of the dataset.

5.1.4.3. Second Scenario: Manageability and Dependability

Based on the initial hypothesis the Manageability and Dependability are the key concepts to be measured by the model. In the following figures there are two metrics:

- *Manageability:* measured by seconds Y-axis and file format in X-axis. The content is the variable number of files 10, 100 and 1000 files

- *Dependability:* Shows the difference between the files format (jpg, vsx, txt, doc) X-axis and the Y-axis seconds for 10 files as an example for the tendency.

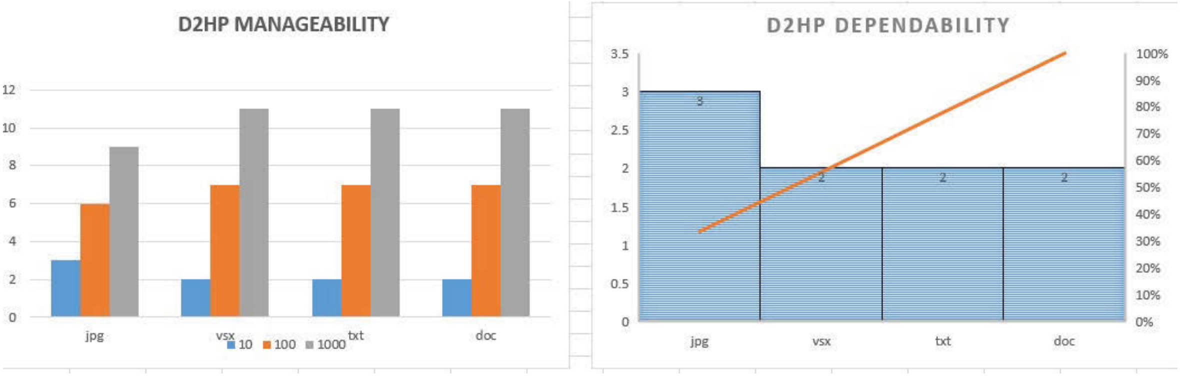


Figure 5.15.: Manageability and Dependability

6. DHP Vision and Benchmarking Approach

Conceptual Architecture Having performed the analysis on stakeholder quality concerns; several competing system architectures had been considered (Tiered architecture, distributed objects architecture, and pipes and filters). Eventually the architecture that had been selected was the event driven architecture.

The design of an EA should be a task of the enterprise´s strategic management. Ross et. al, even claim, that enterprise architecture should be a strategy itself.

Mentioned in the publication Gordon et al. (2015), the information formats show one result about digital age.

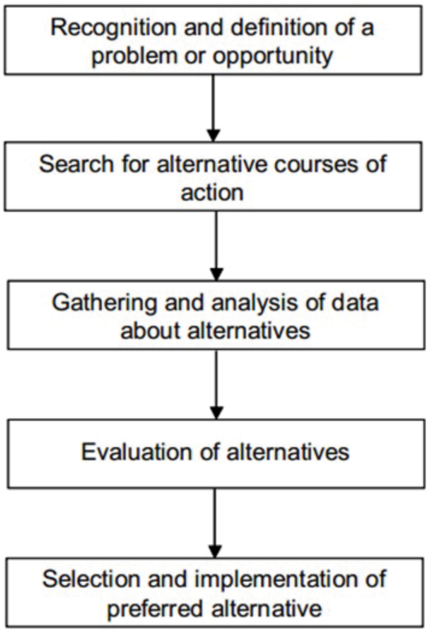

Recognition and definition of a problem or opportunity
Search for alternative courses of action
Gathering and analysis of data about alternatives
Evaluation of alternatives
Selection and implementation of preferred alternative

Figure 6.1.: Functional entities

6.1. Case Studies

The statements between behavior and data preservation are the appropriate resources to evaluate through the case studies.

'Migration is the method of repeated conversion of files or objects. Emulation denotes the duplication of the functionality of systems, be it software, hardware parts, or legacy computer systems as a whole, needed to display, access, or edit a certain document' Strodl et al. (2007).

Creating open data is considered an important goal in the research community. Open data is said to ensure accountability in research by allowing others access to researchers' data and meth

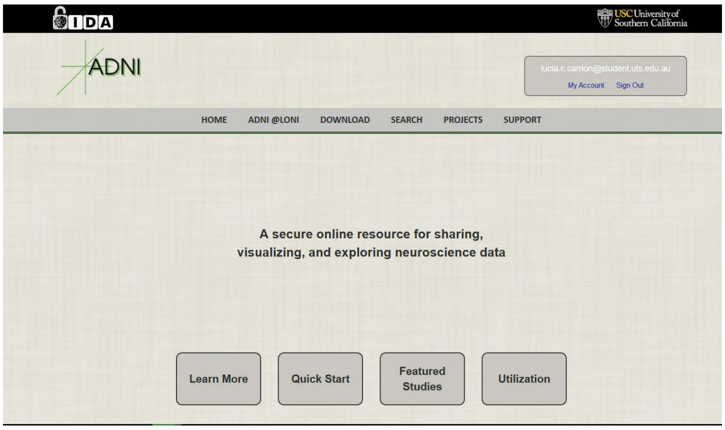
There are several data sets applicable to validate the DHP model. However, the main data sets used for experimentation with the proposed model are:

- Health (Medical Records) Jack et al. (2008)
- Government (Public Archives)Garfinkel et al. (2009)
- Business (Resource Management) Chaczko et al. (2014)
- Geological Data Carrion Gordon et al. (2017)

6.1.1. Medical Records

Clinician documentation with fully integrated data systems support. Prior notes and data are input for the following note and decisions. Machine analyzes input and displays suggested diagnoses and problem list, and test and treatment recommendations based on various levels of evidence: CPG – clinical practice guidelines, UTD – Up to Date®, DCDM – Dynamic Clinical Data Mining. Incorporating Data Celi et al. (2015).

Overwhelmingly, the most important characteristic of the electronic note is its potential for the creation and reception of what we term "bidirectional data streams" to inform both decision-making and research. By bidirectional data exchange, we mean that electronic notes have the potential to provide data streams to the entirety of the EHR database and vice versa. The data from the note can be recorded, stored, accessed, retrieved, and mined for a variety of real-time and future uses. This process should be an automatic and intrinsic property of clinical information systems. The incoming data stream is currently produced by the data that is slated for import into the note according to the software requirements of the application and the locally available interfaces.



A secure online resource for sharing,
visualizing, and exploring neuroscience data

Figure 6.2.: ADNI Front Page

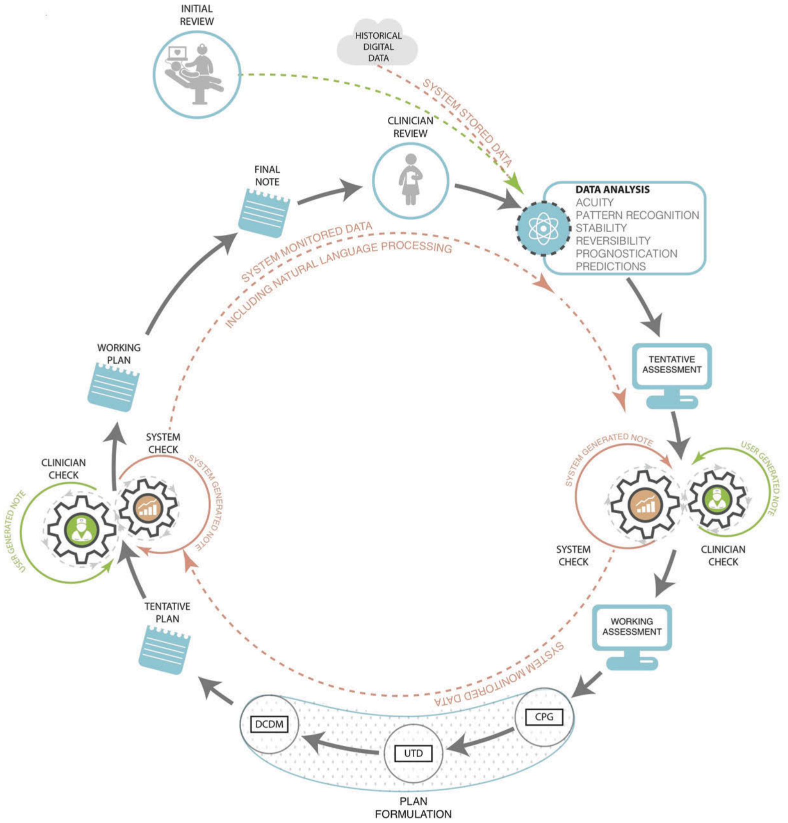


Figure 6.3.: ADNI Front Page. Source:Celi et al. (2015)

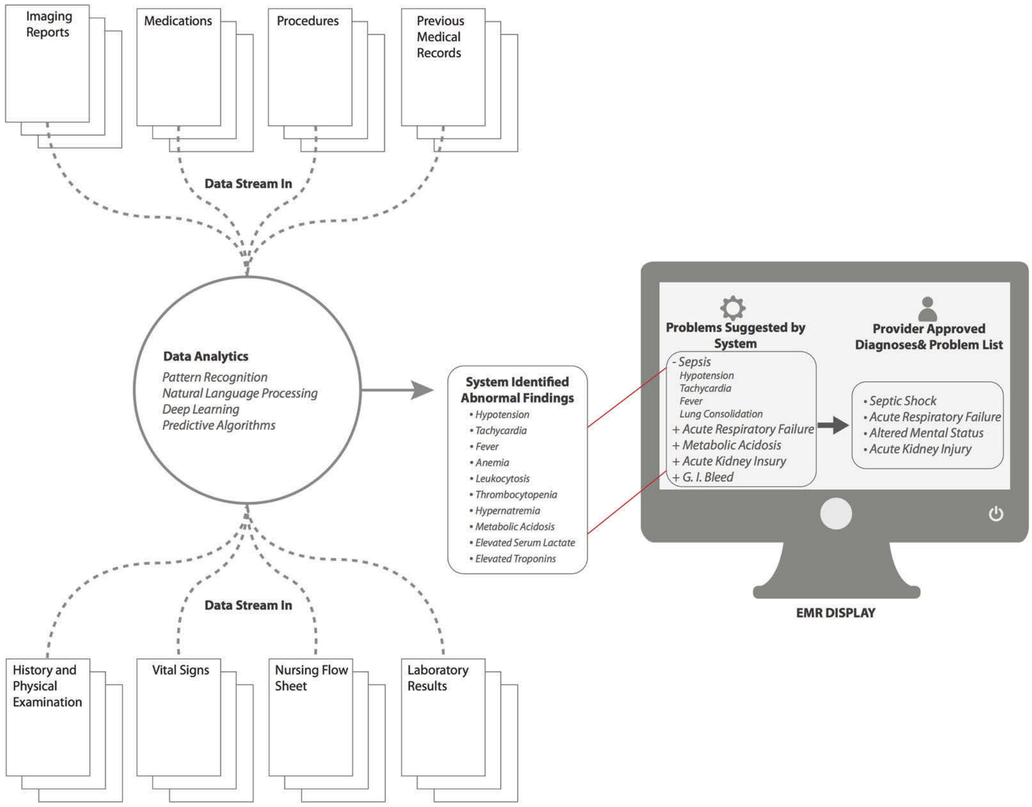
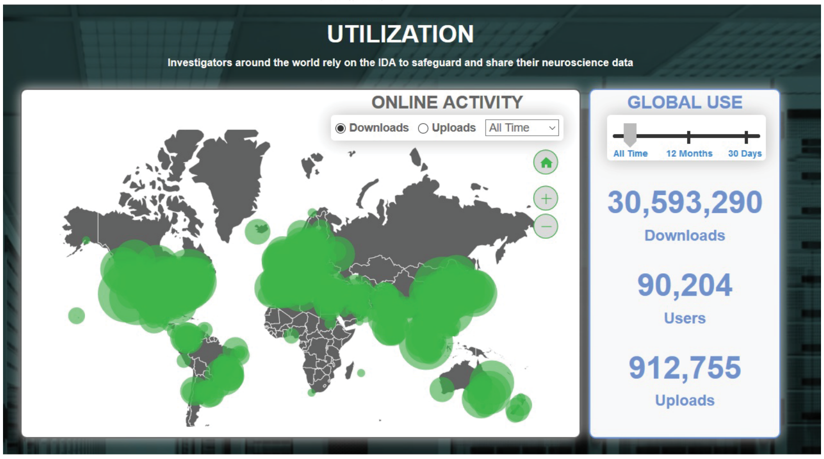


Figure 6.4.: Data Analytics Health Records (Celi, 2015)

Showed in the publication Carrion Gordon & Chaczko (2017) there are options in the management of Medical Records. There are many benefits to open datasets. However, privacy concerns have blocked the widespread creation of open health data. The opportunity to access to documented methods and case studies for the creation of public-use health data.

The Image and Data Archive (IDA) data sets was generated by Alzheimer's Disease Neuroimaging Initiative (ADNI). ADNI initiative also provides tools and resources for identifying, integrating, searching, visualizing and sharing a diverse range of neuroscience data, helping facilitate collaborations between scientists worldwide. There are many benefits to open datasets. However, privacy concerns have concerned the widespread creation of open health data.

It was possible to ensure that the probability of re-identification for a large longitudinal dataset was acceptably low when it was released for a global user community in support of an analytics competition.



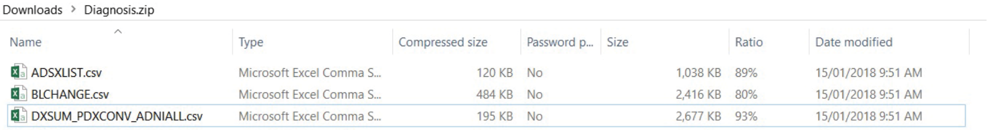
Investigators around the world rely on the IDA to safeguard and share their neuroscience data

30,593,290
Downloads

90,204
Users

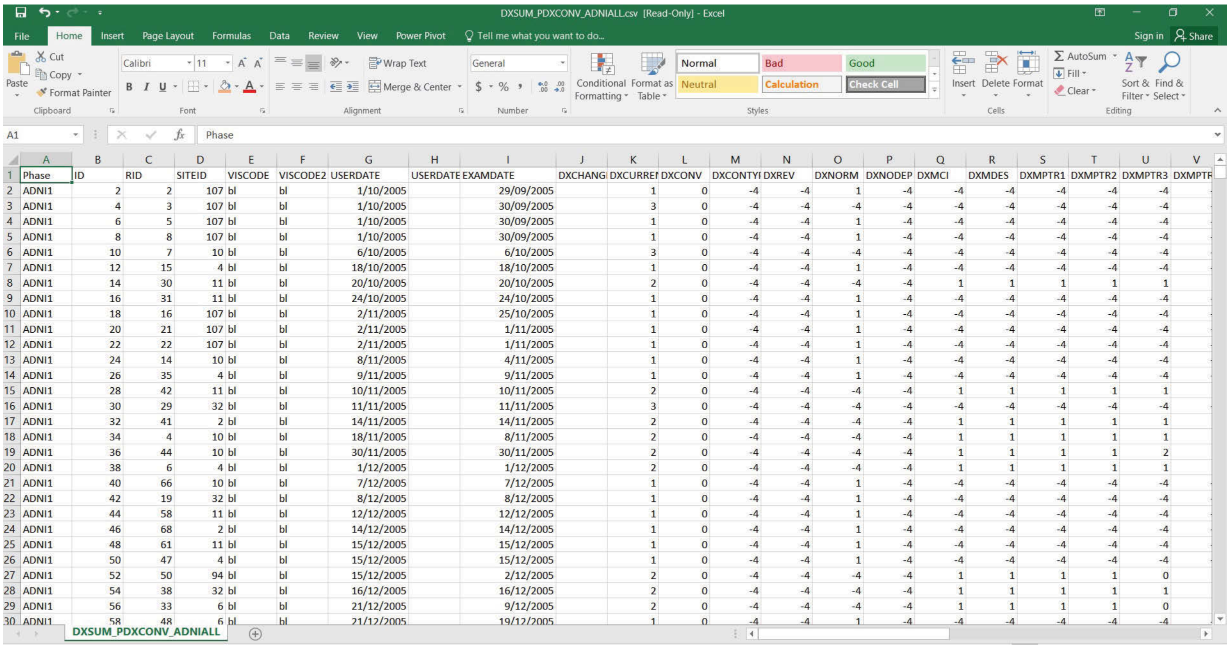
912,755
Uploads

Figure 6.5.: Utilization of ADNI



Name	Type	Compressed size	Password p...	Size	Ratio	Date modified
ADSXLIST.csv	Microsoft Excel Comma S...	120 KB	No	1,038 KB	89%	15/01/2018 9:51 AM
BLCHANGE.csv	Microsoft Excel Comma S...	484 KB	No	2,416 KB	80%	15/01/2018 9:51 AM
DXSUM_PDXCONV_ADNIALL.csv	Microsoft Excel Comma S...	195 KB	No	2,677 KB	93%	15/01/2018 9:51 AM

Figure 6.6.: Examples ADNI Databases



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	Phase	ID	RID	SITEID	VISCODE	VISCODE2	USERDATE	USERDATE	EXAMDATE	DXCHANG	DXCURREI	DXCONV	DXCONTYI	DXREV	DXNORM	DXNODEP	DXMCI	DXMDES	DXMPTR1	DXMPTR2	DXMPTR3	DXMPTR
2	ADNI1	2	2	107	bl	bl	1/10/2005		29/09/2005		1	0	-4	-4	1	-4	-4	-4	-4	-4	-4	
3	ADNI1	4	3	107	bl	bl	1/10/2005		30/09/2005		3	0	-4	-4	-4	-4	-4	-4	-4	-4	-4	
4	ADNI1	6	5	107	bl	bl	1/10/2005		30/09/2005		1	0	-4	-4	1	-4	-4	-4	-4	-4	-4	
5	ADNI1	8	8	107	bl	bl	1/10/2005		30/09/2005		1	0	-4	-4	1	-4	-4	-4	-4	-4	-4	
6	ADNI1	10	7	10	bl	bl	6/10/2005		6/10/2005		3	0	-4	-4	-4	-4	-4	-4	-4	-4	-4	
7	ADNI1	12	15	4	bl	bl	18/10/2005		18/10/2005		1	0	-4	-4	1	-4	-4	-4	-4	-4	-4	
8	ADNI1	14	30	11	bl	bl	20/10/2005		20/10/2005		2	0	-4	-4	-4	-4	1	1	1	1	1	
9	ADNI1	16	31	11	bl	bl	24/10/2005		24/10/2005		1	0	-4	-4	1	-4	-4	-4	-4	-4	-4	
10	ADNI1	18	16	107	bl	bl	2/11/2005		25/10/2005		1	0	-4	-4	1	-4	-4	-4	-4	-4	-4	
11	ADNI1	20	21	107	bl	bl	2/11/2005		1/11/2005		1	0	-4	-4	1	-4	-4	-4	-4	-4	-4	
12	ADNI1	22	22	107	bl	bl	2/11/2005		1/11/2005		1	0	-4	-4	1	-4	-4	-4	-4	-4	-4	
13	ADNI1	24	14	10	bl	bl	8/11/2005		4/11/2005		1	0	-4	-4	1	-4	-4	-4	-4	-4	-4	
14	ADNI1	26	35	4	bl	bl	9/11/2005		9/11/2005		1	0	-4	-4	1	-4	-4	-4	-4	-4	-4	
15	ADNI1	28	42	11	bl	bl	10/11/2005		10/11/2005		2	0	-4	-4	-4	-4	1	1	1	1	1	
16	ADNI1	30	29	32	bl	bl	11/11/2005		11/11/2005		3	0	-4	-4	-4	-4	-4	-4	-4	-4	-4	
17	ADNI1	32	41	2	bl	bl	14/11/2005		14/11/2005		2	0	-4	-4	-4	-4	1	1	1	1	1	
18	ADNI1	34	4	10	bl	bl	18/11/2005		8/11/2005		2	0	-4	-4	-4	-4	1	1	1	1	1	
19	ADNI1	36	44	10	bl	bl	30/11/2005		30/11/2005		2	0	-4	-4	-4	-4	1	1	1	1	2	
20	ADNI1	38	6	4	bl	bl	1/12/2005		1/12/2005		2	0	-4	-4	-4	-4	1	1	1	1	1	
21	ADNI1	40	66	10	bl	bl	7/12/2005		7/12/2005		1	0	-4	-4	1	-4	-4	-4	-4	-4	-4	
22	ADNI1	42	19	32	bl	bl	8/12/2005		8/12/2005		1	0	-4	-4	1	-4	-4	-4	-4	-4	-4	
23	ADNI1	44	58	11	bl	bl	12/12/2005		12/12/2005		1	0	-4	-4	1	-4	-4	-4	-4	-4	-4	
24	ADNI1	46	68	2	bl	bl	14/12/2005		14/12/2005		1	0	-4	-4	1	-4	-4	-4	-4	-4	-4	
25	ADNI1	48	61	11	bl	bl	15/12/2005		15/12/2005		1	0	-4	-4	1	-4	-4	-4	-4	-4	-4	
26	ADNI1	50	47	4	bl	bl	15/12/2005		15/12/2005		1	0	-4	-4	1	-4	-4	-4	-4	-4	-4	
27	ADNI1	52	50	94	bl	bl	15/12/2005		2/12/2005		2	0	-4	-4	-4	-4	1	1	1	1	0	
28	ADNI1	54	38	32	bl	bl	16/12/2005		16/12/2005		2	0	-4	-4	-4	-4	1	1	1	1	1	
29	ADNI1	56	33	6	bl	bl	21/12/2005		9/12/2005		2	0	-4	-4	-4	-4	1	1	1	1	0	
30	ADNI1	58	48	6	bl	bl	21/12/2005		19/12/2005		1	0	-4	-4	1	-4	-4	-4	-4	-4	-4	

Figure 6.7.: Fields Generated by ADNI

6.1.2. Government Records

The Dataset was released a corpus of 1 million documents are freely available for research and freely redistributed. These documents were obtained by performing searches for words randomly chosen from the Unix dictionary, numbers randomly chosen between 1 and 1 million and randomized combinations of the two, for documents of specified file types that resided on web servers in the .gov domain using the Yahoo and Google search engines.

Each file in the corpus is presented as a numbered file with a file extension (e.g. 0000001.jpg). The file extension is typically the file extension that was provided to us when the file was downloaded.

The corpus is available in several ways:Garfinkel et al. (2009)

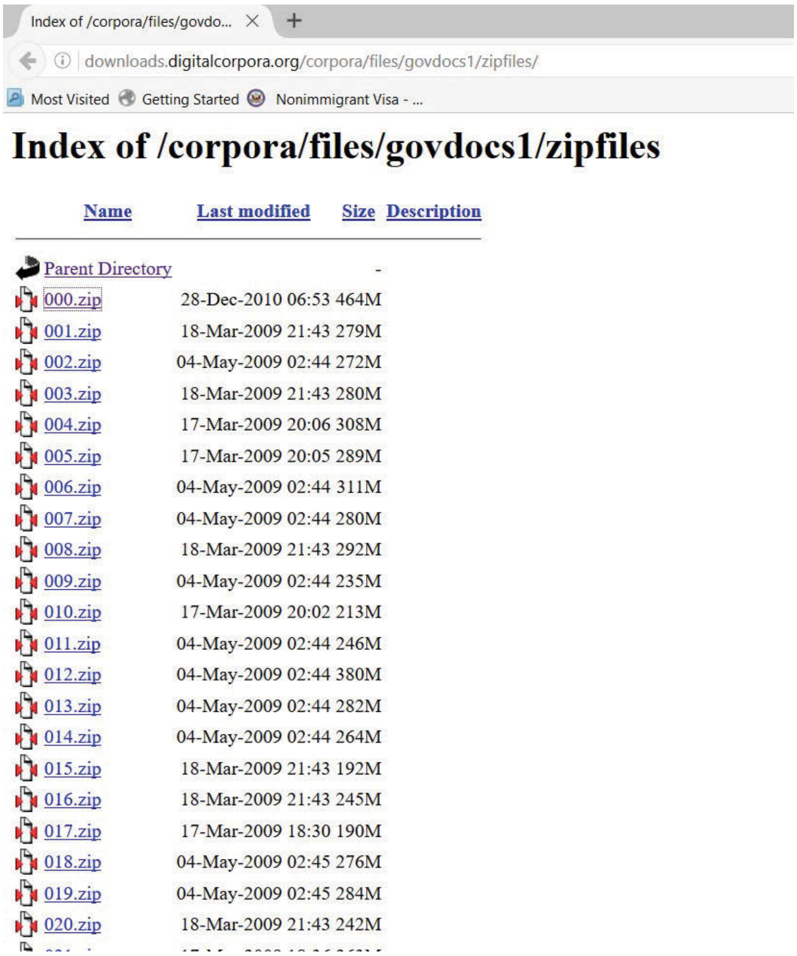
- A set of 1000 directories, with 1000 files in each directory, downloadable from our server at http://downloads.digitalcorpora.org/corpora/files/govdocs1/.

- A set of 1000 ZIP files, each with 1000 files, downloadable from our server at http://downloads.digitalcorpora.org/corpora/files/govdocs1/zipfiles/.

- A a tar file with 109,282 JPEG pictures from the govdocs1m corpus.

- As a set of 10 subset “threads” (subset0.zip through subset9.zip), each one containing 1000 randomly chosen documents.

- These subsets were specifically created for to facilitate pilot studies and student research projects with the rationale that it’s easier to work with 1000 files than with 1 million. Students are encouraged to use one subset for development and another subset for testing. A contextual feature list of data from the digitalcorpora can be found at website.

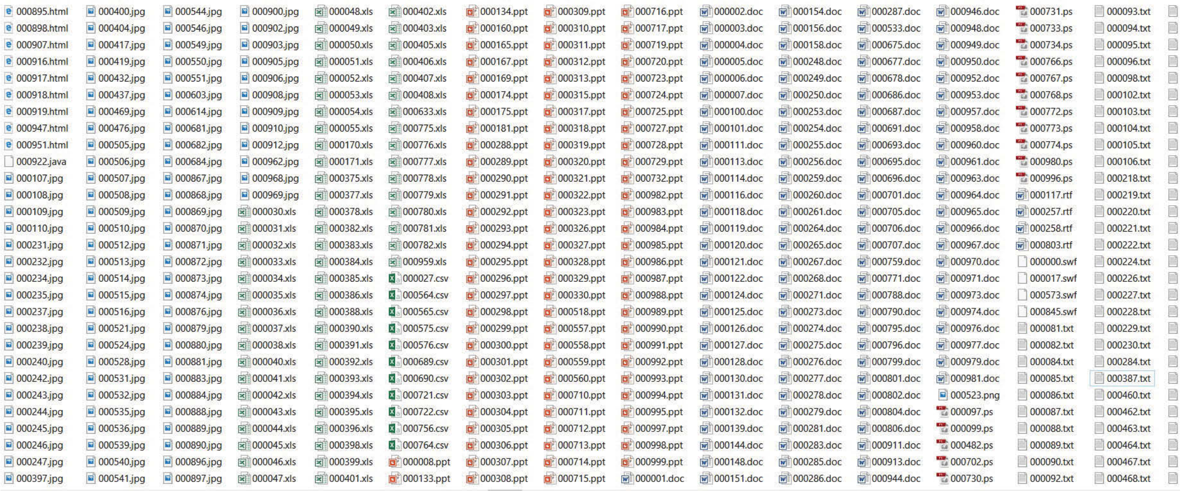


Index of /corpora/files/govdo...
downloads.digitalcorpora.org/corpora/files/govdocs1/zipfiles/
Most Visited
Getting Started
Nonimmigrant Visa - ...

Index of /corpora/files/govdocs1/zipfiles

Name	Last modified	Size	Description
Parent Directory		-	
000.zip	28-Dec-2010 06:53	464M	
001.zip	18-Mar-2009 21:43	279M	
002.zip	04-May-2009 02:44	272M	
003.zip	18-Mar-2009 21:43	280M	
004.zip	17-Mar-2009 20:06	308M	
005.zip	17-Mar-2009 20:05	289M	
006.zip	04-May-2009 02:44	311M	
007.zip	04-May-2009 02:44	280M	
008.zip	18-Mar-2009 21:43	292M	
009.zip	04-May-2009 02:44	235M	
010.zip	17-Mar-2009 20:02	213M	
011.zip	04-May-2009 02:44	246M	
012.zip	04-May-2009 02:44	380M	
013.zip	04-May-2009 02:44	282M	
014.zip	04-May-2009 02:44	264M	
015.zip	18-Mar-2009 21:43	192M	
016.zip	18-Mar-2009 21:43	245M	
017.zip	17-Mar-2009 18:30	190M	
018.zip	04-May-2009 02:45	276M	
019.zip	04-May-2009 02:45	284M	
020.zip	18-Mar-2009 21:43	242M	

Figure 6.8.: Gov1 Entries



000895.html
000898.html
000907.html
000916.html
000917.html
000918.html
000919.html
000947.html
000951.html
000922.java
000107.jpg
000108.jpg
000109.jpg
000110.jpg
000231.jpg
000232.jpg
000234.jpg
000235.jpg
000237.jpg
000238.jpg
000239.jpg
000240.jpg
000242.jpg
000243.jpg
000244.jpg
000245.jpg
000246.jpg
000247.jpg
000397.jpg
000400.jpg
000404.jpg
000417.jpg
000419.jpg
000432.jpg
000437.jpg
000469.jpg
000476.jpg
000505.jpg
000506.jpg
000507.jpg
000508.jpg
000509.jpg
000510.jpg
000512.jpg
000513.jpg
000514.jpg
000515.jpg
000516.jpg
000521.jpg
000524.jpg
000528.jpg
000531.jpg
000532.jpg
000535.jpg
000536.jpg
000539.jpg
000540.jpg
000541.jpg
000544.jpg
000546.jpg
000549.jpg
000550.jpg
000551.jpg
000603.jpg
000614.jpg
000681.jpg
000682.jpg
000684.jpg
000867.jpg
000868.jpg
000869.jpg
000870.jpg
000871.jpg
000872.jpg
000873.jpg
000874.jpg
000876.jpg
000879.jpg
000880.jpg
000881.jpg
000883.jpg
000884.jpg
000888.jpg
000889.jpg
000890.jpg
000896.jpg
000897.jpg
000900.jpg
000902.jpg
000903.jpg
000905.jpg
000906.jpg
000908.jpg
000909.jpg
000910.jpg
000912.jpg
000962.jpg
000968.jpg
000969.jpg
000030.xls
000031.xls
000032.xls
000033.xls
000034.xls
000035.xls
000036.xls
000037.xls
000038.xls
000040.xls
000041.xls
000042.xls
000043.xls
000044.xls
000045.xls
000046.xls
000047.xls
000048.xls
000049.xls
000050.xls
000051.xls
000052.xls
000053.xls
000054.xls
000055.xls
000170.xls
000171.xls
000375.xls
000377.xls
000378.xls
000382.xls
000383.xls
000384.xls
000385.xls
000386.xls
000388.xls
000390.xls
000391.xls
000392.xls
000393.xls
000394.xls
000395.xls
000396.xls
000398.xls
000399.xls
000401.xls
000402.xls
000403.xls
000405.xls
000406.xls
000407.xls
000408.xls
000633.xls
000775.xls
000776.xls
000777.xls
000778.xls
000779.xls
000780.xls
000781.xls
000782.xls
000959.xls
000027.csv
000564.csv
000565.csv
000575.csv
000576.csv
000689.csv
000690.csv
000721.csv
000722.csv
000756.csv
000764.csv
000008.ppt
000133.ppt
000134.ppt
000160.ppt
000165.ppt
000167.ppt
000169.ppt
000174.ppt
000175.ppt
000181.ppt
000288.ppt
000289.ppt
000290.ppt
000291.ppt
000292.ppt
000293.ppt
000294.ppt
000295.ppt
000296.ppt
000297.ppt
000298.ppt
000299.ppt
000300.ppt
000301.ppt
000302.ppt
000303.ppt
000304.ppt
000305.ppt
000306.ppt
000307.ppt
000308.ppt
000309.ppt
000310.ppt
000311.ppt
000312.ppt
000313.ppt
000315.ppt
000317.ppt
000318.ppt
000319.ppt
000320.ppt
000321.ppt
000322.ppt
000323.ppt
000326.ppt
000327.ppt
000328.ppt
000329.ppt
000330.ppt
000518.ppt
000557.ppt
000558.ppt
000559.ppt
000560.ppt
000710.ppt
000711.ppt
000712.ppt
000713.ppt
000714.ppt
000715.ppt
000716.ppt
000717.ppt
000719.ppt
000720.ppt
000723.ppt
000724.ppt
000725.ppt
000727.ppt
000728.ppt
000729.ppt
000732.ppt
000982.ppt
000983.ppt
000984.ppt
000985.ppt
000986.ppt
000987.ppt
000988.ppt
000989.ppt
000990.ppt
000991.ppt
000992.ppt
000993.ppt
000994.ppt
000995.ppt
000997.ppt
000998.ppt
000999.ppt
000001.doc
000002.doc
000003.doc
000004.doc
000005.doc
000006.doc
000007.doc
000100.doc
000101.doc
000111.doc
000113.doc
000114.doc
000116.doc
000118.doc
000119.doc
000120.doc
000121.doc
000122.doc
000124.doc
000125.doc
000126.doc
000127.doc
000128.doc
000130.doc
000131.doc
000132.doc
000139.doc
000144.doc
000148.doc
000151.doc
000154.doc
000156.doc
000158.doc
000248.doc
000249.doc
000250.doc
000253.doc
000254.doc
000255.doc
000256.doc
000259.doc
000260.doc
000261.doc
000264.doc
000265.doc
000267.doc
000268.doc
000271.doc
000273.doc
000274.doc
000275.doc
000276.doc
000277.doc
000278.doc
000279.doc
000281.doc
000283.doc
000285.doc
000286.doc
000287.doc
000533.doc
000675.doc
000677.doc
000678.doc
000686.doc
000687.doc
000691.doc
000693.doc
000695.doc
000696.doc
000701.doc
000705.doc
000706.doc
000707.doc
000759.doc
000771.doc
000788.doc
000790.doc
000795.doc
000796.doc
000799.doc
000801.doc
000802.doc
000804.doc
000806.doc
000911.doc
000913.doc
000944.doc
000946.doc
000948.doc
000949.doc
000950.doc
000952.doc
000953.doc
000957.doc
000958.doc
000960.doc
000961.doc
000963.doc
000964.doc
000965.doc
000966.doc
000967.doc
000970.doc
000971.doc
000973.doc
000974.doc
000976.doc
000977.doc
000979.doc
000981.doc
000523.png
000097.ps
000099.ps
000482.ps
000702.ps
000730.ps
000731.ps
000733.ps
000734.ps
000766.ps
000767.ps
000768.ps
000772.ps
000773.ps
000774.ps
000980.ps
000996.ps
000117.rtf
000257.rtf
000258.rtf
000803.rtf
000000.swf
000017.swf
000573.swf
000845.swf
000081.txt
000082.txt
000084.txt
000085.txt
000086.txt
000087.txt
000088.txt
000089.txt
000090.txt
000092.txt
000093.txt
000094.txt
000095.txt
000096.txt
000098.txt
000102.txt
000103.txt
000104.txt
000105.txt
000106.txt
000218.txt
000219.txt
000220.txt
000221.txt
000222.txt
000224.txt
000226.txt
000227.txt
000228.txt
000229.txt
000230.txt
000284.txt
000387.txt
000460.txt
000462.txt
000463.txt
000464.txt
000467.txt
000468.txt

Figure 6.9.: Example Gov1 Dataset

6.2. Validation and Verification

After the process of Modelling and Design, the application with the third part application, the last part to cover the complete cycle is to validate and verify the model. Through the methodology, the description for the validation divided it in two manners, qualitative and quantitative method. To prove the efficiency of the model and after diverse analysis, benchmarking is the best alternative for the validation. The reasons are the following:

1. DHP model is a dynamic proposal, so the validation has to be dynamic as well, with the analysis of the competitors and in the context of each case study.
2. Benchmarking as a method for verification and validation, through the comparison of results between cases or competitors, in business context, explain the behavior of the dataset that has to be validated.
3. Through Benchmarking the model can certified in terms of Level of Success. The base for the adaptation of the model are to principal factor to cover the needs from the user.

The graphics specified in the Figure 6.10, Figure 6.11 and Figure 6.12 explain the reports generated with the original tool. The source is the Risk Assessment from Western Australia Government.

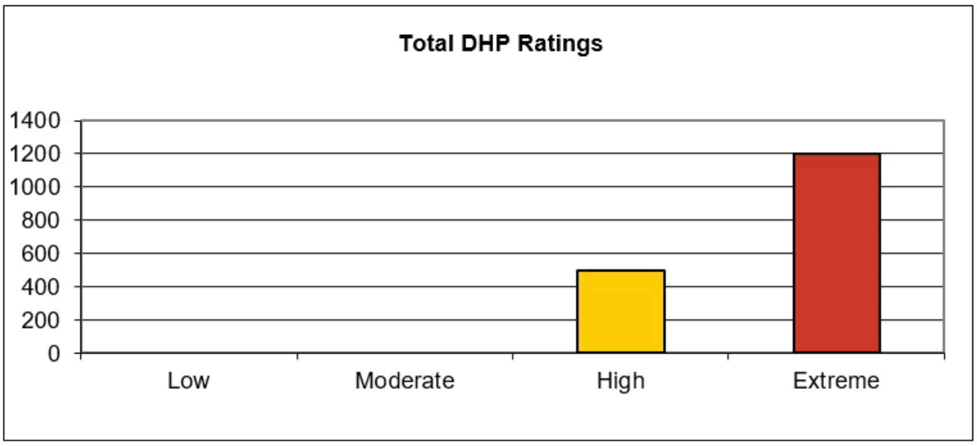


Figure 6.10.: DHP Ratings

The plot area through the concepts define the area of the principal stakeholders. In the Figure 6.11,the IoT community is the summary of the relationship with the case studies in the DHP situation.

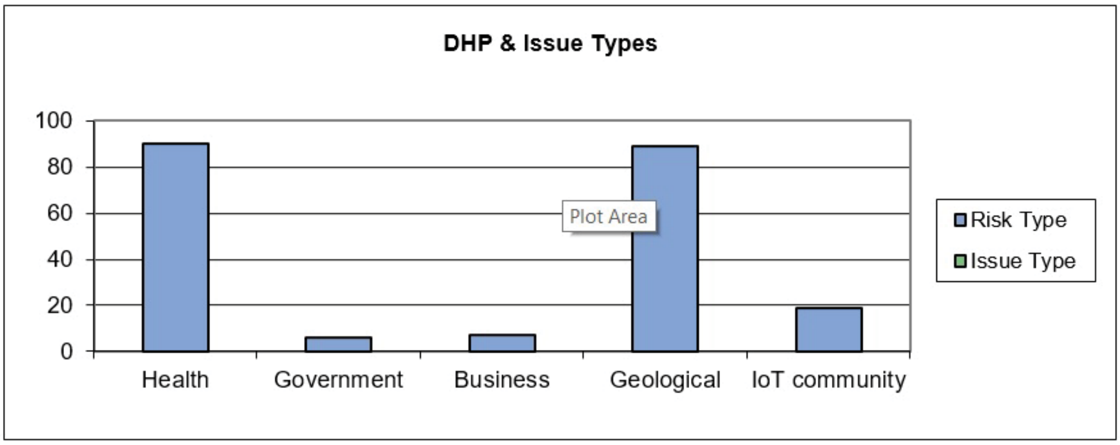
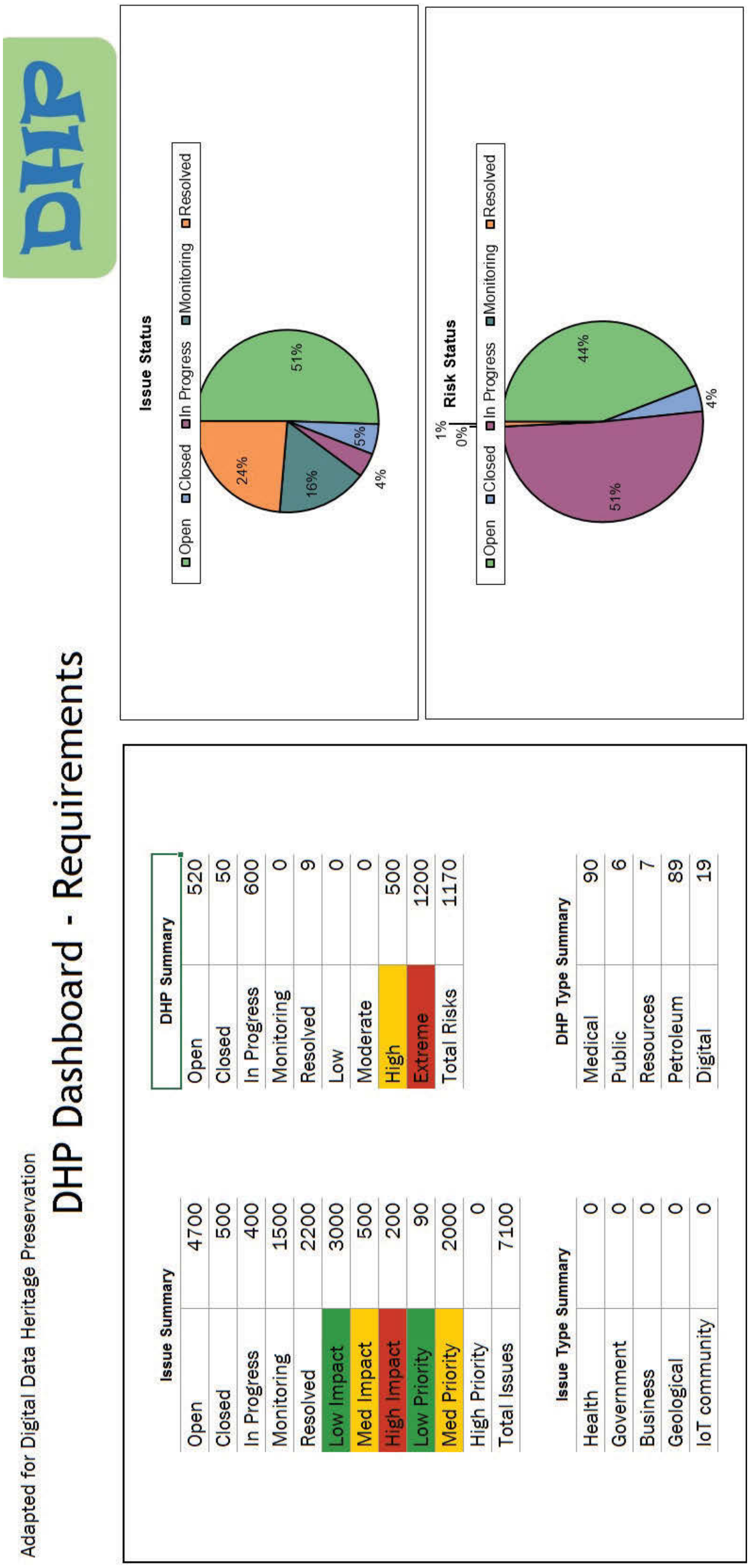


Figure 6.11.: DHP Issue and Types

.



Adapted for Digital Data Heritage Preservation

DHP Dashboard - Requirements

Issue Summary	
Open	4700
Closed	500
In Progress	400
Monitoring	1500
Resolved	2200
Low Impact	3000
Med Impact	500
High Impact	200
Low Priority	90
Med Priority	2000
High Priority	0
Total Issues	7100

DHP Summary	
Open	520
Closed	50
In Progress	600
Monitoring	0
Resolved	9
Low	0
Moderate	0
High	500
Extreme	1200
Total Risks	1170

Issue Type Summary	
Health	0
Government	0
Business	0
Geological	0
IoT community	0

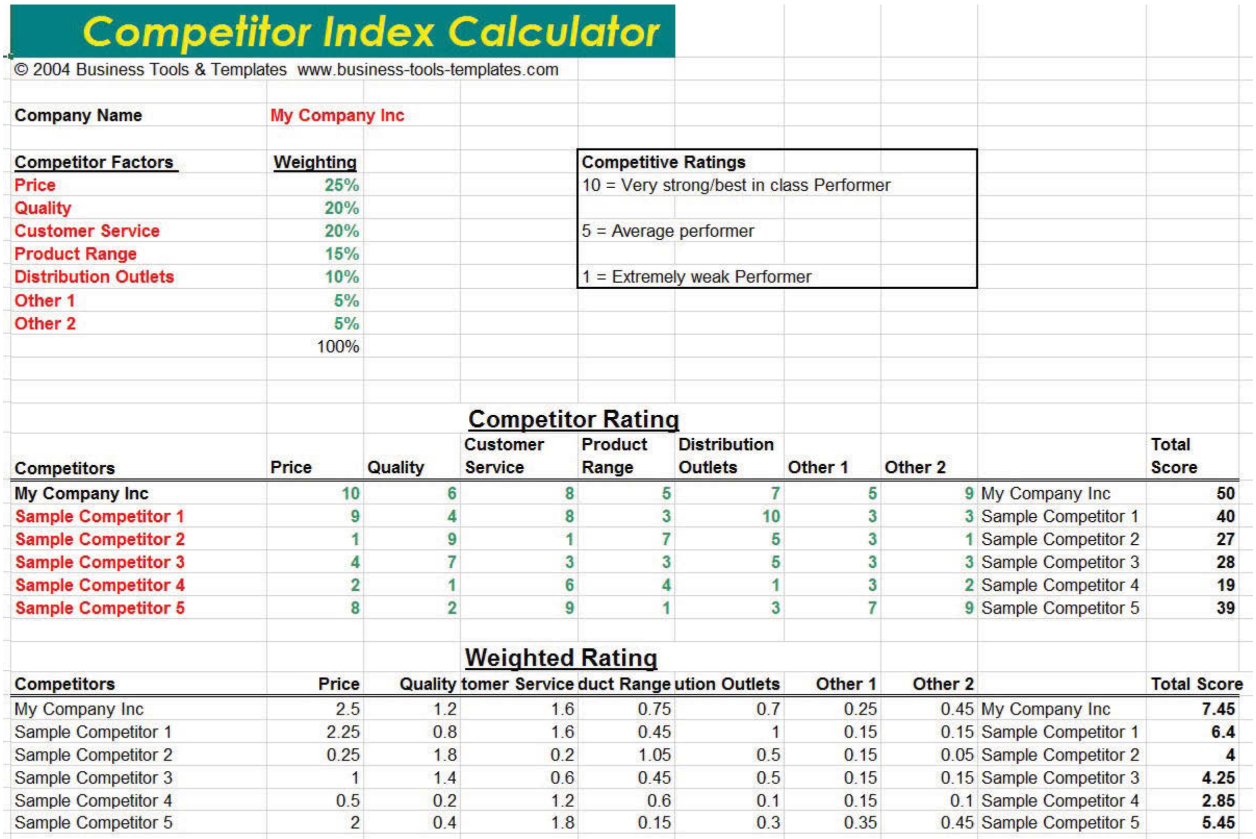
DHP Type Summary	
Medical	90
Public	6
Resources	7
Petroleum	89
Digital	19

Figure 6.12.: Validation Dashboard

6.2.1. Traceability Matrix

This way of validation is inspired in the Competitor Index Calculator. Understanding the competitive strengths and weaknesses is a prerequisite to developing a winning strategy.

The Competitor Rating Calculator, adapted as DHP Model Traceability Matrix allows to rate the offering against that of up to five main competitors or cases. Competitors can be rated on the basis of the most important factors in the purchasing decision.



Competitor Index Calculator

© 2004 Business Tools & Templates www.business-tools-templates.com

Company Name My Company Inc

Competitor Factors	Weighting
Price	25%
Quality	20%
Customer Service	20%
Product Range	15%
Distribution Outlets	10%
Other 1	5%
Other 2	5%
	100%

Competitive Ratings
10 = Very strong/best in class Performer
5 = Average performer
1 = Extremely weak Performer

Competitor Rating

Competitors	Price	Quality	Customer Service	Product Range	Distribution Outlets	Other 1	Other 2		Total Score
My Company Inc	10	6	8	5	7	5	9	My Company Inc	50
Sample Competitor 1	9	4	8	3	10	3	3	Sample Competitor 1	40
Sample Competitor 2	1	9	1	7	5	3	1	Sample Competitor 2	27
Sample Competitor 3	4	7	3	3	5	3	3	Sample Competitor 3	28
Sample Competitor 4	2	1	6	4	1	3	2	Sample Competitor 4	19
Sample Competitor 5	8	2	9	1	3	7	9	Sample Competitor 5	39

Weighted Rating

Competitors	Price	Quality	tomer Service	duct Range	ution Outlets	Other 1	Other 2		Total Score
My Company Inc	2.5	1.2	1.6	0.75	0.7	0.25	0.45	My Company Inc	7.45
Sample Competitor 1	2.25	0.8	1.6	0.45	1	0.15	0.15	Sample Competitor 1	6.4
Sample Competitor 2	0.25	1.8	0.2	1.05	0.5	0.15	0.05	Sample Competitor 2	4
Sample Competitor 3	1	1.4	0.6	0.45	0.5	0.15	0.15	Sample Competitor 3	4.25
Sample Competitor 4	0.5	0.2	1.2	0.6	0.1	0.15	0.1	Sample Competitor 4	2.85
Sample Competitor 5	2	0.4	1.8	0.15	0.3	0.35	0.45	Sample Competitor 5	5.45

Figure 6.13.: Competitor Index Calculator. Source: Business Tool Templates

Bar and radar charts are automatically generated based on the ratings input as it is shown in Figure 6.14, Figure 6.15, Figure 6.16, Figure 6.17, Figure 6.18 and Figure 6.19.

The ratings are then adjusted on the basis of the importance/weighting you have given to each of the factors in the purchasing decision. The weighted ratings are automatically calculated and bar and radar charts are generated.

A Total Competitive Score is also calculated and a corresponding Bar Chart is produced. A Total Adjusted Competitive Score is also calculated and a corresponding Bar Chart is produced.

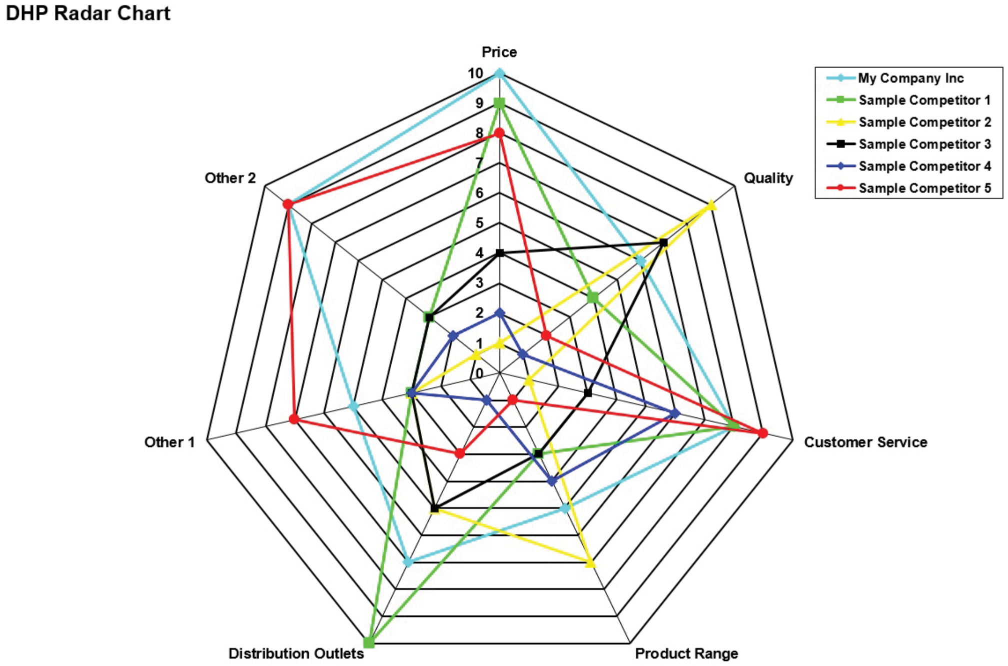


Figure 6.14.: DHP Radar Chart

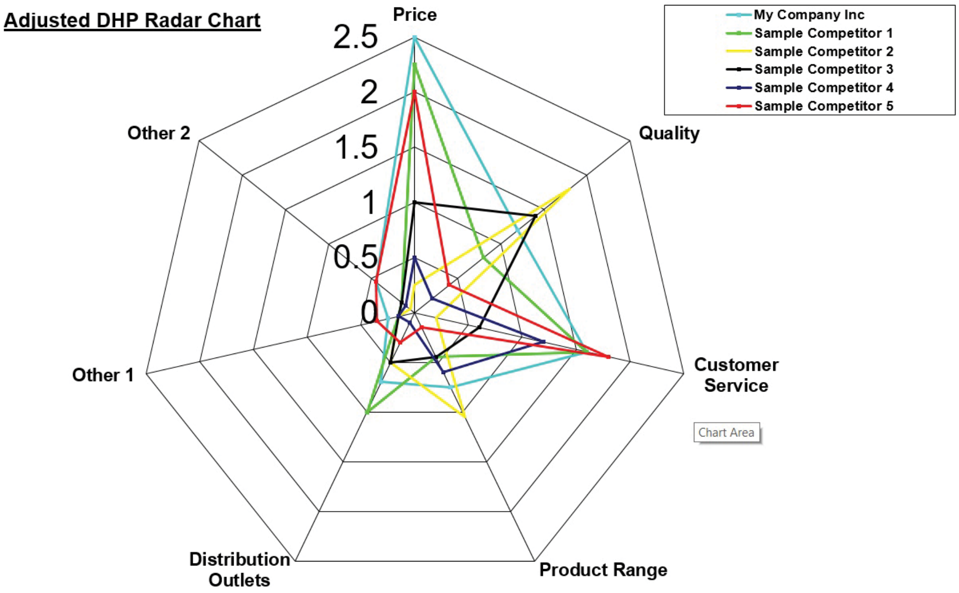


Figure 6.15.: Adjusted DHP Radar Chart

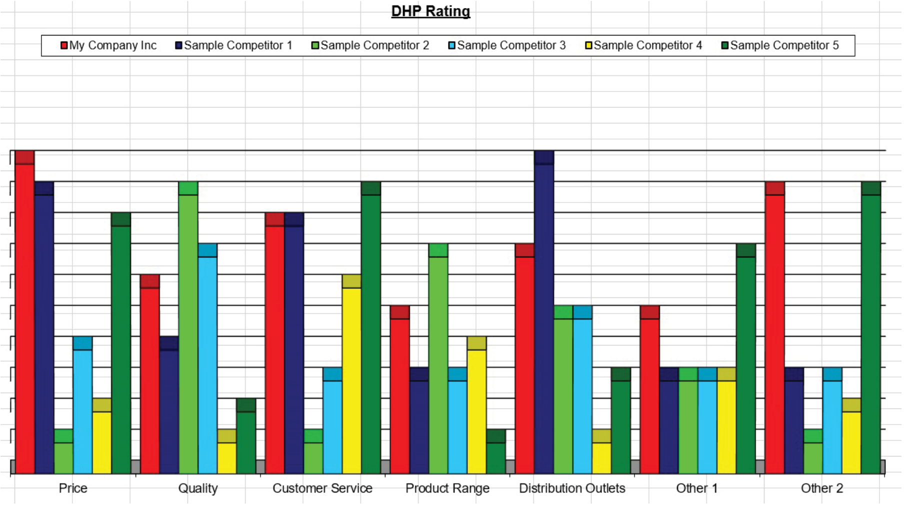


Figure 6.16.: DHP Rating

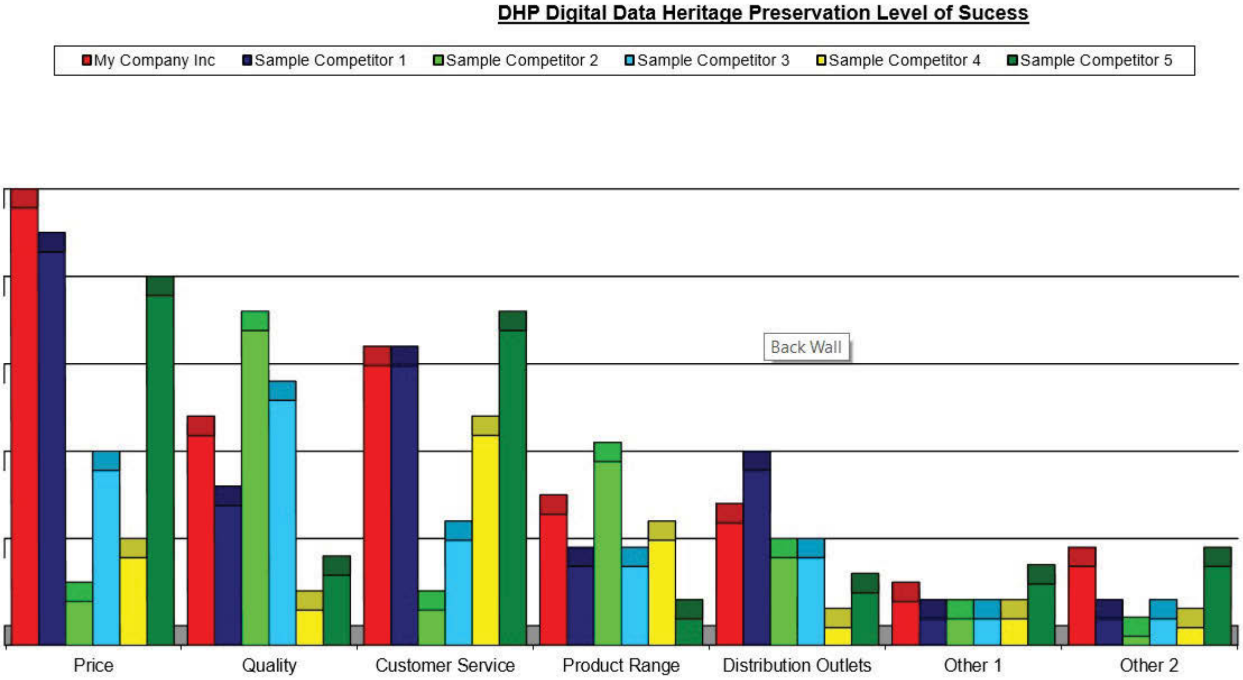


Figure 6.17.: DHP Level of Success

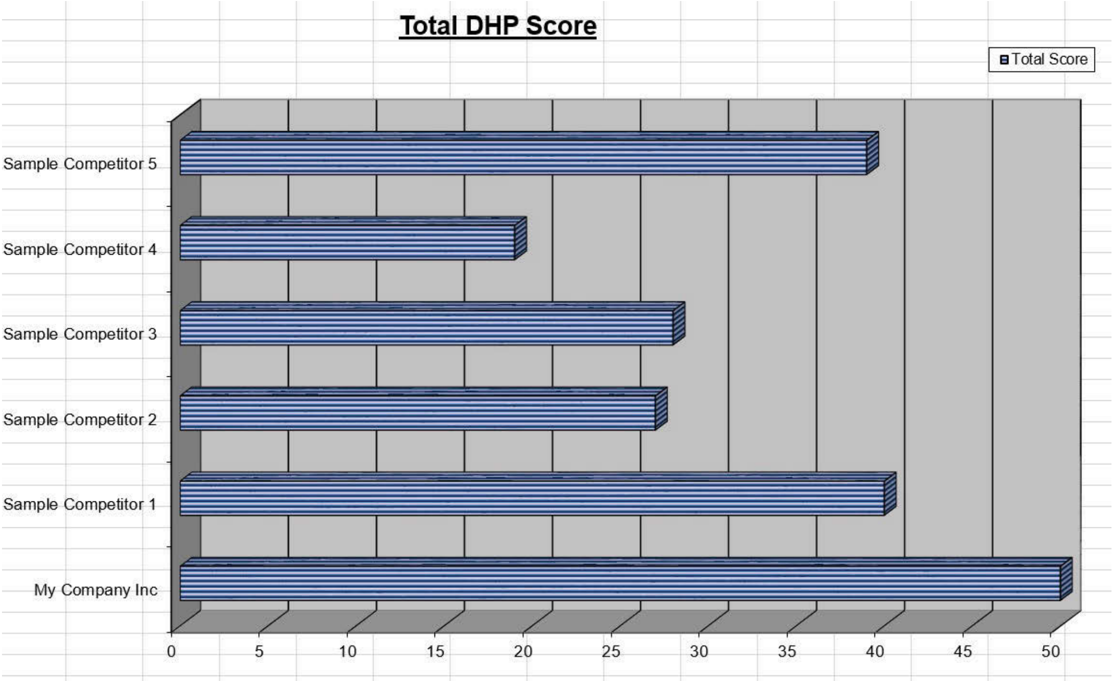


Figure 6.18.: Total DHP Score

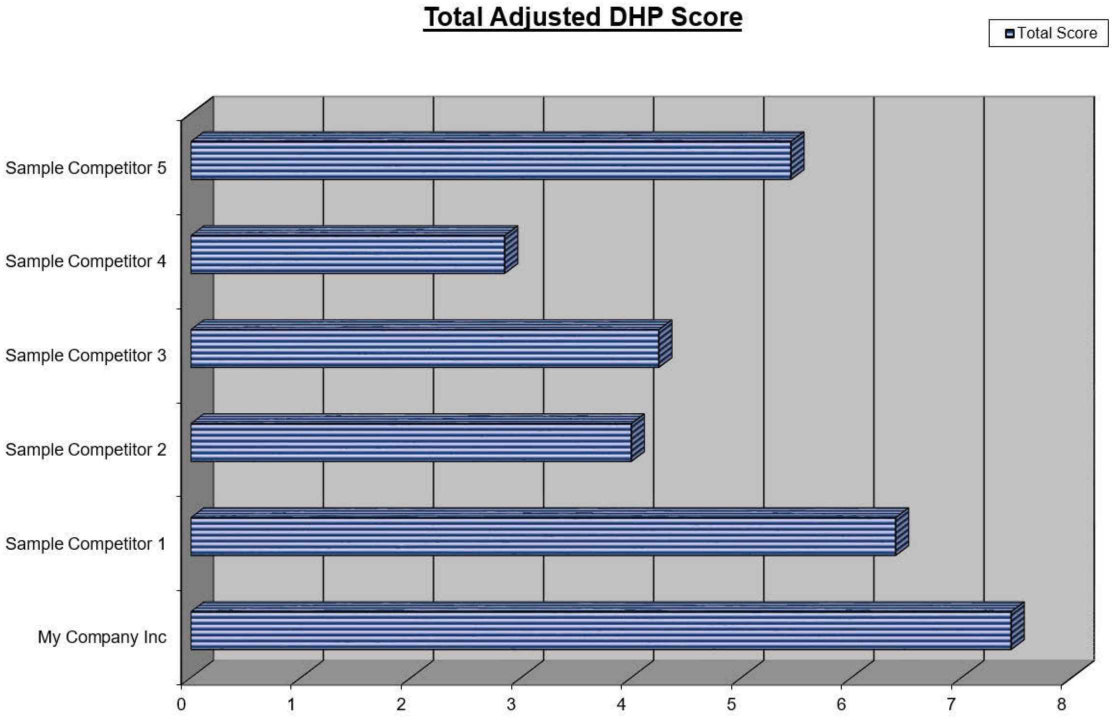


Figure 6.19.: Total Adjusted DHP Score

The following data should be entered:

Edit Company Name, Edit the Competitor Factors you consider most important and the Weighting for each in the Purchasing Decision. Start with the most important factor and assign the weighting to each. Note that the weighting must add up to a total of 100 . It is possible to edit the Name of each Competitor (you may analyse up to 5 competitors)

For each Competitor Factor rate each competitor on a scale of 1 to 10 as follows:

Competitive Ratings 10 = Very strong/best in class Performer

5 = Average performer

1 = Extremely weak Performer

The weighting ratings are calculated for each competitor based on the weighting of each factor in the purchasing decision

A Total Competitive Score is calculated and Bar and Radar Charts are automatically generated.

6.2.2. DHP Analysis Tool

In the "Value Drivers" tab, select & define the value drivers or key buying criteria for the market segment you wish to analyze. In the "Weighting" tab, weigh each of the value drivers or the key buying criteria according to the priorities of your target customers.

Using the "Value Rankings" tab, rank your products versus your competitors for each of the value drivers or key buying criteria.

In the "Price & Market Share" tab, input a pricing rank and estimated market share for each company.

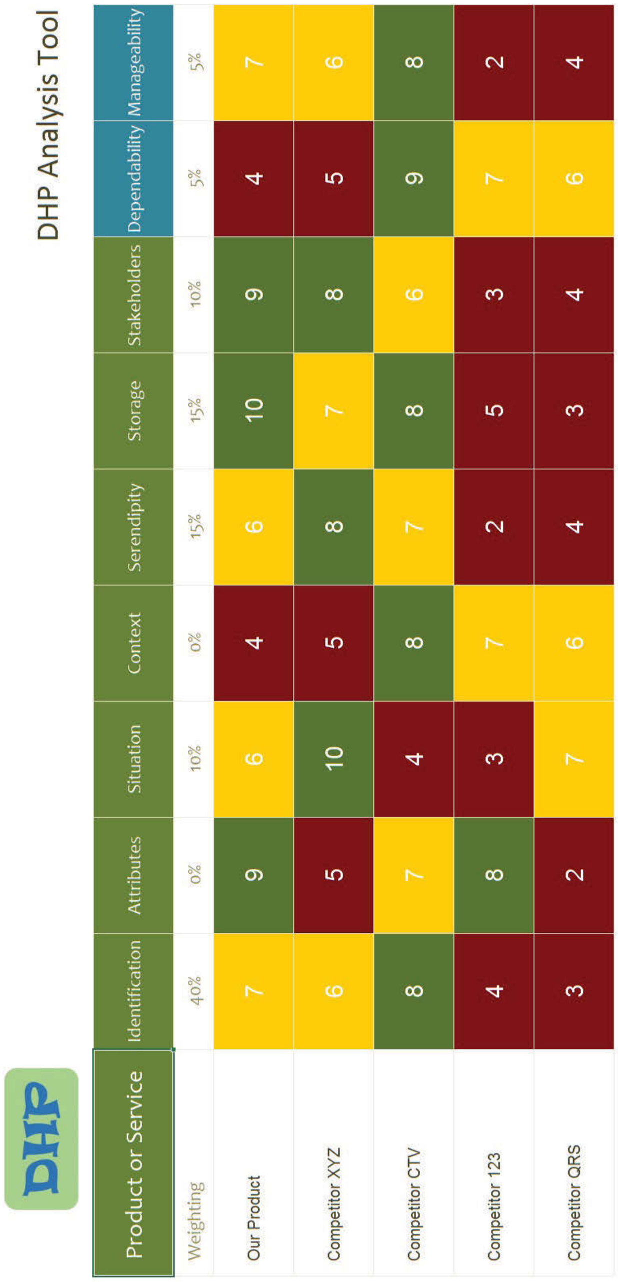
Evaluate how each competitor is differentiated in the "Differentiation Chart" tab, where the rankings from the value driver scoring are displayed visually.

View the "Positioning Map" tab to determine your market position In relation to your competitors with regard to value, price and market share.

Provide a brief description of your product and your top 4 competitors in the additional tabs below.

The adaptation of the tool is called DHP Analysis Tool as is shown in the Figure 6.20. The columns are determined by the same attributes from the original DHP Model.

It is also added two columns with the details of the Requirements, Dependability and Manageability.



DHP Analysis Tool

Product or Service	Identification	Attributes	Situation	Context	Serendipity	Storage	Stakeholders	Dependability	Manageability
Weighting	40%	0%	10%	0%	15%	15%	10%	5%	5%
Our Product	7	9	6	4	6	10	9	4	7
Competitor XYZ	6	5	10	5	8	7	8	5	6
Competitor CTV	8	7	4	8	7	8	6	9	8
Competitor 123	4	8	3	7	2	5	3	7	2
Competitor QRS	3	2	7	6	4	3	4	6	4

Figure 6.20.: DHP Analysis Tool

The representation as spider representation define the differentiation by value driven. It is a suggested representation because it gives the necessary details. It is shown in Figure 6.21.

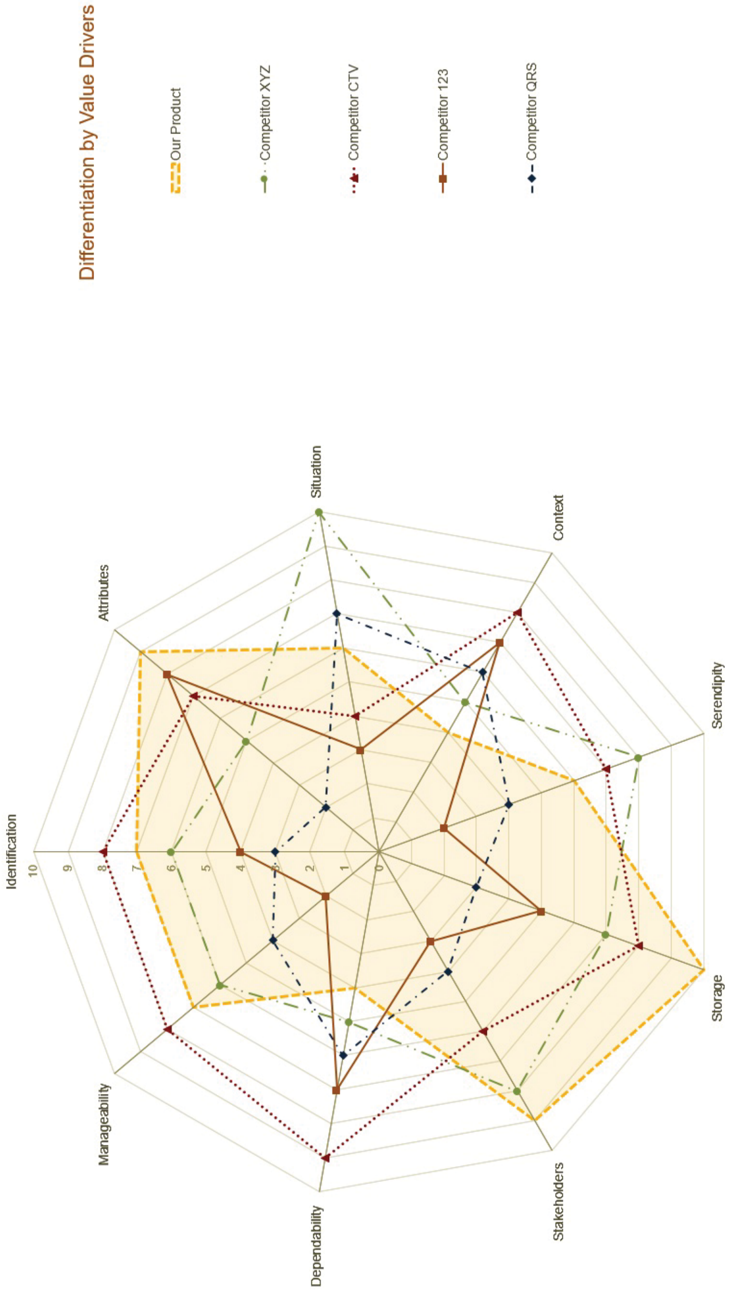


Figure 6.21.: Differentiation by Value Driven

There is another manner of representation. The value proposition represented in the Figure 6.22 shows the weight of the competitor or case studies and solutions related with DHP framework.

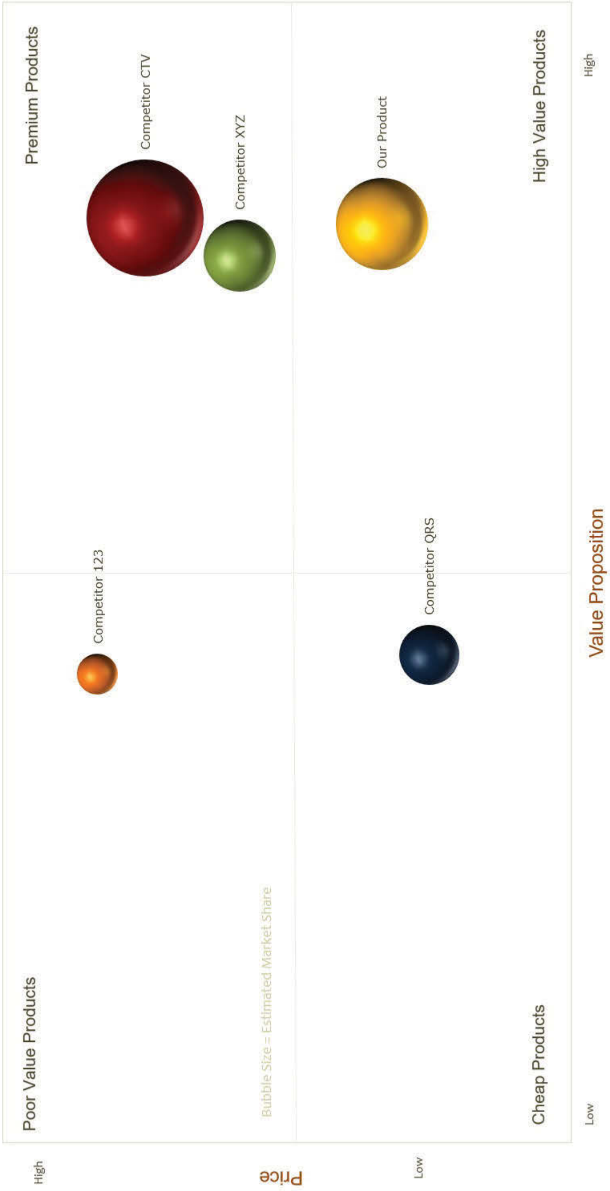
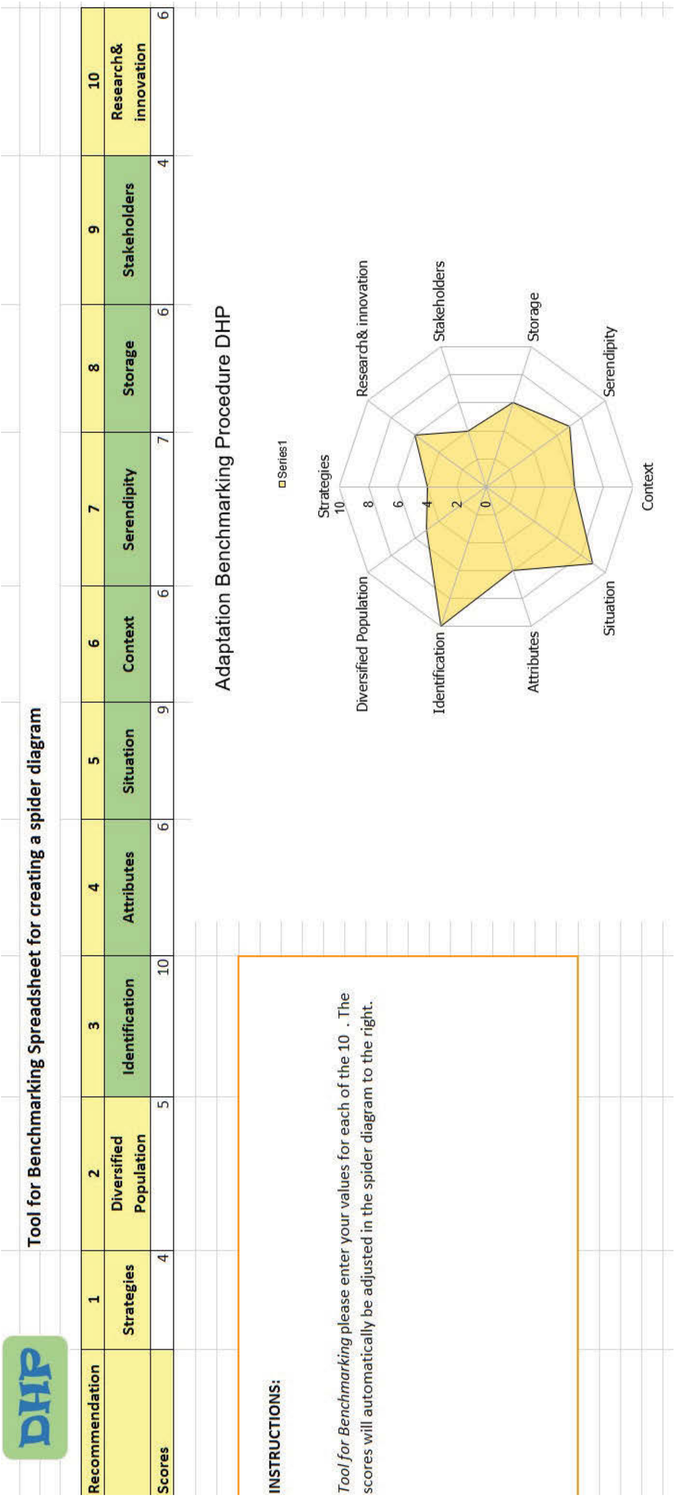


Figure 6.22.: Value Proposition

6.2.3. DHP Compliance Matrix

This is the Benchmarking with spider diagram representation. The scores define the form of the result and expresses in correct manner the tendency of the attributes and the relationships between the objects and concept. It is clarified in the Figure 6.23 below.



Tool for Benchmarking Spreadsheet for creating a spider diagram

Recommendation	1	2	3	4	5	6	7	8	9	10
	Strategies	Diversified Population	Identification	Attributes	Situation	Context	Serendipity	Storage	Stakeholders	Research& innovation
Scores	4	5	10	6	9	6	7	6	4	6

INSTRUCTIONS:

Tool for Benchmarking please enter your values for each of the 10 . The scores will automatically be adjusted in the spider diagram to the right.

Figure 6.23.: DHP Representation in Benchmarking

6.2.4. DHP Quality

The Heritage of the collected information define the quality of the Data. At this stage, the definition of Heritage involved the presence not only the content. It is the express by itself the real meaning of the data. The perception of the importance and relevance of the information is measured through the definitions and metamodel that is proposed. As it is mentioned in the publication Gordon et al. (2015) Digital Heritage Preservation explores the specific concept and context.

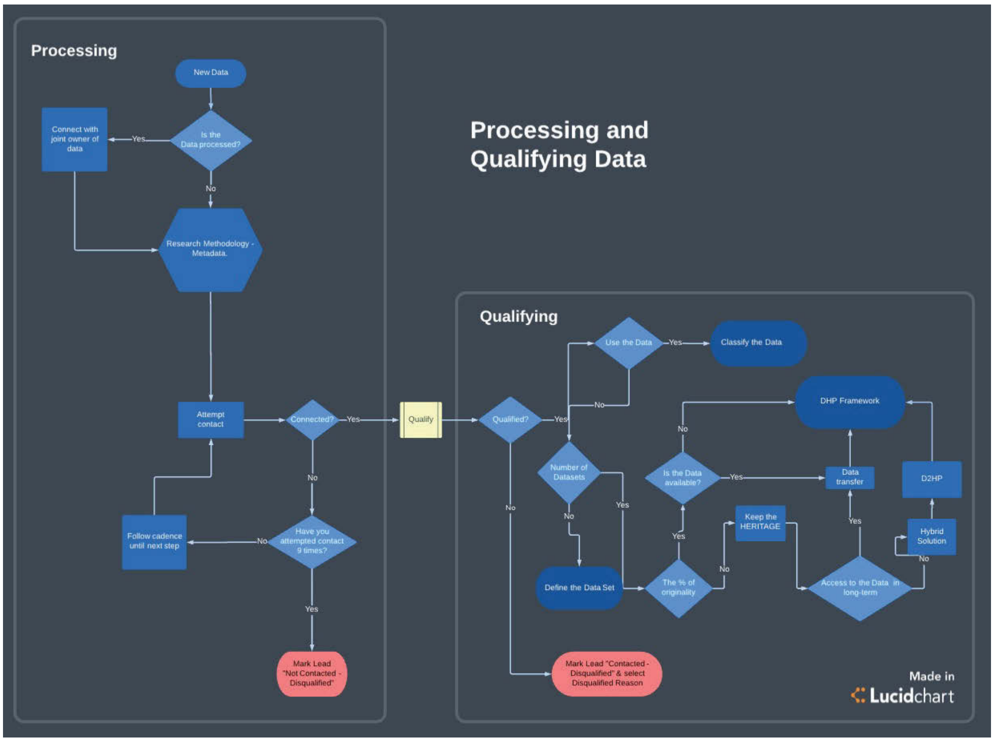


Figure 6.24.: Qualifying and Processing Model

Part III.

Contributions Digital Data Heritage Preservation (DHP)

7. Research Contribution

The prime goal of this study was to design and develop a theoretical model of Digital Heritage Preservation (DHP), that facilitates the manageability and dependability. Both of these quality attributes are considered to be the key features of the proposed model. A practical implementation of the DHP model based on DROID open source framework is called the D2HP. This solution integrates various Digital Preservation tools and Data Heritage components.

The association between Heritage and Digital Preservation terms, is initially reported and addressed in SOTA. A common interpretation of Heritage concept is somehow restricted. In this work, a new concept of Data Heritage was introduced to address the specifics of data creation and its originality. This research redefines the Heritage and sees it as a New concept in the world of IoT, Big Data and the Digital Transformation as a whole. The Data Heritage considers multiple dimensions, relations and the evolution of data. The concept addresses such attributes as: the originality, the source, the ownership, the value, the accesibility, a future use of data and many others. The Data Heritage is not only set in the Cultural Ecosystem, but it can be considered as a key element of every environment that IoT components interact with.

The adaptation of DROID framework provides a strong experimentation base and a standarised framework for the management of the Digital Objects defined by the DHP model. The proposed DHP model combined with several data preservation technologies represents an innovation to traditional concepts of data representation and data handling. This proposal helps to improve and resolve Digital Transformation when dealing with IoT and Big Data challenges (data connectivity and dimensionality). Through conducted experiments, case studies and action research, it was demonstrated that the proposed Data Heritage concepts and DHP models are feasible and can be highly effective as demonstrated in Part II of the thesis.

The new definition of the Data Heritage and related concepts together with experiments involving the DHP model, represent the real contribution of this thesis. The experiments with the DHP modelling consider the optimisation of processing time and effectiveness of data representation.

7.1. Discussion

The challenging parts of this research involve finding suitable methodologies and solutions for data storage and management that can be applied in the IoT domain. This work aims to offer a significant contribution to the IoT community as a new knowledge and a technological innovation. The literature review provides discussions on various modelling approaches and technological advances related to Data Preservation explored thus far.

This research provides the rationale for considering the Data Preservation as an integral part of data intensive IoT architecture. As the number of stakeholders in the IoT domain is growing rapidly, the appreciation for adequate Data Preservation methods and technologies has to match these fast developments.

The central part of this work focus on the data modelling and design methodologies for the implementation of DHP Framework. The outcomes of the main validation experiment involving the DROID based implementation, demonstrates the inner workings, the meaning and the value of Heritage Preservation approach. However, the DHP model approach considers the use of adaptive processes for Data Preservation, hence it is expected the model can be adopted for many IoT applications.

This thesis discusses many Data Preservation concepts, theories, methodologies, techniques, tools and demonstrates case studies related to Digital Transformation in context of IoT.

Although necessary, the algorithmic and statistical methods are applied at a general level when evaluating and benchmarking the characteristics of DHP prototype.

7.2. Challenges

The principal challenges of the implementation and modelling of Digital Data Heritage Preservation are divided in the following areas.

- Interpretation of the Concepts

The DHP as a model is theoretical where the definition of the concepts is fundamental. There are initiatives of Digital Preservation, but the real challenge is to give the position to the term Heritage inside the Computer Sciences initiatives. The cultural and historical interpretation of Heritage are many and valuables. But the reason of this recompilation of ideas, and innovative proposal, is to add the Heritage as a term for inherent to the technological field.

- Misunderstanding of Digital Preservation

In SOTA there are explanations and studies about Digital Preservation. The tendency from physical to digital items are not absolute. Assuming this process and being realistic, the physicality of the World in this Era is not automatic and fast. The relation between physicality and digitalisation, depends on the field and changes time by time. There is not one answer around this analysis. The standardization in every case is the best approach for DHP solution. The ideas in the research support the strong strategies for standardisation. If the scientific part is clear and the investigation accomplishes the principal objective, to show the key concepts.

- Methodologies and Modelling

The analysis of different tendencies as HeuristicLab and OMILAB modelling bring a new ideas across the creation of new DHP model. Despite of the limitations and the huge amount of concepts, the idea of standarisation be reflected in the application and combination of the new concepts.

- Marketshare for the proposed modelt

. The strategies for commercialization have to define: the niche of the market, the receptor, the affections to the community and the results. The deadlines are also important, because the best solution is not only the best investigation, also could be the solution that take place in the appropriate time. Definition of the rights and property laws. It is fundamental to be aware to know the implications of the Intellectual Property. Furthermore, the alliances and strategies are good to have arguments. The innovation never ends. In general, I also believe that to focus in the new markets could be a right way to learn. I strongly believe that the relation between research and commercial field are undisputed.

- Business Approach

The long-term maintenance of the solution means in business point of view, it has to have sustainable duration. Sometimes, to mix the fields in the knowledge could have better results than only focus in one solution. For example, technology and health, mechanics and biotechnology will give us an idea. A significant challenge represents the problem, how DHP can assist the need of Digital Preservation field. There is a need to develop further ICT solutions using IoT technologies. Medicine has finally entered an era in which clinical digitization implementations and data analytic systems are converging. It is important to recognize the power of data in other domains and are beginning to apply it to the health records, applying digitalisation as a necessary but insufficient tool for this purpose in some cases.

7.3. Future Work

The future work should be focused in the following statements:

- Implementation of the new dataset and different fields as experimental complementary solution.
- Integration of network capabilities with the connection to IoT devices.
- Algorithmical and Statistical approach has to be developed in strong way.

Heritage and DHP Framework

There seems to be a gap in tradition approaches to digital data preservation and planning procedures, as these methods do not integrate digital data preservation with aspects of data heritage which can be in anyway useful. In order to mitigate limitations of processes that use solely data preservation mechanisms, an integrated framework for digital data heritage and preservation is proposed. This new framework offers a consolidated and systematic approach for redesigning how digital data is managed in business enterprises. The methodology involves several process activities and adopts classical Design Patterns for digital data preservation and heritage.

Among key activities are: Preparation for re-engineering, Mapping and analysing as-as processes, Designing to-be process, Implementation re-engineered process, Continuous improvement. Among many Design Patterns were found particularly suitable for the method the following: Creational Patterns (Prototype and Singleton), Structural Patterns (Adapter, Composite and Proxy) and Behavioral Patterns (Visitor and State Machine).

Most of the concepts as the purpose of the information give the idea on how to create knowledge and the new understanding of the fields. The visualisation of data would generate the new knowledge.

The main topic is related to the preservation and management.

The following statements guided the research:

- Specific kind of tools benefit the systematic evaluation supported by benchmarking method
- The relevant sample tasks for testing tools and effective data sets to support this testing
- Opportunities arise for collaborative efforts to facilitate benchmarking and share the resulting evidence
- Concrete actions can be taken to establish benchmarking as a useful development of methods

7.4. Final Conclusion

- The Digital Transformation trends around the IoT, the edge computing, the analytics, 5G, AI and blockchain technologies largely depend on methods of data sensing, acquisition, processing and visualisation. However, the quality and dimensionality of data itself seems to be neglected. This work aims to highlight these issues and suggestes suitable solutions around the Digital Data Heritage Preservation concepts.
- The rationale for using Digital Data Preservation tools is to maintain digital artifacts for a long period of time. Using the D2HP model allows to keep the evidence of Heritage inherent attributes of Digital Data in a massive scale and dynamic system solution (Big Data) in order to preserve the nature, ownership, meaning and context of the data creation.
- As demonstrated in chapters 5 and 6, the D2HP framework passed dependability, manageability and usability tests with good results. Here it was validated that the dynamic composition of Digital Data Heritage and DHP framework can integrate within Digital Transformation processes as hypothesised.

- In chapter 6, section 6.1, several case studies demonstrate the basic DHP concepts and their applicability in the development of IoT ecosystem. However, it is the integration of the Digital Heritage concepts and the HeriTAG solution that offers the most valuable contribution to the practice of Digital Transformation.

Bibliography

Abdul-Qawy, A. S., Pramod, P., Magesh, E. & Srinivasulu, T. (2015), 'The internet of things (iot): An overview', *International Journal of engineering Research and Applications* **1**(5), 71–82.

Abdul-Qawy, A. S., Pramod, P., Magesh, E. & Srinivasulu, T. (n.d.), 'The internet of things (iot): An overview'.

Adak, S., Batra, V. S., Bhardwaj, D. N., Kamesam, P. V., Kankar, P., Kurhekar, M. P. & Srivastava, B. (2002), A system for knowledge management in bioinformatics, *in* 'Proceedings of the eleventh international conference on Information and knowledge management', ACM, pp. 638–641.

Albani, M. (2012), Long term preservation of earth observation data: The challenge and the cooperation activities, *in* 'Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International', pp. 7279–7282.

Andal-Ancion, A., Cartwright, P. A. & Yip, G. S. (2003), 'The digital transformation of traditional business', *MIT Sloan Management Review* **44**(4), 34–41.

Archenaa, J. & Anita, E. M. (2015), 'A survey of big data analytics in healthcare and government', *Procedia Computer Science* **50**, 408–413.

Bathurst (2014), 'Bathurst Council'.
URL: *http://www.bathurst.nsw.gov.au/building-and/heritage/what-is-heritage.html*

Beagrie, N. (2003), *National Digital Preservation Initiatives: An Overview of Developments in Australia, the Netherlands, and the United Kingdom and of Related International Activity. Strategies and Tools for the Digital Library.*, ERIC.

Becker, C., Antunes, G., Barateiro, J., Vieira, R. & Borbinha, J. (2011), Modeling digital preservation capabilities in enterprise architecture, *in* 'Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times', ACM, pp. 84–93.

Becker, C., Gon, #231, alo Antunes, Jos, #233, Barateiro, Vieira, R. & Borbinha (2011), ‘Modeling digital preservation capabilities in enterprise architecture’.

Becker, C., Rauber, A., Heydegger, V., Schnasse, J. & Thaller, M. (2008), ‘A generic XML language for characterising objects to support digital preservation’.

Belle, A., Thiagarajan, R., Soroushmehr, S., Navidi, F., Beard, D. A. & Najarian, K. (2015), ‘Big data analytics in healthcare’, *BioMed research international* **2015**.

Bengio, Y., Courville, A. C. & Vincent, P. (2012), ‘Unsupervised feature learning and deep learning: A review and new perspectives’, *CoRR, abs/1206.5538* **1**.

Berman, S. J. (2012*a*), ‘Digital transformation: opportunities to create new business models’, *Strategy & Leadership* **40**(2), 16–24.

Berman, S. J. (2012*b*), ‘Digital transformation: opportunities to create new business models’, *Strategy & Leadership* **40**(2), 16–24.

Bork, D., Karagiannis, D. & Hawryszkiewycz, I. T. (2017), ‘Supporting customized design thinking using a metamodel-based approach’.

Bowersox, D. J., Closs, D. J. & Drayer, R. W. (2005), ‘The digital transformation: technology and beyond’, *Supply Chain Management Review* **9**(1), 22–29.

Brody, T., Carr, L., Hey, J., Brown, A. & Hitchcock, S. (2008), ‘Pronom-roar: Adding format profiles to a repository registry to inform preservation services’, *International Journal of Digital Curation* **2**(2).

Brown, A., Katuu, S., Sebina, P. & Seles, A. (2009), ‘Module 4: Preserving electronic records’, *London: International Records Management Trust, available at: http://www. irmt. org/documents/e duc_training/term% 20modules/IR MT% 20TERM% 20Module* **204**.

Carrion Gordon, L. (2014), ‘Digital preservation strategy: Planning procedure to preserve critical information of heritage’, *APCASE 2014* .

Carrion Gordon, L. & Chaczko, Z. (2017), ‘The metamode of heritage preservation for medical big data’.

Carrion Gordon, L., Flores, M., Escobar, A. & Horna, L. (2017), ‘Estimation of the contaminant risk level of petroleum residues applying fda techniques’, *Latin American Journal of Computing* .

Castelli, D., Taylor, S. J. R. & Zoppi, F. (2010), Open knowledge on e-Infrastructures: the BELIEF project Digital Library, *in* ‘IST-Africa, 2010’, pp. 1–15.

Celi, L. A., Marshall, J. D., Lai, Y. & Stone, D. J. (2015), 'Disrupting electronic health records systems: the next generation', *JMIR medical informatics* **3**(4).

Chaczko, Z., Carrion, L., Alenazy, W. & Mu, M. (2014), Development of an expert system to assist in resource management, *in* 'Information Technology Based Higher Education and Training (ITHET), 2014', IEEE, pp. 1–6.

Challa, S., Gulrez, T., Chaczko, Z. & Paranesha, T. N. (2005), Opportunistic information fusion: a new paradigm for next generation networked sensing systems, *in* 'Information Fusion, 2005 8th International Conference on', Vol. 1, p. 8 pp.

Chandras, C., Weaver, T., Zouberakis, M., Hancock, J. M., Schofield, P. N. & Aidinis, V. (2008), Digital preservation - financial sustainability of biological data and material resources, *in* 'BioInformatics and BioEngineering, 2008. BIBE 2008. 8th IEEE International Conference on', pp. 1–6.

Chang, S.-F., Sikora, T. & Purl, A. (2001), 'Overview of the mpeg-7 standard', *IEEE Transactions on circuits and systems for video technology* **11**(6), 688–695.

Chowdhary, P., Bhaskaran, K., Caswell, N. S., Chang, H., Chao, T., Chen, S.-K., Dikun, M., Lei, H., Jeng, J.-J., Kapoor, S. et al. (2006), 'Model driven development for business performance management', *IBM Systems Journal* **45**(3), 587–605.

Coallition, D. P. C. D. P. (2012), 'Introduction - Definitions and Concepts'.
URL: *http://www.dpconline.org/advice/preservationhandbook/introduction/definitions-and-concepts?q=definitions*

Commission, D. P. et al. (2002), 'Preservation management of digital materials: the handbook'.

Crespi, V., Galstyan, A. & Lerman, K. (2005), Comparative analysis of top-down and bottom-up methodologies for multi-agent system design, *in* 'Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems', ACM, pp. 1159–1160.

Cunliffe, A. (2011), 'Dissecting the digital preservation software platform'.

Day, M. (1998), 'Cedars: Digital preservation and metadata'.

Dhillon, I. S., Mallela, S. & Kumar, R. (2003), 'A divisive information theoretic feature clustering algorithm for text classification', *The Journal of Machine Learning Research* **3**, 1265–1287.

Doyle, J., Viktor, H. & Paquet, E. (2009), 'Long-term digital preservation: preserving authenticity and usability of 3-D data', *International Journal on Digital*

Libraries **10**(1), 33–47.
URL: *http://dx.doi.org/10.1007/s00799-009-0051-7*

Drosou, M., Jagadish, H., Pitoura, E. & Stoyanovich, J. (2017), 'Diversity in big data: A review', *Big data* **5**(2), 73–84.

Duranti, L. & Thibodeau, K. (2006), 'The concept of record in interactive, experiential and dynamic environments: the view of interpares', *Archival Science* **6**(1), 13–68.

Duretec, K., Kulmukhametov, A., Rauber, A. & Becker, C. (2015), 'Benchmarks for digital preservation tools'.

Eakin, L., Friedl, A., Schonfeld, R. & Choudhury, S. (2009), 'A selective literature review on digital preservation sustainability'.

Engelsman, W., Jonkers, H. & Quartel, D. (2011), 'ArchiMateÂ® Extension for Modeling and Managing Motivation, Principles, and Requirements in TOGAFÂ®'.

Fausett, L. (1994), *Fundamentals of neural networks: architectures, algorithms, and applications*, Prentice-Hall, Inc.

Fayad, M. E., Sanchez, H. A., Hegde, S. G., Basia, A. & Vakil, A. (2014), *Software patterns, knowledge maps, and domain analysis*, CRC Press.

Ferreira, M., Baptista, A. A. & Ramalho, J. C. (2006), 'A foundation for automatic digital preservation', *Ariadne* (48).

Fill, H.-G. (2016), 'Semantic-based modeling for information systems using the semfis platform'.

Fill, H.-G. & Karagiannis, D. (2013), 'On the conceptualisation of modelling methods using the adoxx meta modelling platform', *Enterprise Modelling and Information Systems Architectures–International Journal of Conceptual Modeling* **8**(1), 4–25.

Fiorini, R. A. (2015), A cybernetics update for competitive deep learning system, *in* 'Proceedings 2nd International Electronic Conference on Entropy and Its Applications. Retrieved from http://sciforum. net/conference/ecea-2/paper/3277'.

Frank, U. (2011), Some guidelines for the conception of domain-specific modelling languages., *in* 'EMISA', Vol. 190, pp. 93–106.

Garagnani, S. & Manferdini, A. M. (2013), 'Parametric accuracy: Building information modeling process applied to the cultural heritage preservation', *International*

Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences **5**, W1.

Garfinkel, S., Farrell, P., Roussev, V. & Dinolt, G. (2009), 'Bringing science to digital forensics with standardized forensic corpora', *digital investigation* **6**, S2–S11.

Gartner, R. & Lavoie, B. (2013), 'Preservation metadata', *DPC Technol Watch Rep. doi* **10**.

Gipp, B., Meuschke, N., Beel, J. & Breitinger, C. (2016), Using the blockchain of cryptocurrencies for timestamping digital cultural heritage, *in* 'Proceedings of the Workshop on Web Archiving and Digital Libraries (WADL) held in conjunction with the 16th ACM/IEEE Joint Conference on Digital Libraries (JCDL)'.

Gordon, L. C. & Chaczko, Z. (2017*a*), Ontological metamodel for consistency of data heritage preservation (dhp), *in* '2017 25th International Conference on Systems Engineering (ICSEng)', IEEE, pp. 438–442.

Gordon, L. C. & Chaczko, Z. (2017*b*), Ontological metamodel for consistency of data heritage preservation (dhp), *in* '2017 25th International Conference on Systems Engineering (ICSEng)', IEEE, pp. 438–442.

Gordon, L. C., Chaczko, Z. & Resconi, G. (2015), Standardized mapping model for heritage preservation and serendipity in cloud, *in* 'International Conference on Computer Aided Systems Theory', Springer, pp. 110–117.

Goth, G. (2012), 'Preserving digital data', *Communications of the ACM* **55**(4), 11.

Gracy, K. F. & Kahn, M. B. (2012), 'Preservation in the digital age', *Library Resources & Technical Services* **56**(1), 25.

Grefen, P., Mehandjiev, N., Kouvas, G., Weichhart, G. & Eshuis, R. (2009), 'Dynamic business network process management in instant virtual enterprises', *Computers in Industry* **60**(2), 86–103.

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S. & Lew, M. S. (2016), 'Deep learning for visual understanding: A review', *Neurocomputing* **187**, 27–48.

Hedges, M., Hasan, A. & Blanke, T. (2007*a*), Management and preservation of research data with irods, *in* 'Proceedings of the ACM first workshop on CyberInfrastructure: information management in eScience', ACM, pp. 17–22.

Hedges, M., Hasan, A. & Blanke, T. (2007*b*), 'Management and preservation of research data with iRODS'.

Hegarty, R., Lamb, D. J. & Attwood, A. (2014), Digital evidence challenges in the internet of things., *in* 'INC', pp. 163–172.

Higgins, S. (2011), 'Digital curation: the emergence of a new discipline', *International Journal of Digital Curation* **6**(2), 78–88.

Hodges, D. & Lunau, C. D. (1999), 'The national library of canadaâs digital library initiatives', *Library Hi Tech* **17**(2), 152–164.

Holden, M. T. & Lynch, P. (2004), 'Choosing the appropriate methodology: Understanding research philosophy', *The marketing review* **4**(4), 397–409.

Hu, T., Tan, C. L., Tang, Y., Sung, S. Y., Xiong, H. & Qu, C. (2008), 'Co-clustering bipartite with pattern preservation for topic extraction', *International Journal on Artificial Intelligence Tools* **17**(01), 87–107.

Jack, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P. J., L Whitwell, J., Ward, C. et al. (2008), 'The alzheimer's disease neuroimaging initiative (adni): Mri methods', *Journal of magnetic resonance imaging* **27**(4), 685–691.

Jones, S., McCann, P. & Kim, Y. (2009), A practical guide to implementing the dcc testbed methodology, Technical report, Digital Curation Centre.

Junghanns, M., Petermann, A., Neumann, M. & Rahm, E. (2017), Management and analysis of big graph data: Current systems and open challenges, *in* 'Handbook of Big Data Technologies', Springer, pp. 457–505.

Kane, G. C., Palmer, D., Phillips, A. N., Kiron, D. & Buckley, N. (2015), 'Strategy, not technology, drives digital transformation', *MIT Sloan Management Review and Deloitte University Press* **14**.

Kaner, M. & Karni, R. (2004), 'A capability maturity model for knowledge-based decisionmaking', *Information, Knowledge, Systems Management* **4**(4), 225–252.

Kaplan, F. (2015*a*), 'A map for big data research in digital humanities', *Frontiers in Digital Humanities* **2**, 1.

Kaplan, F. (2015*b*), 'A map for big data research in digital humanities', *Frontiers in Digital Humanities* **2**, 1.

Karagiannis, D., Buchmann, R. A., Burzynski, P., Reimer, U. & Walch, M. (2016), Fundamental conceptual modeling languages in omilab, *in* 'Domain-Specific Conceptual Modeling', Springer, pp. 3–30.

Karagiannis, D. & Kühn, H. (2002), Metamodelling platforms, *in* 'EC-Web', Vol. 2455, p. 182.

Karagiannis, D., Mayr, H. C. & Mylopoulos, J. (2016), *Domain-Specific Conceptual Modeling: Concepts, Methods and Tools*, Springer.

Kaur, R., Thakur, A., Saini, H. S. & Kumar, R. (2015), Enhanced steganographic method preserving base quality of information using lsb, parity and spread spectrum technique, *in* 'Advanced Computing & Communication Technologies (ACCT), 2015 Fifth International Conference on', IEEE, pp. 148–152.

Kejser, U. B. g., Nielsen, A. B. & Thirifays, A. (2011*a*), 'Cost model for digital preservation: Cost of digital migration', *International Journal of Digital Curation* **6**(1), 255–267.

Kejser, U. B., Nielsen, A. B. & Thirifays, A. (2011*b*), 'Cost model for digital preservation: Cost of digital migration', *International Journal of Digital Curation* **6**(1), 255–267.

Kinchin, I. M., Hay, D. B. & Adams, A. (2000), 'How a qualitative approach to concept map analysis can be used to aid learning by illustrating patterns of conceptual development', *Educational research* **42**(1), 43–57.

Kohonen, T. (1989), Self-organizing feature maps, *in* 'Self-organization and associative memory', Springer, pp. 119–157.

Koller, D., Frischer, B. & Humphreys, G. (2010), 'Research challenges for digital archives of 3D cultural heritage models', *J. Comput. Cult. Herit.* **2**(3), 1–17.

Lavoie, B. F. & Gartner, R. (2005), *Preservation metadata*, OCLC.

Lee, C. (2016), Big data in management research, Technical report.

Lee, D. (2007), 'Practical maintenance of evolving metadata for digital preservation: algorithmic solution and system support', *International Journal on Digital Libraries* **6**(4), 313–326.
URL: *http://dx.doi.org/10.1007/s00799-007-0014-9*

Lee, K.-H., Slattery, O., Lu, R., Tang, X. & McCrary, V. (2002), 'The state of the art and practice in digital preservation', *Journal of research of the National institute of standards and technology* **107**(1), 93.

Lorie, R. A. (2001), Long term preservation of digital information, *in* 'Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries', ACM, pp. 346–352.

Maemura, E., Moles, N. & Becker, C. (2017), 'Organizational assessment frameworks for digital preservation: A literature review and mapping', *Journal of the Association for Information Science and Technology* **68**(7), 1619–1637.

Mazonka, O. (n.d.), 'Blockchain: Simple explanation, 2016', *URL http://jrxv.net/x/16/blockchain-gentle-introduction. pdf.[Online* .

McCay-Peet, L. (2014*a*), 'INVESTIGATING WORK-RELATED SERENDIPITY, WHAT INFLUENCES IT, AND HOW IT MAY BE FACILITATED IN DIGITAL ENVIRONMENTS'.

McCay-Peet, L. (2014*b*), 'Investigating work-related serendipity, what influences it, and how it may be facilitated in digital environments'.

McDonald, J. (2008), 'Design Patterns'.
URL: *www.dzone.com*

Mehta, S. (n.d.), 'Service design approach; an opportunity for indian it industry to move up the value chain'.

Mettler, T. & Rohner, P. (2009), Situational maturity models as instrumental artifacts for organizational design, *in* 'Proceedings of the 4th international conference on design science research in information systems and technology', ACM, p. 22.

Mishra, P. & Kumar Sharma, R. (2014), 'A hybrid framework based on sipoc and six sigma dmaic for improving process dimensions in supply chain network', *International Journal of Quality & Reliability Management* **31**(5), 522–546.

Mita, A. (2016), 'Digital powrr (preserving digital objects with restricted resources) http://digitalpowrr. niu. edu', *Technical Services Quarterly* **33**(1), 94–96.

Muir, A. (2004), 'Digital preservation: awareness, responsibility and rights issues', *Journal of Information Science* **30**(1), 73–92.

Nasir, S. A. M. & Noor, N. L. M. (2010), Integrating ontology-based approach in knowledge management system (kms): Construction of batik heritage ontology, *in* 'Science and Social Research (CSSR), 2010 International Conference on', IEEE, pp. 674–679.

Neumayer, R., Rauber, A., Ross, S. & Strodl, S. (2008), 'Fostering Collaboration with DigitalPreservationEurope'.

O'Donnell, K. et al. (2010), 'Taming digital records with the warrior princess: Developing a xena preservation interface for trim', *Archives and Manuscripts* **38**(2), 37.

Ogleby, C. L. (1998), ‘The "Truthlikeness" of Virtual Reality Reconstructions of Architectural Heritage: Concepts and Metada’.

Organisation, C. (1997), ‘Glossary of world heritage terms’.

Park, E. (2001), ‘Understanding" authenticity" in records and information management: Analyzing practitioner constructs’, *The American Archivist* **64**(2), 270–291.

Patel, K. & McCarthy, M. P. (2000), *Digital transformation: the essentials of e-business leadership*, McGraw-Hill Professional.

Phillips, M. E. (1999), ‘The national library of australia: Ensuring long-term access to online publications’, *Journal of Electronic Publishing* **4**(4).

Pittl, B. & Bork, D. (2017), Modeling digital enterprise ecosystems with archimate: A mobility provision case study, *in* ‘International Conference on Serviceology’, Springer, pp. 178–189.

Podio, F. L. (1992), Research on methods for determining optical disk media life expectancy estimates, *in* ‘Optical Data Storage’, Vol. 1663, International Society for Optics and Photonics, pp. 447–456.

Prakash, S., Shanmugam, V. & Murugesan, A. (2012), ‘Privacy preserving combinatorial function for multi-partitioned data sets’, *International Journal of Computer Applications* **44**(8), 8–10.

PRAKASH, V. S. & SHANMUGAM, A. (2013), ‘ENHANCED PRIVACY PRESERVATION WITH PERTURBED DATA USING FEATURE SELECTION’, *Journal of Theoretical & Applied Information Technology* **58**(3).

Quenault, H. (2004), Vers: Building a digital record heritage, *in* ‘Archiving Conference’, Vol. 2004, Society for Imaging Science and Technology, pp. 2–7.

Rajasekar, A., Wan, M., Moore, R. & Schroeder, W. (2006), A prototype rule-based distributed data management system, *in* ‘HPDC workshop on Next Generation Distributed Data Management’, Vol. 102.

Ramakrishnan, N. & Grama, A. Y. (1999), ‘Data mining: From serendipity to science’, *Computer* **32**(8), 34–37.

Reiter, L., Rinner, O., Picotti, P., Hüttenhain, R., Beck, M., Brusniak, M.-Y., Hengartner, M. O. & Aebersold, R. (2011), ‘mprophet: automated data processing and statistical validation for large-scale srm experiments’, *Nature methods* **8**(5), 430–435.

Rogers, C. (2012), ‘Digital records pathways: topics in digital preservation’.

Rogers, C. (2015), ‘Diplomatics of born digital documents–considering documentary form in a digital environment’, *Records Management Journal* **25**(1), 6–20.

Rössler, M. (2002), UNESCO World Heritage Centre Background Document on UNESCO World Heritage Cultural Landscapes, *in* ‘prepared for the FAO Workshop and Steering Committee Meeting of the GIAHS (Globally Important Ingenious Agricultural Heritage Systems) project, Rome’, pp. 5–7.

Russell, K. (2000), ‘Digital preservation and the cedars project experience’, *New Review of Academic Librarianship* **6**(1), 139–154.

Ryan, J. & Silvanto, S. (2009), ‘The world heritage list: The making and management of a brand’, *Place Branding and Public Diplomacy* **5**(4), 290–300.

Sabatier, P. A. (1986), ‘Top-down and bottom-up approaches to implementation research: a critical analysis and suggested synthesis’, *Journal of public policy* **6**(01), 21–48.

Saracevic, T. & Dalbello, M. (2001), ‘In: Proceedings of the american society for information science and technology,(2001), vol. 38, pp. 209-223. a survey of digital library education’, *Proceedings of the American Society for Information Science and Technology* **38**, 209–223.

Schuurman, N. & Leszczynski, A. (2006), ‘Ontology-based metadata’, *Transactions in GIS* **10**(5), 709–726.

Serenko, A. & Bontis, N. (2009), ‘Global ranking of knowledge management and intellectual capital academic journals’, *Journal of Knowledge Management* **13**(1), 4–15.

Simon, H. A. (1979), ‘Rational decision making in business organizations’, *The American economic review* **69**(4), 493–513.

Sinz, E. J. & Bork, D. (n.d.), ‘Design of a som business process modelling tool based on the adoxx meta-modelling platform’.

Steinberg, A. N. & Rogova, G. (2008), Situation and context in data fusion and natural language understanding, *in* ‘Information Fusion, 2008 11th International Conference on’, pp. 1–8.

Strodl, S., Becker, C., Neumayer, R. & Rauber, A. (2007), ‘How to choose a digital preservation strategy: evaluating a preservation planning procedure’.

Strodl, S., Petrov, P., Rauber, A. et al. (2011), 'Research on digital preservation within projects co-funded by the european union in the ict programme', *Vienna University of Technology, Tech. Rep* .

Syerina Azlin Md, N. & Noor, N. L. M. (2010), Integrating ontology-based approach in Knowledge Management System (KMS): construction of Batik Heritage Ontology, *in* 'Science and Social Research (CSSR), 2010 International Conference on', pp. 674–679.

Thalmann, S., Seeber, I., Maier, R., Ren, #233, Peinl, Pawlowski, J. M., Hetmank, L., Kruse, P. & Bick, M. (2012), 'Ontology-based standardization on knowledge exchange in social knowledge management environments'.

Thibodeau, K. (2002), 'Overview of technological approaches to digital preservation and challenges in coming years', *The state of digital preservation: an international perspective* pp. 4–31.

Tuomi, I. (1999), Data is more than knowledge: Implications of the reversed knowledge hierarchy for knowledge management and organizational memory, *in* 'Systems Sciences, 1999. HICSS-32. Proceedings of the 32nd Annual Hawaii International Conference on', IEEE, pp. 12–pp.

UNESCO (n.d.), 'Information Document Glossary of World Heritage Terms (June, 1996)'.
URL: *http://whc.unesco.org/archive/gloss96.htm*

Vail, E. F. (1999), 'Knowledge mapping: getting started with knowledge management', *Information Systems Management* **16**, 10–23.

Verheul, I. (2006), *Networking for digital preservation: current practice in 15 national libraries*, Vol. 119, Walter de Gruyter.

Vesanto, J., Himberg, J., Alhoniemi, E., Parhankangas, J. et al. (1999), Self-organizing map in matlab: the som toolbox, *in* 'Proceedings of the Matlab DSP conference', Vol. 99, pp. 16–17.

Wagner, S., Beham, A., Kronberger, G., Kommenda, M., Pitzer, E., Kofler, M., Vonolfen, S., Winkler, S., Dorfer, V. & Affenzeller, M. (2010), Heuristiclab 3.3: A unified approach to metaheuristic optimization, *in* 'Actas del séptimo congreso español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados (MAEB'2010)', p. 8.

Wagner, S. & Kronberger, G. (2012), Algorithm and experiment design with heuristic lab: an open source optimization environment for research and education, *in*

‘Proceedings of the 14th annual conference companion on Genetic and evolutionary computation’, ACM, pp. 1287–1316.

Walken, C. (2014), ‘Math41112/61112 ergodic theory’, *Course notes, retrieved July* **14**.

Weaver, R., Meier, W. M. & Duerr, R. M. (2008), Maintaining Data Records: Practical Decisions Required For Data Set Prioritization, Preservation, and Access, *in* ‘Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International’, Vol. 3, pp. III – 617–III – 619.

Wendler, R. (1999), ‘Ldi update: metadata in the library’, *Library Notes* **1286**, 4–5.

Wheeldon, J. & Faubert, J. (2009), ‘Framing experience: Concept maps, mind maps, and data collection in qualitative research’, *International Journal of Qualitative Methods* **8**(3), 68–83.

Xiong, H., Steinbach, M., Ruslim, A. & Kumar, V. (2009), ‘Characterizing pattern preserving clustering’, *Knowledge & Information Systems* **19**(3), 311–336.

Yu, B. & Singh, M. P. (2002), An agent-based approach to knowledge management, *in* ‘Proceedings of the eleventh international conference on Information and knowledge management’, ACM, pp. 642–644.

Zhang, Y., Qiu, M., Tsai, C.-W., Hassan, M. M. & Alamri, A. (2017), ‘Health-cps: Healthcare cyber-physical system assisted by cloud and big data’, *IEEE Systems Journal* **11**(1), 88–95.

Zins, C. (2007), ‘Conceptual approaches for defining data, information, and knowledge’, *Journal of the Association for Information Science and Technology* **58**(4), 479–493.

Zysman, J., Murray, J., Feldman, S., Nielsen, N. C. & Kushida, K. E. (2011), ‘Services with everything: the ict-enabled digital transformation of services’.

Appendix

Chapter 4
Digital Patterns for Heritage and Data Preservation Standards

Lucia Carrion Gordon and Zenon Chaczko

Abstract This research covers the digital preservation concept and the terms around this definition. According to this contextualization it is related with the management of data in the different fields but with the same issues and concerns. The appropriate way to keep information through the time is with preservation parameters. The managing of massive amounts of critical data involves designing, modeling, processing and implementation of accurate system. The methods for the treatment of data have to consider two dimensions that this chapter has focused: access dimension and cognitive dimension. Both of them have relevance to get results because at the same time, ensure the correct data preservation. On the other side data management has to consider digital resources and digital artefacts and how they affect the process of preservation. In this chapter there are approaches with other authors referencing current and future studies. This chapter has been organized as follow. First in the introduction we can find ideas about the types of data in the different fields nowadays. Second, the main part, the proposal of methods in reference with other authors. Finally the conclusions and future projects based on the findings. This study focus on the development of framework, beginning with the methodology and projected through the application of specific patterns such a model for sustainable solution.

Keywords Digital preservation · Knowledge management · Heritage · Data management · Digital patterns

4.1 Introduction

Through the time, the information and how it can be organised and classified show the challenge in the Era of Data Management. But increasingly our possessions and communications are no longer material, now they are digital and dependent on

L.C. Gordon (✉) · Z. Chaczko
FEIT Faculty of Engineering and IT, University of Technology, NSW, Australia
e-mail: Lucia.CarrionGordon@uts.edu.au

Z. Chaczko
e-mail: Zenon.Chaczko@uts.edu.au

G. Borowik et al. (eds.), *Computational Intelligence and Efficiency in Engineering Systems*, Studies in Computational Intelligence 595,
DOI 10.1007/978-3-319-15720-7_4

the technology to make them accessible. As a consequence of this big growth, to gather the requirements of data preservation involved the study and analysis inside the enterprises for constructing the profile of the customer with any kind of data. Based on that, the classification of data is important inside the Government Entities, Business Enterprises, Health Field and Cultural Area as a principal sectors.

Increasing regulatory compliance mandates are forcing enterprises to seek new approaches to managing reference data. Sometimes the approach of tracking reference data in spreadsheets and doing manual reconciliation is both timeconsuming and prone to human error. As organizations merge and businesses evolve, reference data must be continually mapped and merged as applications are linked and integrated, accuracy and consistency, realize improved data quality, strategy lets organizations adapt reference data as the business evolves.

A further motivation is to provide and improve data referencing. There are two parts: Physical (related with accessibility) and Cognitive (related with contextualization). The management of artifact have to be considered as a technological component. The main criteria is the challenge of sort out the information or data processed in smart and efficient way. The knowledge management give us the best direction of the processes of standardization. Between the principal requirements of the users are involved the authenticity and reliability. The development of the right process is independent of the type and amount of data. Nowadays the requirements of the appropriate media for the data storage should be solved easily. The needs in storage media, migration, conversion, emulation and management strategies have been focused as a consequence of the appropriate implementation of processes. This research proposed the following parts: definition of the methodology, modeling the process, implementing framework. The strategy and application of policies are important considerations at the moment of applied in case studies. Preservation responsibilities for integrating long-term preservation into planning, administration, system architectures and resource allocation are inherent in the parts of this research.

Other considerations above this topic are the applications of Digital Preservation applied in Cloud Computing architecture. Consequently referencing is a challenge because it means to identify metadata and classify them while the conventional solutions have been very limited. The investigation of scalable solution with the identification of metadata and emphasize the role of methodology and patterns should be useful to provide guidelines for protecting resources from dealing with obsolescence, responsibilities, methods of preservation, cost, and metadata formats. Based on the experience of National Library of Australia, the principal factor that we have to consider in the application of preservation is to provide frameworks related to digital heritage. It is one applications basically with the use of archives. The principal actors in this process of preservation are users and custodians. Knowledge Preservation has been applied in many areas focus on standards, scalable designing, adaptive and survivable network applications.

Furthermore, the reason for using the frameworks is the standardization of the process and the results. It does not mean the uniformity of the data, so the easy access to maintain the originality to ensure the preservation of the information. Preservation strategies and specific software tools for emulation or migration must always be chosen according to requirements of individual institutions or users. Software tools sometimes guarantee full traceability and documentation of all elements influencing the final result. Utility Analysis and its ability to integrate inhomogeneous criteria sets is used to evaluate different strategies.The framework is often a layered structure indicating what kind of relations can or should be built and how they would interrelate.

4.2 Heritage Concepts

The heritage term is defining as the crucial and central part of the research, we can refer it to "heritage is those items and places that are valued by the community and is conserved and preserved for future generations" [2]. The concept is much wider than historical buildings. It includes items and places with natural heritage significance and Aboriginal heritage significance. "The heritage value of a place is also known as its cultural significance which means its aesthetic, historic, scientific, social or spiritual value for past, present or future generation" [2].

One of the principal keywords in this research is Heritage. UNESCO is one of the Entities that refers in an accurate way to this definition. The expert meeting denied a heritage route as "composed of tangible elements of which the cultural significance comes from exchanges and a multi-dimensional dialogue across countries or regions that illustrate the interaction of movement, along the route, in space and time" [10].

But the question is what is heritage and which parameters defining the artifact or the information as a heritage? The context and the interpretation of data is the answer.

4.3 General Frameworks for Heritage

The development of the preservation framework is related with the value of information. "Value has always been the reason underlying heritage conservation. It is selfevident that no society makes an effort to conserve what it does not value" [2]. The Value of the information is located in the second level of important after the concept of Heritage. If we can define the result that we want, how we can manage and measure the value of that information? The principal ways are the perception, interpretation and contextualization.

4.3.1 Value Derived from Individual Perceptions

Depends on the Entity or Enterprise that information is classified, the study can consider the specific perceptions adapted to the needs of the end user. For example, there are geological, archeological, economic, social matters that could be managed by this methodology. These characteristics are considered in the case studies (Fig. 4.1).

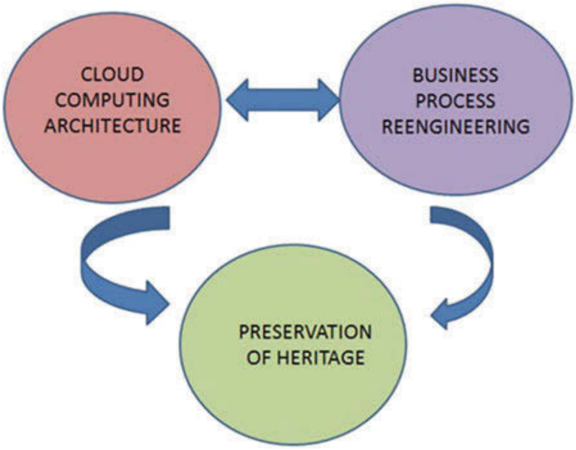


Fig. 4.1 Areas of information management

4.3.2 Process of Preservation

The main concern around researching and definition of data preservation is the technique or methodology used. More than using technology as the solution, the principal consideration is how we can apply and verify the whole process. The definition of cycles and steps to follow has been considered as a principal part for the first consolidation of the preservation knowledge. The different perspective for the information management supports the idea to achieve general solution. The following figure shows us which could be the areas that cover the information management. There are three areas: Cloud Computing Architecture, Business Process Reengineering and Preservation of Heritage.

4.4 Architecture Vision

The formatting of data provides the unique result called Digital Age of the information. The Knowledge Management and Ontology are techniques for analyzing information. One concern of digitalization would be the format, standards and migration of the data. It should be solved with the use of Architectural Methodology and with the development of a fast prototype. This requires the definition of the sequential process.

First, should be considering a Framework as a whole front end and for reception of the information in a basic way. Using the Open Group Architectural Framework (TOGAF) [7] it is found to be more suitable as a result of dissemination of the data. Meanwhile, it is associated with an Ontology and Knowledge Management terms to be more specific.

Second, the Methodology with an architectural vision using concepts of Architectural Development Method (ADM). It has been identified in terms of enterprise description for validating information of several types of data.

Third, the conceptualization of patterns for a centralization of the preservation knowledge providing a unique result: the digital age of the data.

The connection is also with the artefacts and the correct use of them. The recovering of the information is another issue has to consider. Between the techniques we used "encapsulation, migration and emulation" [6]. The evaluation is based on "PRIST model defining by Privacy, Rights, Integrity, Security and Trust" [5] across PC considerations related with Physical and Cognitive characteristic of data.

According to the author [3] the correct use of the interface and the exact generation of metadata, are the key considerations to follow around the process of optimal data preservation (Fig. 4.2).

$$TOGAF + \mathbf{ADM} + \mathbf{PRIST} \tag{4.1}$$

Formula Open Group Architectural Framework (TOGAF) + Architectural Development Method (ADM) [7] + Privacy, Rights, Integrity, Security and Trust (PRIST model) [5].

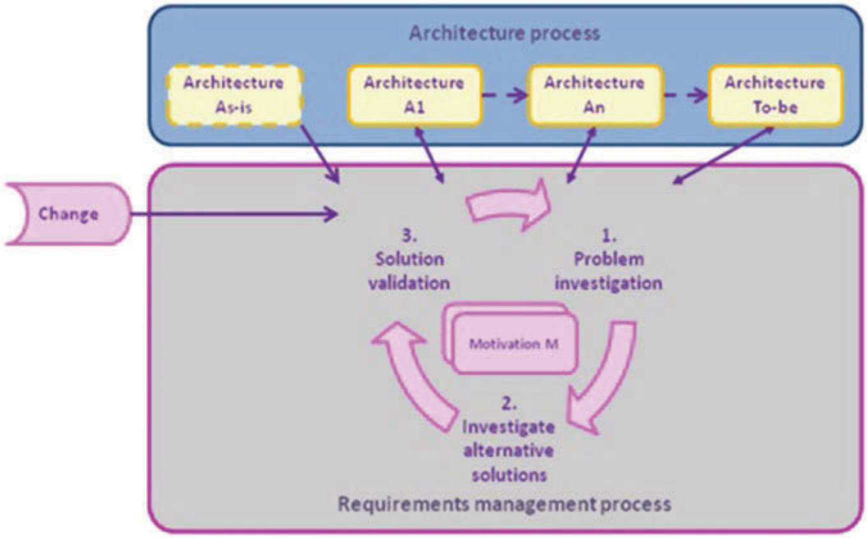


Fig. 4.2 Application of the requirements cycle [7]

4.5 Evaluation of Preservation and Planning Procedure

Nowadays Electronic publications, and a collection of multimedia art are the principal issues that need to preserve their data holdings, simulation models, or studies over time. The other matters that we have to consider is the legal constraints, to guarantee the accessibility and usability over time. Moreover the Digital Preservation is considering as a research discipline.

The two methods related in this study is: migration and emulation. There are often hold very valuable data in complex structures.

Migration is the method of repeated conversion of files or objects. Emulation denotes the duplication of the functionality of systems, be it software, hardware parts, or legacy computer systems as a whole, needed to display, access, or edit a certain document [9].

The integration of inhomogeneous criteria sets is used to evaluate different strategies. For example web archive collections, collections of scientific publications, and electronic multimedia art. The collaborative research program is funding research in different aspects of digital preservation, including collection practices, risk analyses, legal and policy issues, and technology. In Europe two new research projects. Scientific information is born-digital or only available in digital form. At the moment libraries, archive and scientific institutions are primarily dealing with the challenge of long term preservation [9].

The Reference Model for an Open Archival Information System (OAIS) was published 2002 by the Consultative Committee for Space Data Systems (CCSDS). ISO 14721:2003 defines an OAIS as "an archive, consisting of an organization of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community" [9].

The OAIS model further provides a framework for describing and comparing different long term preservation strategies and techniques. Submission Information Package (SIP) Archival Information Package (AIP) Ingest also extracts descriptive information from the AIPs and coordinates updates to Archival Storage and Data Management Access functions include access control, request coordination, response generation in the form of Dissemination Information Packages (DIPs) and delivery of the responses to consumers. The development of preservation strategies and standards as well as packaging designs and plans (Fig. 4.3).

The facility to integrated into existing archival environments is one of the advantages of this process.

The Preservation Planning is a crucial decision process, depending on both object characteristics as well as institutional requirements. One of the methods used is the Utility Analysis approach developed at the Vienna University of Technology and the Dutch testbed designed by the National Archive of the Netherlands.

Two web archives, one coming from a library context, the other one from an archiving institution, two collections of electronic publications with scientific provenance, coming from three European national libraries and a large collection of born-digital multimedia art [9].

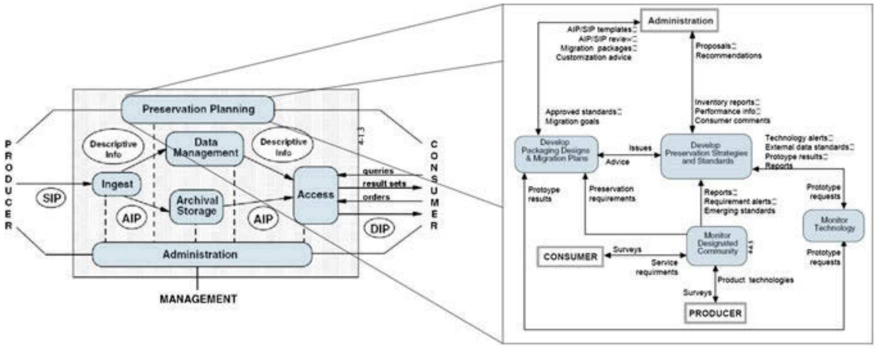


Fig. 4.3 Functional entities of the OAIS reference model [9]

The main idea of this research is related the management of heritage. It could be useful as a cultural case study related directly in terms of preservation. It exposes the experimental methodology and a valid analysis of the results. The reliability and accuracy of data are very strategic points in this article because the knowledge base is wide and complete.

On the other hand is important to differentiate the terms data, information, knowledge and knowledge management is almost as large as a number of authors contributing to the field. “Data is a set of discrete, objective facts”. “Information is data that has been organized or given structure that is, placed in context and thus endowed with meaning”. “Knowledge is information combined with experience, context, interpretation and reflection that is ready to apply to decisions and actions” [8].

4.5.1 Analysis of Capability Maturity Model (CMM)

The CMM measures and manages the improvement in software development processes. Preservation as a process with the heritage follow these steps. There are several levels:

Level 1—Initial, Level 2—Repeatable, Level 3—Defined, Level 4—Managed, Level 5—Optimizing

According to the methods and modelling, there are general approaches that give us the decision making maturity in different level as it shows in the figure below (Figs. 4.4 and 4.5).

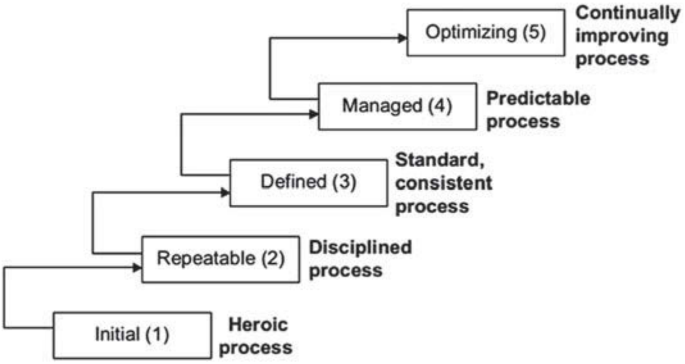
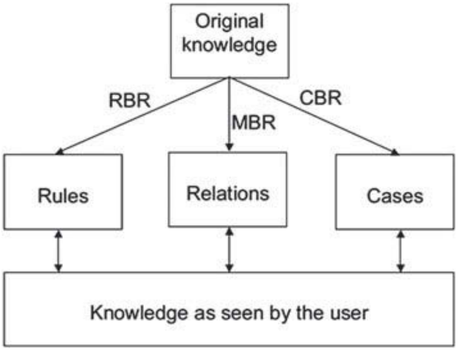


Fig. 4.4 The capability maturity model [8]

Fig. 4.5 General approaches to handling and presenting knowledge [8]



4.5.2 Heritage and Digital Data Preservation Framework

There seems to be a gap in tradition approaches to digital data preservation and planning procedures, as these methods do not integrate digital data preservation with aspects of data heritage which can be in anyway useful. In order to mitigate limitations of processes that use solely data preservation mechanisms, an integrated framework for digital data heritage and preservation is proposed. This new framework offers a consolidated and systematic approach for redesigning how digital data is managed in business enterprises. The methodology involves several process activities and adopts classical Design Patterns for digital data preservation and heritage. Among key activities are: Preparation for re-engineering, Mapping and analysing as-as processes, Designing to-be process, Implementation re-engineered process, Continuous improvement. Among many Design Patterns were found particularly suitable for the method the following:

- Creational Patterns (Prototype and Singleton)
- Structural Patterns (Adapter, Composite and Proxy) and
- Behavioral Patterns (Visitor and State Machine)

4.5.3 Modelling Heritage and Data Preservation

The new framework's principal components are represented in 5 + 2 dimensional model. That combines 5 dimensions of: Privacy, Rights, Integrity, Security, Trust that are denoted as PRIST [5] and two additional Cognitive (Authenticity) and Physical (Access) dimensions denoted as CP.

The new model formulas can be expressed as follows:

$$M_A + M_B = 7PC \tag{4.2}$$

$$PRIST + CP = DiADM(DIGITAL\_AGE\_DATA\_MODEL) \tag{4.3}$$

There are three dimensions for the Preservation such as:

Cognitive Dimension: Higher-level knowledge, meaning, situation, understandability, context, interpretability, cataloguing (Fig. 4.6).

Physical Dimension: Access control, Historical Retrieval, Digital Rights Management, extraction, interface, storage.

Processes: Techniques, Approach.

When representing data in Privacy, Rights, Integrity, Security and Trust dimensions, the proposed model needs to contextualize the interpretation, conservation and preservation of the natural and cultural heritage of the real world artifacts.

Thus, the model needs to incorporate physical and cognitive dimensions to form a unified model we can be called the Digital Age of Data Model or simply DiADM.

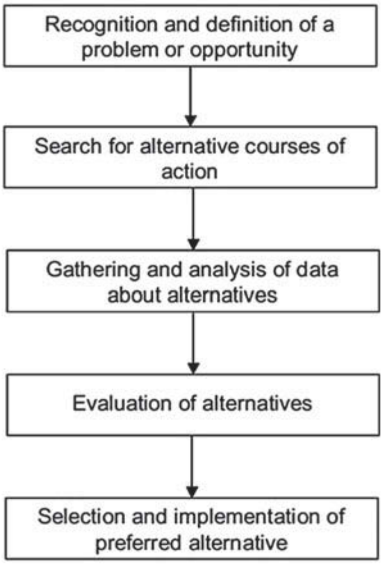

Recognition and definition of a problem or opportunity
Search for alternative courses of action
Gathering and analysis of data about alternatives
Evaluation of alternatives
Selection and implementation of preferred alternative

Fig. 4.6 Rational economic model of decisionmaking [8]

During the evaluation of the approach the following question need to be addressed to ensure positive outcomes:

1. How is this material to be kept for the future?
2. Does making it accessible involve risks for the right to privacy?
3. Are we going to treat it as heritage?
4. It is possible to encoded contextualization of information to preserve originality?
5. Can all dimensions in this model keep and preserve information without damaging it?

In order to validate the proposed framework, the next step is to generate DiADM model in context, obtain concrete results and evaluate the techniques. The DiADM contextualization will depend on the case studies scenarios.

4.6 Evaluation Methods

Based on the exposed statements, the basic techniques are related with genetic and biological cases of study. The relationship between these terms is given for the behaviour and the treatment of data. There are examples referenced by known authors. The sustainability of the preservation of the information give us the discussion about the appropriate resources infrastructure.

Research by Chandras [4] explains the relationship between the terms for preservation and genetics. Nowadays Biological databases are useful for determining the interaction of biological molecules and process. The financial issues known as the important matters for evaluation of the preservation patterns is useful for testing the methodology presented. The web data concept and the use of data warehouse are the principal tendency for the management of information. The behaviour of BRC Biological Resource Center reflects the best approximation of the concepts. In the context of the studies of the maintaining Data Records [11]. "An operational climate data service must ensure that all climate data are preserved and made available to users. In addition to the climate data; metadata; production software source code; documentation on the data, metadata and data formats; calibration/validation information; and QA information will also be archived. Regular backup of data and the capability to migrate any or all of this information to new media are also important." The challenges are specialized for deciding to manage the data sets that have to be measured. The activity to generate information and in different types and also integrate the quality for the three dimensions impact.

Based on the weighting factors use of data sets and resources decisions are very important. Research related with the evaluation of a given data set is based upon inputs from three components: the Maturity Matrix, Data Set Activity levels, and Stakeholder Input. The experience related with infrastructure is one of the possibilities to adopt new kinds of the aggregation of the contents like a case of study of Libraries. The mature Matrix Criteria has the specific considerations given by the term Preservation. According to the author the management of the information is divided in three parts of the Data Set Evaluation Criteria: Science Maturity, Preservation Maturity and Social Impact. The other example is the Safety Net for Scientific Data. There are proposal of how the society and the researches in this case can preserve the information based on Data Archiving Polices Molecular Ecology. This technique is after the study of the basic information and how the databases have to be managed. The considerations with the DNA based microarray technology [1] are given by the treatment of the structured or not structured information.

The schema of XML and how it could be standardized the information is based on the two basic techniques: migration and emulation. This alternative could determine the reliability of the generalized framework and also affect the methodology a patterns that the researching have to establish as a result of the study. The analysis of the sequences of the different behaviours biological and genetic mixes and summarizes the terms as knowledge management and DNA. Otherwise, the patterns are related with this terminology. On the other hand using the Open Group Architectural Framework (TOGAF) [7] the research should base the central part in developing of Standardised Framework. For example Cloud Computing and Service Oriented Architecture (SOA) based on Web services technology designed to assist cultural heritage institutions in the implementation of migration based preservation interventions. SOA delivers a recommendation service and a method to carry out complex format migrations. The recommendation service is supported by three evaluation components that assess the quality of every migration intervention in terms of its performance (Migration Broker), suitability of involved formats (Format Evaluator) and data loss (Object Evaluator). The proposed system has to be able to produce preservation metadata can be used by any enterprise in the Society for documenting preservation interventions and retain objects authenticity. Although it can also be used for other purposes such as comparing file formats or evaluating the performance of conversion applications the best of this technique is Collecting Data for having only one parameter. Metadata conceptualization plays an important role in preservation of digital heritage and archives in the digital objects. The quality as one of the principal issue should be considered like a summary of good authenticity and good reliability. However, as digital objects evolve over time, their associated metadata evolves a consistency issue. Since various functionalities of applications containing digital objects (e.g., digital library, public image repository) are based on metadata, evolving metadata directly affects the quality of such applications. Modern data applications are often large scale (having millions of digital objects) and are constructed by software agents.

4.7 Conclusion

The Framework, Methodology and Architecture are the principal parts that study explains. The digital age of the data is the initiative to standardise the solution for preserve information. The consolidation of data means to have accessibility and understanding of the information during the recovering cycle. It could be the next step of the study, as the validation of the results and method to find information.

The case studies with the genetic and biological references are useful for the conclusion in behaviour of the information. The relationship between many types of data is not relevant when the study is the access of the information. The process of Collecting Data is strategic and should determine the way that it has to be managed, classified and treated.

Preservation patterns are the final stepto consider to reach the standardization of the data. Across the theories, the technology should be oriented on the time that they are implemented.

Data must be securely stored and freely available to the research community depends on the context.

Long term sustainability requires adequate and reliable sources of finding data to preserve. The academic-commercial partnership may appear to have potential should corporations become involved in this collaborative effort. To prolonge financial sustainability is vital for data preservation and development of this kind of model.

References

1. Adak, S., et al.: A system for knowledge management in bioinformatics. In: Proceedings of the Eleventh International Conference on Information and Knowledge Management, (2002) pp. 638–641 Virginia, ACM
2. Bathurst, http://www.bathurst.nsw.gov.au/building-and/heritage/what-is-heritage.html
3. Castelli, D., et al.: Open knowledge on e-Infrastructures: the BELIEF project Digital Library. IST-Africa (2010)
4. Chandras, C., et al.: Digital preservation—financial sustainability of biological data and material resources. In: 8th IEEE International Conference on BioInformatics and BioEngineering (BIBE) (2008)
5. Challa, S., Gulrez, T., Chaczko, Z., Paranesha, T.N.: Opportunistic information fusion: a new paradigm for next generation networked sensing systems. In: 8th International Conference on Information Fusion, vol. 1, p. 8 (2005)
6. Doyle, J., et al.: Long-term digital preservation: preserving authenticity and usability of 3-D data. Int. J. Digit. Libr. **10**(1), 33–47 (2009)
7. Engelsman, W., Jonkers, H., Quartel, D.: ArchiMate® extension for modeling and managing motivation, principles, and requirements in TOGAF®. White Paper, The Open Group (2011)
8. Kaner, M., Karni, R.: A capability maturity model for knowledge-based decisionmaking. Inf. Knowl. Syst. Manag. **4**(4), 225–252 (2004)
9. Strodl, S., et al.: How to choose a digital preservation strategy: evaluating a preservation planning procedure. In: Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries. pp. 29–38. Vancouver, ACM (2007)

10. UNESCO. Information Document Glossary of World Heritage Terms. http://whc.unesco.org/archive/gloss96.htm (1996)
11. Weaver, R., et al.: Maintaining data records: practical decisions required for data set prioritization, preservation, and access. In:IEEE International Geoscience and Remote Sensing Symposium (IGARSS) (2008)

Chapter 24
Bio-informatics with Genetic Steganography Technique

Raniyah Wazirali and Zenon Chaczko

Abstract Biological systems have been a rich source of stimulation for computer security specialists. A wide number of approaches have been proposed over the last decade for encoding data using deoxyribonucleic acid (DNA), giving rise to the developing area of DNA data hiding. In this work, a new data hiding technique based upon DNA characteristic have been developed. DNA matrix has been used to represent the secret message. After that DNA matrix converted to QR (Quick Response) representation that offers a broad scope of practical usage. In addition, the paper provides an idea of choosing the optimal locations of the QR in order to obtain rightmost position. A new system based on the genetic algorithm has been developed. Finally, we embed the QR codes into the the most appropriate location by applying the Haar Wavelet technique on the resulting DNA signals and LSB with assist of the GA in order to reduce the error between the cover image and the stego image. Experimental results have presented a high PSNR which indicate a high level of quality in stego image with high capacity.

24.1 Introduction

For many years, computer security professionals have been intrigued by biological systems. Many Biological systems in the past have been a great source of inspiration for many computer professionals all over the world [13]. Therefore, over the previous decade, a number of strategies have been proposed concerning the use of deoxyribo-nucleic acid (DNA) in encoding information. This has led to an emerging field known as DNA data embedding [8]. In this field, data hiding is based upon DNA characteris-tics of a particular individual. Therefore, the secret message is

Raniyah Waziralii · Zenon Chaczko
Faculty of Engeneering and Information Technology, University of Technology, Sydney
e-mail: `Raniyah.A.Wazirali@student.uts.edu.au`, `zenon.chaczko@uts.edu.au`

represented by the DNA matrix. Furthermore, conversion of the DNA matrix into Quick Response (QR) opens the door to a variety of practical applications. When biologists discovered the structure of DNA, they started to change their perspective regarding information sci-ence. This is because it became apparent that information processing could be used not only to process, but also to represent and understand structures, activities and distribution of living organisms [16]. This has also led to many changes in the field of computer science. Some of these changes include DNA-based cryptography, molecu-lar DNA computing and DNA Steganography. Steganography has been described as the science of developing hidden messages within harmless messages or carriers [14]. Carriers could be videos, sound tracks, images and DNA. Data hiding within the DNA is based on its apparent randomness. Furthermore, it's tremendous storage capability as well as the ability to synthesize sequences in whatever length makes it the best medium of hiding data. This paper discusses some of the approaches used in DNA Steganography. Furthermore, the paper proposes to find the most appropriate positions on the spatial domain and frequency domain that provide slightest statistic features disruption on the spatial domain and frequency domain. Based on Genetic Algorithm (GA), GA affords a technique for hiding QR in the rightmost location which assurance minimum disturbance on the cover images. Figure 24.1 shows the flowchart for the proposed method.

24.2 Relared Works

The first successful DNA based hiding technique utilizes microdots similar to the ones used during the World War II. Furthermore, an encryption algorithm has already been developed. This algorithm uses a plaintext binary encoded message and mixes it with dummy strands in order to achieve information hiding founded on DNA binary sequences [12]. The other approach is a reversible scheme for information hiding that utilizes reversible contrast technology. The approach that will be discussed in detail, in this paper is the DNA_QR technique. This data hiding method utilizes a steganography algorithm that hides DNA encrypted secret messages within QR codes that are then randomized and finally embedded within a common image [10] as shown in Figure 24.1. QR codes are two dimensional matrix barcodes useful in encoding information. Compared to regular barcodes, QR codes are more common because they offer a higher storage capacity. DNA_QR combines a strong DNA encryption algorithm with a two-stage data hiding technique making the whole process very hard to break [19].

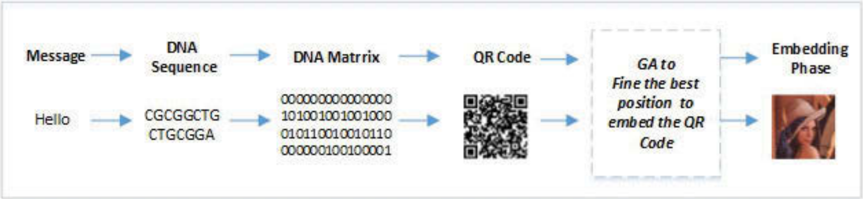


Fig. 24.1 Flowchart of the Proposed Method of DNA with QR Code

24.2.1 DNA Encryption

DNA encryption is an information hiding techniques inspired by the micro-dots technique used during World War II. Through this technique, scientists are able to produce artificial DNA strands containing secret messages [1, 15]. Table 24.1 displays a triplet that is used in encoding individual characters or numbers. DNA encryption utilizes a simple substitution algorithm that encodes individual char-acters into DNA sequences through Eq. 24.1 where the decoding through Eq. 24.2.

$$Encoding : X \longrightarrow Y \tag{24.1}$$

$$X \in [A,B,\ldots,Z,1,2,\ldots,0,”.”,”,”,”;”,”:”]$$

$$Y \in [xyz :: x,y,z \in [A,C,G,T]]$$

The function used for decoding is:

$$Decoding : Y \longrightarrow X \tag{24.2}$$

Table 24.1 Triplet Character Encoding Rule

Character = Triplet			
1 = ACC	A = CGA	K = AGG	U = CTG
2 = TAG	B = CCA	L = TGC	V = CCT
3 = GCA	C = GTT	M = TCC	W = CCG
4 = GAG	D = TTG	N = TCT	X = CTA
5 = AGA	E = GGC	O = GGA	Y = AAA
6 = TTA	F = GGT	P = GTG	Z = CTT
7 = ACA	G = TTT	Q = ACC	= ATA
8 = AGG	H = CGC	R = TCA	, = TCG
9 = GCG	I = ATG	S = ACG	. = GAT
0 = ACT	J = AGT	T = TTC	: = GCT

24.3 Polynomial Representation of DNA Matrix

In the following sections, DNA sequences are represented as a two-dimensional signal images using matrices. After that, a sparse polynomial representation of the DNA is considered. Lastly, these formulations will be discussed as QR formats. Firstly, various scholars note that the best way of storing information is through the use of QR images instead of DNA sequences, this is because the resulting database is not only compact but is also easier to traverse when performing searches. However, apart from QR code, DNA sequence can also be represented in the form of a matrix consisting of four rows. Each row represents the presence of a DNA base, be it T, G, C or A. For instance, a DNA sequence with five rows and a base of TCCGATAACGACT is represented as follows in Eq. 24.3:

$$\begin{array}{c} A \\ C \\ G \\ T \end{array} = \begin{array}{ccccccccccccc} 0&0&0&0&1&0&1&1&0&0&1&0&0 \\ 0&1&1&0&0&0&0&0&1&0&0&1&0 \\ 0&0&0&1&0&0&0&0&0&1&0&0&0 \\ 1&0&0&0&0&1&0&0&0&0&0&0&1 \end{array} \tag{24.3}$$

In the above matrix, the first row is a representation of occurrence of molecule A in the DNA sequence. The other rows represent molecules C, G, and T, therefore, the matrix can be visually represented using dots signifying occurrence of digit one and blanks signifying occurrence of digit zero. Secondly, DNA encryption involves the development of sparse polynomials [6]. Actually, the DNA matrix represented above is a sparse matrix made up of four sparse vectors in each one of the bases (base A, base C, base G and base T). Therefore, row vectors can be utilized to create four sparse polynomials using formula 24.4:

$$p_9 x_0 = \sum_{i=1}^{n} a_i x^i \tag{24.4}$$

In the above case, a_i represents $\{0, 1\}$, whereas n represents Genome length. Therefore, the previously documented matrix results in Eqs. 24.5–24.8 sparse polynomials:

$$P_A = X^5 + X^7 + X^8 + X^{11} \tag{24.5}$$

$$P_C = X^2 + X^3 + X^9 + X^{12} \tag{24.6}$$

$$P_G = X^4 + X^{10} \tag{24.7}$$

$$P_C = X^1 + X^6 + X^{13} \tag{24.8}$$

Therefore, the power of each of the terms can be stored within each polynomial in-stead of storing the whole sequence. The following are the resulting Eqs. 24.9–24.12:

$$P_A = (5,7,8,11) \tag{24.9}$$

$$P_C = (2,3,9,12) \tag{24.10}$$

$$P_G = (4,10) \tag{24.11}$$

$$P_T = (1,6,13) \tag{24.12}$$

DNA sequences have millions of bases. Therefore, trying to represent each one of the bases becomes impossible. What DNA encryption does is that it only stores sparse polynomial numbers made up of very big numbers. The number of storage bits re-quired is reduced by storing only the difference between powers. This is a common practice in lossless data compression . Therefore, when this technique is applied to the discussed DNA string, the following equations are obtained:

$$P_{dA} = (5,2,1,3) \tag{24.13}$$

$$P_{dC} = (2,1,6,3) \tag{24.14}$$

$$P_{dG} = (4,6) \tag{24.15}$$

$$P_{dT} = (1,5,7) \tag{24.16}$$

When multiples of the lengths of the DNA string are augmented with the four polynomials above forming one polynomial, the following is the result show in 24.17.

$$P = (5,12,20,31,33,36,45,57,61,71,72,78,91) \tag{24.17}$$

Finally, the difference sparse polynomial is given by the following Eq. 24.18:

$$p_d = (5,2,1,3,2,1,6,3,4,6,1,5,7) \tag{24.18}$$

This last vector occupies only 52 bits compared to the original vectors that occupied 104 bits. This is a compression ratio of a half. Therefore, representing DNA sequences with sparse polynomials reduces the amount of time consumed traversing the database. This is because each sequence in the database will be stored in the form of four sparse polynomials. These are PdT, PdG, PdC and PdA. At the same time, query sequences will also be represented as four sparse polynomials. These will include PdqT, PdqG, PdqC and PdqA. Hamming distance between all adjacent vectors (hT, hG, hC, hA) are determined with the following equation:

$$H = h_A + h_C + h_G + h_T \tag{24.19}$$

This reveals that the vector having the least hamming distance of (H) will be the closest to the desired query sequence.

24.4 QR Format Design

A QR code is a two dimensional code or a visual matrix code that can be read by Smartphone cameras or QR code readers. These codes are made up of black modules arranged in such a way that they assume a square pattern upon a white background. The encoded information could be a URL, embedded text or any other type of data. Figure 24.2 is an example of a commercial QR code.

Fig. 24.2 QR Code

24.5 GA to Find the Rightmost Position

The genetic algorithm (GA) is a search and optimization approach founded on the doctrines of natural selection and genetics and. It enables population involving may people to evolve under itemized selection rubrics to a state which makes the most of the "fitness" while minimizing the cost function. The technique was advanced by John Holland in 1975 [7]. The genetic algorithm begins with no information of the exact solution and based totally on replies from its progress operators such as reproduction, crossover and mutation to get the most suitable solution. By beginning at some independent ideas and examining in parallel, the GA approach prevents local minima and meets to achieve optimal solutions. Therefore, GAs have been used to be accomplished of finding high performance ranges in multifarious domains without suffering the challenges related with high dimensionality, as may happen with rise decent techniques or approaches that trust on imitative information [7, 18].

24.5.1 Process of Concealing the QR

After producing the Qr code, the GA will search for the rightmost position to embed this QR code with minimum destortion. The GA will help in reduce the error between the cover image and the stego image. The process of finding the best place to embed the QR code as follow [17].

1. Creating a random population of chromosomes; each chromosomes with length equal to the size of the produced QR code
2. Assessing the objective (fitness) function; The evaluation of the given function hinges on the PSRN criterion, which is sup-posed to reach its minimal value for the process to have any meaningful results. PSRN, or Peak Signal to Noise Ratio, being the criterion as the foundation for the fitness function, it is traditionally defined with the help of the functions (24.20, 24.21):

$$PSNR = 10 \times \log_{10} \frac{Max^2}{MSE} \tag{24.20}$$

$$MSE = \frac{1}{m \times n} \sum_{i-1}^{m} \sum_{j-}^{n} (A_{ij} - B_{ij})^2 \tag{24.21}$$

The Peak Signal to Noise Ratio PSNR value is considered to be the fitness function. Any value under 30 dB of PSNR values indicate low quality (i.e., distortion caused by embedding is high). A high and acceptable quality stego image should strive PSNR value of 40 dB, or greater [5]. The higher PSNR value means the minimum changes and the higher quality. The score matrix will evaluate the changes of the cover image for each block.

3. Repeating the steps a–c until the new population is made:

a. Choosing a pair of chromosomes (probability increasing together with the function of fitness);
b. Forming two new strings with a crossover of chromosomes;
c. Mutating the newly obtained chromosomes and plant the new strings in-to the population.

4. Swapping the new and the previous population;

5. As long as the optimum solution can be provided by correcting the value of error with the amount of generations or the maximum amount of generations is attained before it ceased to grow at the point where it serves as the location of the best chromosome, the experiment can be considered successful.

24.6 Embedding DNA_QR in images using DWT and LSB

24.6.0.1 Least Significant Bit Substitution

Two different methods can be used in the creation of QR codes and embedding information within them. The first method is known as LSB. According to various scholars, Least Significant Bit (LSB) substitution is the most commonly used Steganography technique. Presented with any image, LSB can replace least sig-nificant bits of every byte with bits from the hidden message. For a gray-level image, every pixel is made up of 8 bits. Since one pixel is capable of displaying 28 varia-tions, these translate to 256 variations. The main idea behind LSB substitution is embedding confidential data starting from the rightmost bits so that the procedure of embedding the data does not greatly affect original pixel value. Rightmost bits have the smallest weight. LSB can be represented mathematically using the Eq. 24.22

$$x_i^{'} = x_i - x_i mod 2^k + m_i \tag{24.22}$$

In the equation (22), $x_i^{'}$ stands for the stego-image's i^{th} pixel value, x_i stands for the i^{th} pixel value of the original image, whereas m_i stands for the decimal value belonging to the i^{th} block in the confidential data. K stands for the number of LSB that will be substituted. In the extraction process, k-rightmost bits are copied directly. The extraction message can be represented using the Eq. 24.23. Figure 24.3 shows an example of the LSB method.

$$m_i = x_i^{'} mod 2^k \tag{24.23}$$

24.6.0.2 Discrte Wavelet Transform

The second method of embedding secret message is DWT. This paper utilizes the Haar-DWT frequency domain transform technique [4, 9]. Any 2-dimensional Haar-DWT is made up of two operations. These are the horizontal and vertical operations. This section discusses procedures involved in manipulation of 2-dimentional Haar-DWT. Firstly; pixels are scanned from left to right along a horizontal axis. After this, subtraction and addition operations are performed on neighboring pixels. Sums are stored to the left, whereas differences are stored to the right [11]. This is represented in Fig. 24.3. This is repeated until all rows are completed. Pixel sums stand for low frequency sections of the initial image. They are symbolized by ‘L’. Pixel differences symbolize high frequency parts on the initial image. They are represented by ‘H’.

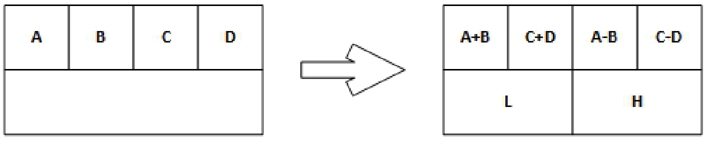


Fig. 24.3 Vertical Operation

Secondly, pixels are scanned from top to bottom along a vertical axis. Addition and subtraction operations are performed on neighboring pixels with sums being stored on top and differences at the bottom as shown in Fig. 24.4. This process is repeated until all columns are completed. Lastly, four sub-bands denoted as HH, LH, HL, and LL will be obtained. The LL sub-band represents low frequency portion. It looks quite similar to the initial image.

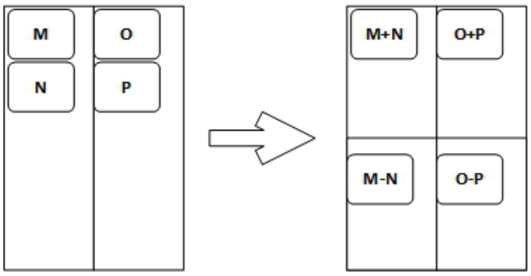


Fig. 24.4 Horizantal Operation

24.7 Performance Analysis

24.7.1 Hiding Capacity

The payload delivered by a steganographic technique illustrates maximum hiding capability of this algorithm. This means that the total number of bits that is capable of being embedded in a cover media is put into consideration. Since this paper is dealing with DNA media, the hiding capacity of the steganographic technique is measured in terms of bits per nucleotide (bpn) [3]. Assuming that total length of the DNA sequence referenced in this paper is —S— reflecting all nucleotides making up its sequence, then the DNA encryption algorithm substitutes each pair of message bases with another pair of bases. This means that any DNA sequence can hide a secret message as long as the length of its sequence [3]. This is because two bits of binary message (M) can be carried by a single nucleotide. Therefore, the algorithm's full payload can be expressed in terms of message size in bits. This fact is represented by the following equation:

$$Capacity = \frac{\beta}{\gamma} \tag{24.24}$$

Where β is the size of message in bits and γ is the size of cover in bases.

24.7.2 Visual Quality of the Stego-image

The fastest way of determining visual quality of a digital image is using perception of the human eye. Even though this criterion is effective, results usually differ from one individual to the other. An objective analysis of a digital image can be obtained using a parameter known as Peak Signal to Noise Ratio (PSNR). It is defined using equation (20 and 21).

In order to calculate PSNR, the dB value must be adopted facilitating quality judgment. The larger value of PSNR the greater the image quality. However, this means that there is little difference between the stego-image and the cover image. Conversely, a smaller PSNR value suggests that the level of distortion between the stego-image and the cover image is high. Due to the use of the GA technique, GA supports to find optimal location to embed the resulted QR code which aim to ensure minimum changes in the cover images.

24.7.3 Robustness Against Attacks

What makes DNA Steganography a robust method of data encoding is that four elements are required in order to crack or determine the secret message. These include the embedding technique, DNA matrix, DNA sequence and break QR code [2]. The Genetic Algorithm provide another layer of security due to the embedding the QR code in un-expected place. The QR code will be embedded in the noisiest location which will make it very strong against the visual attack.

24.8 Experimental Results

Experimental results of the proposed method are presented and discussed in this section. The program was written in Matlab with ZXing library for QR encoding. The images verified in our experiment are all 8-bit images with size 512×512. Figure 24.5 shows the original image of the Lenna.jpg. Figure 24.7 shown the stero image in different situations. Figure 24.6a show the image that have been encoded by DNA and QR code and embedded by LSB. Figure 24.6b show the image that have been encoded by DNA and QR code and embedded by DWT. Evaluation has been done to compare the result of the DNA encryption without QR code to evaluate the quality and the have been shown in Figs. 24.6c and 24.6d. The Performance analysis has been explained in Section 24.7.

Table 24.2 The result of PSNR and MSE for Lenna

	DNA with QR Method		DNA without QR Method	
	LSB	DWT	LSB	DWT
PSNR	66.639	701.8532	57.7745	60.9113
MSE	0.0141	0.0053	0.1086	00527

Table 24.3 The result of PSNR for the method in different images

DNA with QR Method		
	PSNR (LSB)	PSNR(DWT)
Lenna	66.639	70.853
Peppers	65.854	69.325
Baboon	65.584	68.554
Airplane	65.425	68.521
Barbara	64.985	66.344

Fig. 24.5 The original Lenna image

24.9 Conclusion

In conclusion, this paper presents a method that guarantees safety of information as it traverses from one person to another. The main idea behind Steganography is safe and secure data. Therefore, hiding DNA sequences within QR codes offers more practical advantages compared to simply hiding the sequences. Furthermore,

Fig. 24.6 a) DNA and QR code with LSB, b) DNA and QR code with DWT, c) DNA encoding without QR in LSB, d) DNA encoding without QR in DWT

the Steganography algorithm is flexible and can be done through DWT of LSB. DNA-QR can be applied on e-banking. Other methods like cryptography can be used together with DNA-QR in order to add another layer of security to secure information ex-change. For instance, when cryptography is incorporated, information is first coded before being taken through DNA Steganography. At the same time, other forms of media like music or videos can be used together with Steganography in order to offer well-secured data. Therefore, Steganography is a valid method that offers a credible and well-secured data.

Experiments have been done with assist of the GA to find the optimal locations of the metadata in order to obtain rightmost position. A new system based on the genetic algorithm has been developed. A few images have been tested using various size of text to be concealed. The stego images do not have any noticeable distortion on it that can be realized by the naked eyes. The PSNR value is consider as a fitness function for the proposed method. The proposed method has a high value of the PSNR value which indicates high quality of the stego images.

References

1. L. Adleman. Computing with DNA: The manipulation of DNA to solve mathematical problems is redefining what is meant by computation. *Scientific American*, 1998.
2. A. Cheddad, J. Condell, K. Curran, and P. Mc Kevitt. Digital image steganography: Survey and analysis of current methods. *Signal Processing*, 90(3):727–752, Mar. 2010.
3. G. Cui, L. Qin, Y. Wang, and X. Zhang. An encryption scheme using DNA technology. *3rd International Conference on Bio-Inspired Computing: Theories and Applications*, pages 37–42, Sept. 2008.

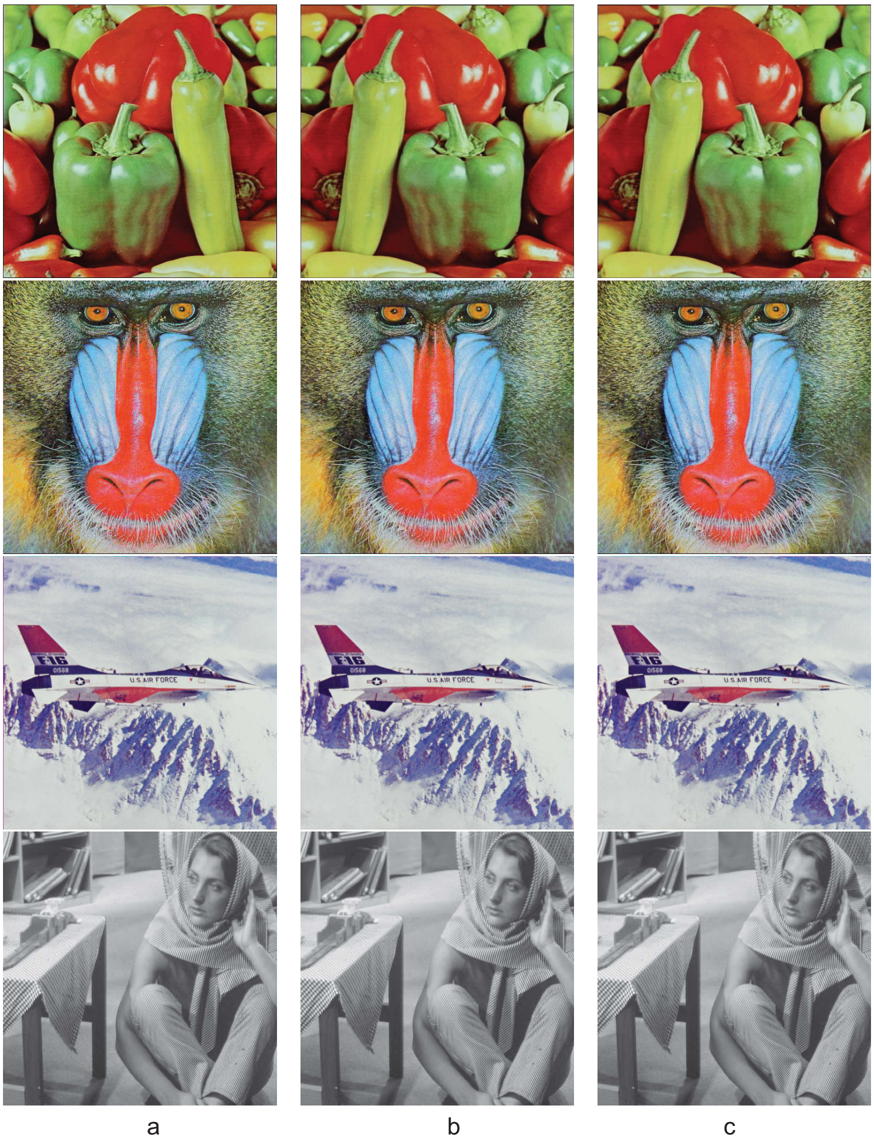


Fig. 24.7 Different varaitey of the proposed method in LSB and DWT: a) the original image, b) the result of DNA with QR coding in LSB method, and c) the result of DNA with QR code in DWT method

4. J. S. Dan Boneh, Christopher Dunworth, Richard J. Lipton. On The Computational Power of DNA. *Discrete Applied Mathematics*, 71:79—-94, 1995.
5. J. Fridrich, M. Goljan, and R. Du. Detecting LSB Steganography in Scale Images. *IEEE MultiMedia*, pages 22–28, 2001.
6. C. Guo, C.-c. Chang, and Z.-h. Wang. A new data hiding scheme based on dna sequence. *International Journal of Innovative Computing, Information and Control*, 8(1):139–149, 2012.
7. J. Holland. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence [Paperback]*. Cambridge, Mass : MIT Press, 1992, 1992.
8. L. Kari. From Micro-soft to Bio-soft : Computing with DNA. *World Scientific*, (1642):1–20, 1997.
9. D. Kumar and S. Singh. Secret data writing using DNA sequences. *2011 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC)*, pages 402–405, Apr. 2011.
10. A. Leier, C. Richter, W. Banzhaf, and H. Rauhe. Cryptography with DNA binary strands. *Biosystems*, 57(1):13–22, June 2000.
11. A. Magdy, M. Saeb, A. B. Mohamed, and A. Khadragi. Haar Wavelet Transform of The Signal Representation of DNA Sequences. *Haar Wavelet Transform of the Signal Representation of DNA Sequences*, 1, 2011.
12. H. Mousa, K. Moustafa, W. Abdel-wahed, and M. Hadhoud. Data Hiding B ased on Contrast Mapping Using DNA Medium. 8(2), 2011.
13. R. G. V. Schyndel, A. Z. Tirkel, and C. F. Osborne. A DIGITAL WATERMARK. 1994.
14. H. Shiu, K. Ng, J. Fang, R. Lee, and C. Huang. Data hiding methods based upon DNA sequences. *Information Sciences*, 180(11):2196–2208, June 2010.
15. X. Wang and Q. Zhang. DNA computing-based cryptography. *2009 Fourth International on Conference on Bio-Inspired Computing*, pages 1–3, Oct. 2009.
16. P. Wayner. *Disappearing cryptography : information hiding : steganography &amp; watermarking*. 2002.
17. R. A. Wazirali, Z. Chaczko, and A. Kale. Digital Multimedia Archiving based on Optimization Steganography System. *IEEE*, 2014.
18. Y.-T. Wu and F. Y. Shih. Genetic algorithm based methodology for breaking the steganalytic systems. *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society*, 36(1):24–31, Feb. 2006.
19. G. Xiao, M. Lu, L. Qin, and X. Lai. New field of cryptography: DNA cryptography. *Chinese Science Bulletin*, 51(12):1413–1420, June 2006.

Preservation Model to Process ’La Bomba del Chota’ as a living cultural heritage

Lucia Carrion Gordon
Faculty of Engineering and
Information Technology (FEIT)
University of Technology Sydney (UTS)
Australia, NSW, Sydney
Email: Lucia.CarrionGordon@uts.edu.au

Maria Gabriela Lopez Yanez
GIA Grupo Itinerante de Artes Guandul
Ecuador, Pichincha, Quito
Email: guanduldanza@gmail.com

***Abstract*—This project focuses on heritage concepts and their importance in every evolving and changing Digital Era where system solutions have to be sustainable, efficient and suitable to the basic needs. The prototype has to cover the principal requirements for the case studies. How to preserve the sociological ideas of dances in Ecuador like “La Bomba” is the best example and challenge to preserve the intangible data. The same idea is applicable with books and music. The History and how to keep it, is the principal mission of Heritage Preservation.**

The dance of La Bomba is rooted on a specific movement system whose main part is the sideward hip movement. La Bomba´s movement system is the surface manifestation of a whole system of knowledge whose principal characteristics are the historical relation of Choteños with their land and their families.

I. Introduction

Ecuador is a country located in the northwest part of South America with about fifteen million inhabitants. In the 2010 Ecuadorian census, mestizos were recognized as the majority (71.9%), followed by indigenous (8%) and Afrodescendant-Ecuadorians or Afroecuadorians (7.2%) (INEC, 2010). Mestizo-Ecuadorians occupy the highest social and economical status while Afroecuadorians rank as the lowest [11]. In the sixteenth century, the Jesuits, a Roman Catholic order of priests, first introduced African slaves in massive numbers to a valley located in the northeast part of the Ecuadorian territory known as Chota. From that time until the present, almost all of the inhabitants of Chota Valley have been Afrodescendants. Currently, Chota Valley is a rural area of approximately 80 km. It is composed of thirty-eight villages with about 25,000 residents representing approximately 2% of the Ecuadorian population. Is in this area where the dance named La Bomba was originated. La Bomba has been performed by Choteños since its origin approximately during the 16th century until the present period for reasons that La Bomba regenerates or revitalizes in Choteños a feeling of being “essentially us”. The deeply felt sense of an “us” or of “being Choteño” is possible because the movement system of La Bomba is meant to provoke specific interactions rooted in a Choteño system of knowledge that has historically existed since the 16th century. Therefore, the permanence of the dance of La Bomba works as an indicator of the vitality and continued presence of a Choteño approach to “being Choteño” in Ecuador.

Is in this area where the dance named La Bomba was originated. La Bomba has been performed by Choteños since its origin approximately during the 16th century until the present period for reasons that La Bomba regenerates or revitalizes in Choteños a feeling of being “essentially us”. The deeply felt sense of an “us” or of “being Choteño” is possible because the movement system of La Bomba is meant to provoke specific interactions rooted in a Choteño system of knowledge that has historically existed since the 16th century. Therefore, the permanence of the dance of La Bomba works as an indicator of the vitality and continued presence of a Choteño approach to “being Choteño” in Ecuador.

1. Background information

A. *Definition of La Bomba*

Although hundreds of Africans were victims of the oppressive slavery system, since their arrival in Ecuador (16th century), they were also active agents of adapting to their new reality. For instance, enslaved people re-established their communities and reformulated their way of living through the creation of new cultural practices such as La Bomba. La Bomba can be defined as a creative expression generated by enslaved Africans in Ecuador –Choteños- and executed by them from enslavement up until the present day. The main characteristic of La Bomba is that it is always performed as a shared experience, this meaning, while executing La Bomba, Choteños are in constant interaction among each other. The shared experience of La Bomba is based on the historical formation of groups of solidarity on which a communal structure and emotional bonds based on reciprocity are generated through a gathering spirit that highlights their own way of interacting among each other.

B. *Origin of La Bomba*

There is no exact date for the origin of La Bomba. Although the first written description of a performance of La Bomba in Chota was done by Hassaurek (1868), other authors [1], [2], [3] affirm that La Bomba originated in Chota Valley much earlier, sometime at the beginning of the Ecuadorian slavery

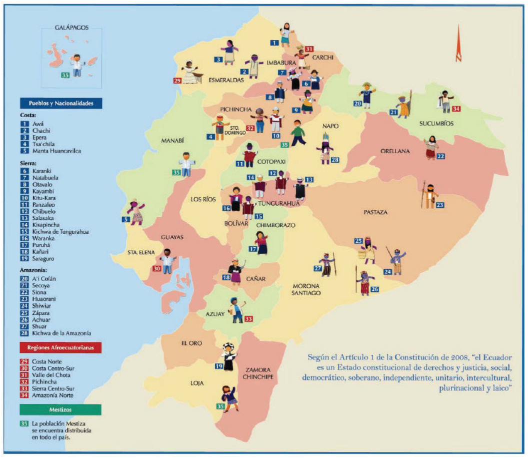


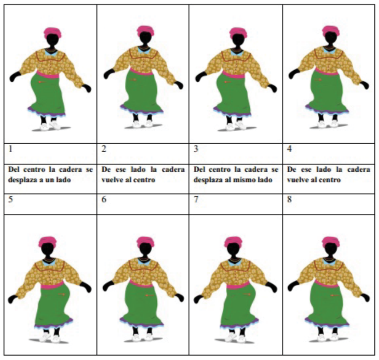
Fig. 1. Map of ethnic groups of Ecuador.

period (16th century). Hassaurek describes an Afro-Choteño performance called Bundi. This description is considered the first description of La Bomba since the author does mention the presence of a drum called La Bomba. Also, although the dance described – Bundi - does not appear to be the dance of La Bomba as it is performed in present days, neither has the name of La Bomba, it does have similarities. A special feature in the performance of La Bomba in Chota observed at the present and highlighted by Hassaurek in 1868, is the inseparable character of the dance, the music and the drum.

Regarding the origin of the name Bomba, in documents of the 18th century of Popayán, Colombia (country located next to Ecuador), Gutierrez (1971 in Valarezo, 2010) located slaves with a name that includes "Bomba". Gutierrez mentions the most remarkable case, a slave whose name was "José Bomba Arará," José of Spanish origin and Arará from a village near Dahomey . Valarezo (2010) suggests that the name Bomba could refer to a specific place within the village of Arará. There is another hypothesis for the origin of the name Bomba posited [4] In the Spanish language, bomba refers to a circle. The author relates the dance's name with the formation that dancers sometimes make, a circle, while dancing La Bomba. This version was also mentioned by Espinoza (2010) during a public conference. However, Bueno (1991) points out that if [4] is correct in his hypothesis, then a number of dances in Ecuador would also be named Bomba since many other dances are also performed in circles.

Map of ethnic groups of Ecuador. The numbers in red (29-34) show the peripheral geographic location of Afroecuadorians. With the exception of Nº32 on which the strong Afroecuadorian presence in the province of Pichincha is showed. Within Pichincha, Afroecuadorians live in peripheral and almost exclusive Afroecuadorian neighborhoods.

In the performance of La Bomba in Chota, two sideward hip movements are executed on each side. 1) from the center, the hip jut out to the side (e.g. right), 2) the hip returns to the center, 3) the hip jut out again toward the same direction as 1 (e.g. right), 4) the hip return to the center, 5) the hip jut out to the opposite side than 1 (e.g. left), 6) the hip returns to the center, 7) the hip jut out again toward the same direction as 5 (e.g. left) and 8) the hip return to the center. Design: Verónica López.



Del centro la cadera se desplaza a un lado

De ese lado la cadera vuelve al centro

Del centro la cadera se desplaza al mismo lado

De ese lado la cadera vuelve al centro

Fig. 2. La Bomba Movements, Design Veronica Lopez [1]

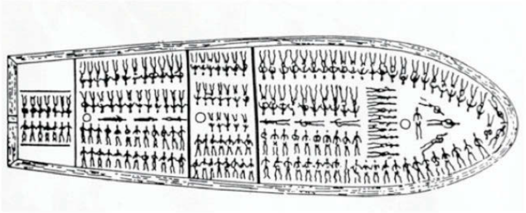


Fig. 3. Slave ship

Slave ship. Cut of the boat that transported slaves in fetal position. Marta Goldberg Archives.[12][12]

C. The movement system of La Bomba

i. Moving the hips: Currently, the principal movement of the dance of La Bomba is a sideward hip movement. Choteño sideward hip movement is essential in the performance of La Bomba in Chota as can be observed in one of the lyrics of a song of La Bomba, "The chicha (fermented beverage) and the alcohol don't have anise, shake it shake it little plant of chili, as women shake it around here" (Coba, 1980; Costales & Peñaherrera, 1959; Hernández, 2005, 2010). While executing the sideward hip movement, the hip follows the percussive rhythm of the music of La Bomba.

D. The knowledge system of La Bomba

i. The hip in La Bomba: A dancer's ability to move her or his hip as taught by the elders is considered a virtue among Choteños. For Choteños, hip are important as one of the most attractive body parts of males and females [4]. For instance, one of the virtues of a female Choteño is her ability to move her hip sideward with elegance, not just while dancing, but also while walking. "The way of walking is one of the characteristics of Choteños. They move their hips more than others" (L. Bolaños, personal communication, July, 2011). The ability to move the hip is intensified and more visible during the performance of La Bomba because this implies not only the ability to move the hip but also to have rhythm [5]

The central role of the hip movements in the performance of La Bomba is essential because these movements are a way that Choteños relate among each other. This becomes clear in one of the oldest descriptions of La Bomba in Chota written by Costales and Peñaherrera [5]. These authors describe in detail the Bottle dance, a dance highlighting the hip movement and the relation among participants in La Bomba.

In this description, the role of the dancer's hip movement as a way to relate among participants can be noticed. "Her hip begin to vibrate while chasing the black man… the black man avoids the hip shove..." (p. 192). Also, the hip movements serve as a way to gain the attention of participants "…lets see, who is the black man who can put up with me…! Slowly…the folds of the red skirt begin to shake…people break out of laughter and happiness… while Doña Custodia, putting her arms on her hip, makes a complete turn" ([5], p. 192).

ii. La Bomba as a space of freedom and creativity: La Bomba performed in Chota Valley represents a Choteño space of freedom and creativity (Peters, 2004). This sense of freedom and creativity while dancing La Bomba among Choteños appears to be historical as expressed through a song of La Bomba compiled by Chalá (2006), "Enough of cutting cane, freedom has arrived, freedom for the blacks, lets go to bombear [the action of performing La Bomba] " (p. 177).

I feel that there is a moment that just exists when we dance Bomba, not other rhythms that are also played in Chota. In that moment, what we do is just not rational. I remember once, it was the last day of the year. I had to work in Quito until the night. I just took the bus and arrived in Juncal [a village of Chota] at around three in the morning. When I saw how Choteños were dancing La Bomba, I just thought, these guys are crazy; it's not normal what they are doing. I was witnessing the best part of the party. The pure Choteños party. That is what I love about Chota. When they dance they are free. They don't measure anything, they don't calculate. It's just giving everything. The energy that they use when they dance, they just use all the energy they have.

Since slavery period, the performance of La Bomba in Chota is sometimes executed in an intimate environment, on which Choteños feel more the sense of freedom while dancing. The sense of generating an intimate environment to perform La Bomba was much stronger in previous generations than nowadays.

iii. La Bomba in Chota as a shared experience: The tight relation of Choteños with their land and their families: Bouisson (1997) names two principles that were imposed by the Jesuits and that would serve as a future reference for Choteños, to not split the members of each family and to not be transferred from one hacienda (estate) to another nor outside of Chota Valley. Because of this Choteño experience with Jesuits and in spite of their conditions due to slavery, Choteños got used to being among their families in their hacienda (estate) and never accepted any other reality (Bouisson, 1997; Savoia & Ocles, 1998).

The historical Choteño' notion of being among relatives, friends and compadres (godparents) in Chota is also reflected in most of Choteño daily activities. Pabón (2007) narrates, "before the electric light came to the villages of Chota Valley [1977], after dinner, while the elders gathered among each other to talk, youngsters used to go to play during the nights of full moon until a specific time, then they will go to sleep or to listen to stories from their parents and grandparents".

Male and female Choteños cooked, rested, cleaned and washed clothes and dishes in groups that included all ages. Several authors (J. Chalá, 2006; O. Chalá, 2004; Costales & Peñaherrera, 1959; De-La-Cruz, 2012; Franco, 2000; Guerrón, 2000; Hernández, 2005; Santillán, 2006; Valarezo, 2010) locate La Bomba as one of the usual activities that is performed by Choteños as a group, a shared experience.

E. Performing "our" Bomba from Chota

Most Choteños refer to La Bomba as something that has historically belonged to them. Within this Choteño sense of belonging with La Bomba two elements are recognized. First, there is an association of La Bomba with what Choteños call "being black", and secondly with their ancestors.

Regarding La Bomba related with Choteño people's ancestors, Karla Aguas, an adult female who is a dancer from the village of Chota, asserts that "La Bomba is from our ancestors…" (Ruggiero, 2009, 45 min 24 s). Gualberto Espinoza says, "La Bomba is passed down, and little by little is gets embedded within the bodies. The ear becomes familiar with the rhythm. It does not matter how strong the influence of other rhythms. La Bomba is always going to be in our memory because it is from our families, from our ancestors, from our parents. The paternal rhythm is always going to be more important" (personal communication, November 2013).

The results show that for Choteños, La Bomba belongs to an "us" as a group. This "us" is Choteño. Choteños refer to themselves as a group of people with African roots, and with ancestral roots traced to Afrodescendants who were born in Chota Valley.

F. The socializing function of La Bomba

The characteristic of La Bomba in Chota that is most often mentioned by Choteños is it's gathering or socializing function. This characteristic appears to be as old as La Bomba itself since. As pointed out by Medina (1996), "during the slavery period Choteños were destined to only do repetitive and tiring works, even surpassing their physical strength, enslavers took from slaves any possibility of developing abilities within their group and almost all possibilities of human interaction where annulled. In this context, La Bomba acted as perhaps the only space in which they could gather and share together" (p. 108). Until the present days, La Bomba integrates and unifies the entire community.

G. Learning La Bomba

Choteños in Chota Valley learn La Bomba from a very young age (See fig. 15) through careful observation of the elders. None of the interviewees affirmed having been taught how to dance or play La Bomba. Some of the interviewees shared that at times they even had to hide themselves while watching elders in order to learn the dance since elders were not always willing to let younger Choteños see their techniques.

Regarding the age at which La Bomba is learned, Teodoro Mendez, an old male Choteño, says, “Before learning how to walk, when we are babies, we are moving our bodies while sitting down, we are hitting the tables, the chairs with a rhythm. We actually learn how to dance La Bomba before we learn how to walk”. Nelly Calderón Chalá, a woman approximately 50 years old of Chota Village says, “We learned to dance La Bomba when we were kids, with that we also learned to speak and to sing” (Ruggiero, 2009, 40 min 58 s).

Regarding the way of learning to dance La Bomba, Belermina states that, “I, by myself, got interested on learning, so I used to look at the elders, to my mom or her friends” (Pabón, 2010). “Nobody told my friends, or me, look! This is how you dance! do this! Do that! No! It wasn’t like that! We learned by ourselves. One used to look at the elders and imitate them” (B. Congo, personal communication, December 2013). Similarly, Eudocia Chalá narrates, “I used to chase my older cousin to watch her dancing, she knew how to dance, so I learned, just by looking at her” (personal communication, December 2012).

H. The current status of La Bomba in Chota

At the present, La Bomba still occupies a central role within the daily group activities of Choteño people (J. Chalá, 2006; O. Chalá, 2004; Costales & Peñaherrera, 1959; De-La-Cruz, 2012; Franco, 2000; Guerrón, 2000; Hernández, 2005; Santillán, 2006; Valarezo, 2010). La Bomba is still being danced very frequently among children, adult and old Choteños in Chota Valley and even in Choteños’ neighborhoods in the main cities of Ecuador, on which each neighborhood usually has a specific disco-bar exclusively for the performance of La Bomba and Salsa. La Bomba is very often performed in Chota Valley. The frequency of performance is each week. In some communities La Bomba is performed during specific times two or three days per week as part of religious celebrations, marriages, baptisms or as a weekend activity. “Any occasion is good to dance La Bomba (, p. 191).

II. Heritage Concepts

The heritage term is defining as the crucial and central part of the research, we can refer it to ’heritage is those items and places that are valued by the community and is conserved and preserved for future generations’ [8]. The concept is much wider than historical buildings. It iDemosntrate, Validate, Benefits Human do different activities, how well includes items and places with natural heritage significance and Aboriginal heritage significance. ’The heritage value of a place is also known as its cultural significance which means its aesthetic, historic, scientific, social or spiritual value for past, present or future generation’[8].

One of the principal keywords in this research is Heritage. UNESCO is one of the Entities that refers in an accurate way to this definition. The expert meeting defined a heritage route as ’composed of tangible elements of which the cultural significance comes from exchanges and a multi-dimensional dialogue across countries or regions that illustrate the interaction of movement, along the route, in space and time’[9]

But the question is what is heritage and which parameters defining the artifact or the information as a heritage? The context and the interpretation of data is the answer.

The heritage term is defining as the crucial and central part of the research, we can refer it to ’heritage is those items and places that are valued by the community and is conserved and preserved for future generations’ [8].

What is heritage? Defining the term like the crucial and central part of the research. People commonly equate heritage with historic buildings but the concept is much wider and includes items and places with natural heritage significance and Aboriginal heritage significance. Although these categories are covered by different pieces of legislation, in reality they often overlap. Heritage is what we inherit, but more specifically what we retain of this inheritance. The concept is much wider than historical buildings. It includes items and places with natural heritage significance and Aboriginal heritage significance. ’The heritage value of a place is also known as its cultural significance which means its aesthetic, historic, scientific, social or spiritual value for past, present or future generation’[8].

One of the principal keywords in this research is Heritage. UNESCO is one of the Entities that refers in an accurate way to this definition. The expert meeting defined a heritage route as ’composed of tangible elements of which the cultural significance comes from exchanges and a multi- dimensional dialogue across countries or regions, and that illustrate the interaction of movement, along the route, in space and time’ [9].

But the question is what is heritage and which parameters defining the artifact or the information as a heritage? The context and the interpretation of data is the answer.

Heritage is not mechanic things. It is related to the syntax, context, meaning and behaviour. This is a misunderstood concept of the Heritage. It is always to pass to the future generations. There is a value to keep in the future generations and it should not be as a issolated element.

Tha data is often in issolation. But the data needs to be with the connections and relationships. It gives the meaning of the information. If that heritage is not preserve in the future the information can be lost.

Ther is an effective serendipituos use. In context of Big Data is a problem. We have the metadata, how I can incorporate in metdata context? The problem of Big Data because the more data we have, we can get with or without Heritage. The most data we can get, the most data could be lost. There are a large Data Sets and they are the new knowledge in the future.

Can have value in the future? That is why the need of the definition of Congnitive, Contextual and Physical

A. General frameworks for heritage

The development of the preservation framework is related with the value of information. 'Value has always been the reason underlying heritage conservation. It is selfevident that no society makes an effort to conserve what it does not value"[8]. The Value of the information is located in the second level of importante after the concept of Heritage. If we can define the result that we want, how we can manage and measure the value of that information? The principal ways are the perception, interpretation and contextualization.

III. Interpretation

Modelling Theory

This chapter demonstrate general background of digital preservation and data structure.

- MODEL DRIVEN APPROACH
- HYPOTHETIC TESTBED
- BUILD THE TOOL
- USE THE TOOL
- SIMULATION

It involves case studies and action studies. Primary Study qualitative research Observation, Data Collections.

In terms of Quantitive evaluation there are following proposals:

- Empirical Experiments
- Testing the ideas
- Case studies propose model

On the other side Qualitaive evaluation is based on:

- Running Experiments

The use os the Tools will be UML, Archimate, Bonita Soft.

IV. General Research Methodologies

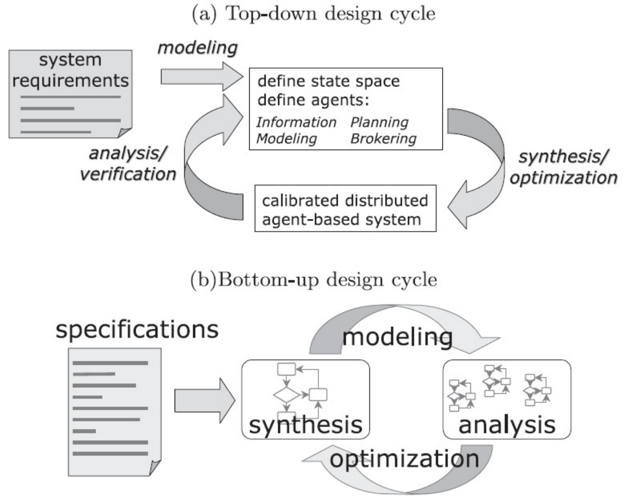
The three known basic patterns in software development are the waterfall, spiral, and prototyping. A software development method is a method you used in software development. This is known as software or software lifecycle process called. Software Development Methodologies

Why are there so many software development methodologies? Are circulating even estimated that 50% of the software goes wrong. A well-known, road pricing mistake. Therefore one is looking for a method that works. For decades they try to find predictable processes to improve productivity and quality. Some models try to systematize and formalize software development. Others apply project management techniques to writing software. Various Software Development Methods.

A. Top Down

The top-down methodology has been recently developed to produce probably perform designs relative to what is achieved in classical centralized control theory. The design process consists of three steps: modeling, synthesis and analysis/optimization.[10]

Traditionally, two alternative design methodologies, called top-down and bottom-up, have been used in building complex systems. Under these conditions the properties of a classical centralized solution to the global specification are expected to hold, up to some tolerable performance degradation, also in a decentralized environment.


(a) Top-down design cycle

(b)Bottom-up design cycle

Fig. 4. Conceptual representation of the two design methodologies top multi agent [9]

V. Cost Model for Digital Preservation

We have applied the OAIS functional entities Ingest, Archival Storage, Data Management, Administration, Preservation Planning, Access, and Common Services. Furthermore we have included the OAIS roles of Producer, Consumer, and Management, as placeholders for external cost factors, which influence the cost of preservation.

- Producer who performs the dance La Bomba
- Entities cost-critical activities
- The basic formula for an activity is the effective time required to complete an activity (measured in pw) multiplied by the wage level, plus purchases (monetary value). [11]

Bottom–Up The bottom-up design methodology is known for producing autonomous, scalable and adaptable systems often requiring minimal (or no) communication. The design process consists of three steps: Synthesis, Modeling and Analysis, and Optimization.

1) Cost equations:

$$Costperactivity = (Time \times Wage) + Purchase \quad (1)$$

$$c(a) = \sum_{i-0}^{N1} t_i * \sum_{i-0}^{N1} W_i + P \quad (2)$$

Costing Preservation Planning and Digital Migrations while the goal is to model the whole lifecycle of digital preservation. The first version of the model only deals with the cost of Preservation Planning and digital migrations.

- The amount of documentation (number of pages) is one of the principal factors

- The complexity of the documentation (low, medium, high)
- The quality of the documentation (low, medium, high),

FI means Formal Interpretation.

$$FI = \#pages \times timeperpage \times complexity \times quality. \quad (3)$$

VI. Conclusion

- Detailed infomation related with La Bomba give us a brief understanding of sociological issues around this dance. The best methodology is how to represent the entities and its interpretation.
- The context, relation and situation of the Serendipitous Heritage are impressive relevant in the research because it gives the sense of the future of the Knowledge in the World. Through the Socio - Technical, Cultural fields, the process of Preservation will do a contribution for the Memories of the World.
- The Business Process Management give us a good approach to the development of Performance and Data Preservation. Through process the increase of data can be justified.
- According to this consideration it is important to mention the type and structure of data. Through the time preserving digital information has a process for designing a practical system for managing massive amounts of critical data. The way to improve the understanding of the methodology, the information has to consider two dimensions: access dimension and cognitive dimension. Both of them have the level of importance in terms of the results. As a methodology of treatment digital preservation, it could be risky even when the strategy could develop a clear idea of digital resources and digital artefacts. The approaches related with other authors have similarities and differences in opinion.

References

[1] Coba, C. A. (1980) Literatura Popular Afroecuatoriana [Afroecuadorian Popular Literature]. Vol. 43. Colección Pendoneros-Serie Cultura Popular. Otavalo: Instituto Otavaleño de Antropología.

[2] Pabón, I. (2007). Identidad afro: Procesos de Construcción en las Comunidades Negras de la Cuenca Chota Mira [Afro Identity: Construction Processes in Black Communities of the Chota-Mira Basin]. Quito: Abya-Ayala.

[3] Coba, C. A. (1985). Danzas y Bailes En El Ecuador [Dances in Ecuador]. Latin American Music Review, 6(6), 166-200.

[4] Guerrón, C. (2000). El color de la Panela [The color of brown sugar]. Universidad Politécnica Salesiana, Ibarra.

[5] Costales, A., & Peñaherrera, P. (1959). Coangue: Historia Cultural y Social de los Negros de El Chota y Mira [Coangue: Cultural and Social History about Black people in Chota and Mira] Llakta, 7.

[6] Aguirre, C. (2005). En Ecuador domina el racismo [In Ecuador racism rules]. El Universo. Retrieved from http://www.eluniverso.com

[7] Bathurst, "Bathurst Council," 2014. [Online]. Available: http://www.bathurst.nsw.gov.au/building-and/heritage/what-is-heritage.html.

[8] UNESCO, "Information Document Glossary of World Heritage Terms (June, 1996)." [Online]. Available: http://whc.unesco.org/archive/gloss96.htm.

[9] P. A. Sabatier, "Top-down and bottom-up approaches to implementation research: a critical analysis and suggested synthesis," J. Public Policy, vol. 6, no. 01, pp. 21–48, 1986.

[10] U. B. Kejser, A. B. Nielsen, and A. Thirifays, "Cost model for digital preservation: Cost of digital migration," Int. J. Digit. Curation, vol. 6, no. 1, pp. 255–267, 2011.

[11] Rahier, J. M. (1998). Blackness, the Racial/Spatial Order, Migrations, and Miss Ecuador 1995-96. American Anthropologist, 100(2).

[12] Llambi, M. (2010). Esclavitud en el Río de La Plata: Historia Negra [Slavery in Río de La Plata: Black History]. El Federal, Año 6, 22-31.

Ontological Metamodel for Consistency of Digital Heritage Preservation (DHP)

Lucia Carrion Gordon, Zenon Chaczko
Faculty of Engineering and
Information Technology (FEIT)
University of Technology Sydney (UTS)
Australia, NSW, Sydney
Email: Lucia.CarrionGordon@uts.edu.au
Email: Zenon.chaczko@uts.edu.au

***Abstract*—In this stage of Data Preservation the challenge is how to keep the attributes of the data and how to preserve the originality. It is like to keep the living part of the data. It is how the concepts of Heritage have sense. Heritage is the concrete data, it gives the interconnection to other aspects of the reality. Nowadays the physical value and the aspects of items complete the relevance of information. The relation between Preservation and Digital patterns of Heritage is well related because of the two aspects to consider: Accessibility and Context.**

Keywords. Data, Preservation, Digital, Heritage, Ontology, Management, Meta-Model

I. Introduction

There are two tendencies around the understanding of the management of the ideas. The ontology and the Epistemology of this study, centralized the future use and the Serendipity tendency of the item. However in the perspective of the nonphysical items there is a World of Physical and Logical and how the Preservation need to look items and how will be the manifestation. Digital Preservation has evolved into a specialized, interdisciplinary research. Through the time the challenge in to jointly develop solutions.. As the patterns and alternative solutions there are Information Retrieval and, Machine Learning or Software Engineering. The real fact of Digital Preservation show us the reality of the understanding of the World about the facility to have digital expressions rather than just physical.

The Heritage of the collected information define the quality of the Data. At this stage, the definition of Heritage involved the presence not only the content. It is the express by itself the real meaning of the data. The perception of the importance and relevance of the information is measured through the definitions and metamodel that is proposed. Digital Data and Heritage Preservation as concepts are related to data management, contextualization and storage. There are many issues and concerns around it. This research explores the precise definition, context and the need of patterns of heritage. The relations, interpretation and context give us the appropriate methods to keep information for a long term use. The management of massive amounts of critical data involves designing, modeling, processing and implementation of accurate systems. The methods to understand data have to consider two dimensions that this research has to focus on: access dimension and cognitive dimension. Both of these dimensions have relevance to get results because at the same time, ensure the correct data preservation. Our cultural heritage, documents and artefacts increase regularly and place Data Management as a crucial issue. The first stage involves exploration and approaches based on review of recent advances.

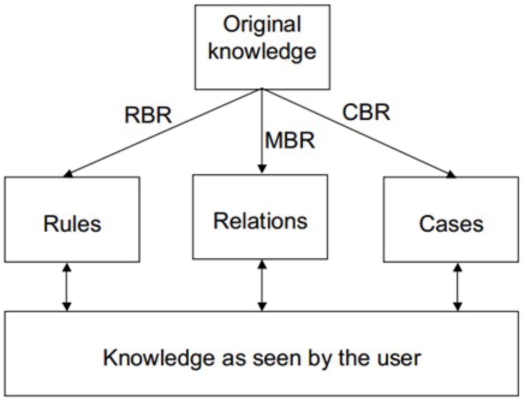
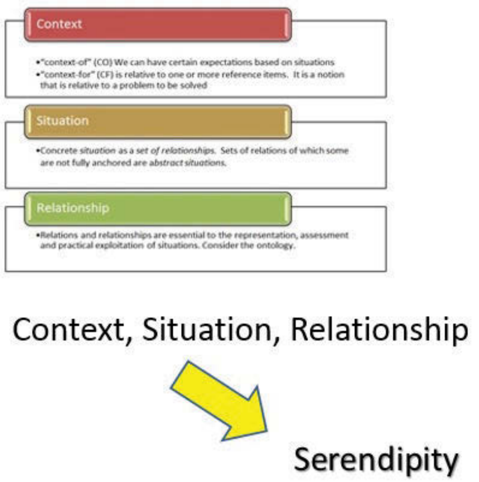


Fig. 1. General Approaches

The second stage involves adaptation of architectural framework and development of software system architecture in order to build the system prototype. Increasing regulatory compliance mandates are forcing enterprises to seek new approaches to managing reference data. Sometimes the approach of tracking reference data in spreadsheets and doing manual reconciliation is both time consuming and prone to human error. As organizations merge and businesses evolve, reference data must be continually mapped and merged as applications are linked and integrated, accuracy and consistency, realize improved data quality, strategy lets organizations adapt reference data as the business evolves.

A. Patterns for DHP

One of the primarily concerns about the explosive amount of information and the complexity of the classification, is how to keep the principal characteristics data. A need to move away the traditional understanding of Heritage reflects the real



Context

- "context-of" (CO) We can have certain expectations based on situations
- "context-for" (CF) is relative to one or more reference items. It is a notion that is relative to a problem to be solved

Situation

- Concrete *situation* as a *set of relationships*. Sets of relations of which some are not fully anchored are *abstract situations*.

Relationship

- Relations and relationships are essential to the representation, assessment and practical exploitation of situations. Consider the ontology.

Fig. 2. Vision of Context

meaning of the data. More artifacts and everyday life tendency is to have less physical representation in the World of Logic. The representation of the items refers to the tendency of more things nonphysical and ow through the Heritage it passes the attributes

What is heritage conservation? A brief overview Heritage conservation doesn't mean freezing a building in time, creating a museum or tying the hands of property owners so they can't do anything with their properties. Instead, it seeks to maintain and thereby increase the value of buildings by keeping their original built form and architectural elements, favouring their restoration rather than replacement and, when restoration is impossible, recreating scale, period and character. Heritage Conservation provides concrete benefits to property owners, to businesses and to the community as a whole: Heritage preservation and designation increases property values, both of the restored building and surrounding properties. Heritage preservation can be a draw to tourism and helps businesses attract customers. Communities, such as Meaford fortunate to have a significant stock of heritage buildings can build their town or city's image around those elements: Toronto's Distillery District, Niagara-on-the-Lake and Merrickville are good examples. Retaining the historic integrity of a neighbourhood or downtown attracts people just for that ambiance alone and that attracts business. A small town without a heritage main street attracts no one. Restoration keeps money within the community, by requiring fewer materials from outside and more labour-intensive work by local trades. With the right programs in place, businesses and building owners can take advantage of government programs and incentives to maintain and restore heritage buildings. Restoration reduces construction and demolition waste and uses less than half the energy of new construction. Heritage preservation is an investment in our community that rewards us today and leaves an invaluable resource for future generations.

Deep learning (deep machine learning, or deep structured learning, or hierarchical learning, or sometimes DL) is a branch of machine learning based on a set of algorithms that attempt to model high-level abstractions in data by using multiple processing layers with complex structures, or otherwise composed of multiple non-linear transformations. Deep learning is part of a broader family of machine learning methods based on learning representations of data. An observation (e.g., an image) can be represented in many ways such as a vector of intensity values per pixel, or in a more abstract way as a set of edges, regions of particular shape, etc.. Some representations make it easier to learn tasks (e.g., face recognition or facial expression recognition [6]) from examples.

One of the promises of deep learning is replacing handcrafted features with efficient algorithms for unsupervised or semi-supervised feature learning and hierarchical feature extraction.[7] Research in this area attempts to make better representations and create models to learn these representations from large-scale unlabeled data. Some of the representations are inspired by advances in neuroscience and are loosely based on interpretation of information processing and communication patterns in a nervous system, such as neural coding which attempts to define a relationship between various stimuli and associated neuronal responses in the brain.[8] Various deep learning architectures such as deep neural networks, convolutional deep neural networks, deep belief networks and recurrent neural networks have been applied to fields like computer vision, automatic speech recognition, natural language processing, audio recognition and bioinformatics where they have been shown to produce state-of-the-art results on various tasks.

Values are the subject of much discussion in contemporary society. In this postmodern, post-ideology, post-nation-state age, the search for values and meaning has become a pressing concern. In the field of cultural heritage conservation, values are critical to deciding what to conserve — what material goods will represent us and our past to future generations — as well as to determining how to conserve. Discussions of values, of how social contexts shape heritage and conservation, and of the imperative of public participation are issues that challenge conventional notions of conservation professionals' responsibilities. Values is most often used in one of two senses: first, as morals, principles, or other ideas that serve as guides to action (individual and collective); and second, in reference to the qualities and characteristics seen in things, in particular the positive characteristics (actual and potential).

B. *Ontology*

The formatting of information provides the unique result as a digital age of the information. Other objective is the knowledge management and Ontology as techniques for analyzing information. Also could develop a system could give a result and next steps with a specific process. One concern of digitalization would be the format, standards and migration of the data. It should be solved with the use of Architectural Methodology and with the development of a fast prototype. This requires the definition of the sequential process. First, should be considering a Framework as a whole front end and

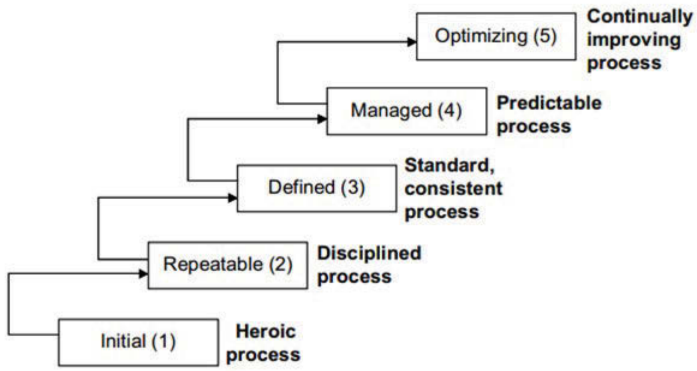


Fig. 3. Mature Model

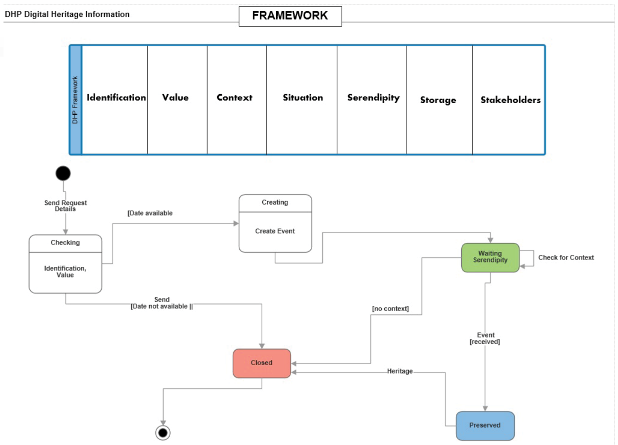


Fig. 4. Context [1]

for reception of the information in a basic way. Using the Open Group Architectural Framework (TOGAF) it is found to be more suitable as a result of dissemination of the data. Meanwhile, it is associated with an Ontology and Knowledge Management terms to be more specific and with a deep sense of definition. Second the Methodology with an architectural vision using concepts of Architectural Development Method (ADM). It has been identified in terms of enterprise description for validating information of several types of data. Third, the conceptualization of patterns for a centralization of the preservation knowledge providing a unique result: the digital age of the data. The connection is also with the artefacts and the correct use of them. The recovering of the information is another issue has to considering between the techniques used for this purpose like: migration and emulation. The challenges and constraints shown by the type of data classified as structured and unstructured information are reflected in the requirements of each field. The evaluation is based on PRIST model defining by Privacy, Rights, Integrity, Security and Trust across PC considerations related with Physical and Cognitive characteristic of data. The authenticity of the information and the reliability of the same is the principal challenge of the study. The concepts of e-infrastructure are useful for the evaluation of requirements in specific cultural matters like Libraries. According to the author the correct use of the interface and the exact generation of metadata, are the key considerations to follow around the process of optimal data preservation.

TOGAF +ADM+PRIST

C. Architectural Framework

Introduce a consolidated, systematic approach to the redesign of a business enterprise. The proposed framework includes

- Identification
- Value
- Context
- Situation
- Serendipity
- Storage
- Stakholders

D. Active Structures

4. Metamodel Explore Metdata Models for Dynamic Data Representation Data Aging are the new relations and mechanism. For the experimentation there are three bases for the advances in this term. 1. Matlab: with accordance with the toolbox, and to have proficiency 2. SOM Self Organizing Maps: Representation and maps 3. Heuristic LAB Through this topic there are specific concepts related with e The Heritage Preservation Digital Data and Historical Analytics. The relation is the IoT manly cloud, because the Heritage Preservation as Data Analytics Accessing Digital Cultural Heritage • So we are getting content-access ... and there will be trickle-down into digital libraries for cultural heritage • Digital Cultural Heritage in 10 years ... • New technologies, new media forms • New forms of Digital Cultural Heritage • New interaction modalities ... tablets, augmented reality, brain sensing from wearables • New ways to navigate where we control • ... but these are "just technology" changes and evolution • Biggest change will be in people, expectations, demands because People want to be in control SOM

The future challenges of the conservation field will stem not only from heritage objects and sites themselves but from the contexts in which society embeds them. These contexts—the values people draw from them, the functions heritage objects serve for society, the uses to which heritage is put—are the real source of the meaning of heritage, and the raison d'être for conservation in all senses.

For the formulation of the Hypothesis the relation between keywords and statements is important consideration for imporvement of the proposal model.

The relation between Software Architecture and Serendipitous Heritage is going to improve Data Preservation Heritage oriented metadata for improving the real usage of the information.

The inclusion of Serendipitous Heritage improves performance, scalability, dependability, manageability and data ac-

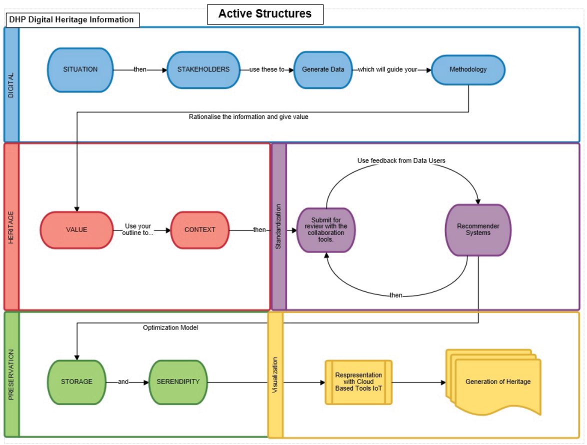


Fig. 5. Active Structures in DHP

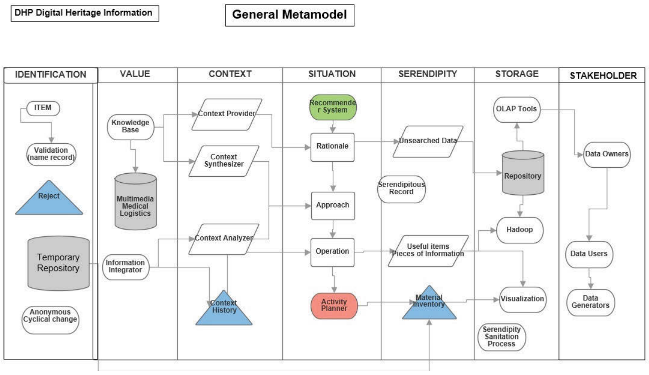


Fig. 6. 'Rational economic model of decisionmaking ' [3]

cess for Digital Data Preservation Mechanisms in Big Data Architectural solutions. The important knowledge, exactly the context situation relationship and concept. The best way uis to demonstrate, validate and show the benefits. The mankind do different activities, how well the Serendipitious Heritage concept will help to grow the meaning of Data in every field.

The massive amount of data and the growth of Big Data drive the society to preserve the information prinicipally related with the lost of key information.The protagonism in the role of metadata and the requeriment that data has to be keep in a long term open the alternative to focus on information management.

II. General DHO Process based on Metamodel

Based on the exposed statements, the Metamodel has to include the seven stages for the frameworkThe relationship between these terms is given for the behaviour and the treatment of data. There are examples referenced by known authors. The sustainability of the preservation of the information give us the discussion about the appropriate resources infrastructure.

A. Metamodel

Nowadays Electronic publications, and a collection of multimedia art are the principal issues that need to preserve their data holdings, simulation models, or studies over time. The other matters that we have to consider is the legal constraints, to guarantee the accessibility and usability over time. Moreover the Digital Preservation is considering as a research discipline.

The two methods related in this study is: migration and emulation. There are often hold very valuable data in complex structures.

Migration is the method of repeated conversion of files or objects. Emulation denotes the duplication of the functionality of systems, be it software, hardware parts, or legacy computer systems as a whole, needed to display, access, or edit a certain document[2].

The integration of inhomogeneous criteria sets is used to evaluate different strategies. For example web archive collections, collections of scientific publications, and electronic multimedia art. The collaborative research programm is funding research in different aspects of digital preservation, including collection practices, risk analyses, legal and policy issues, and technology. In Europe two new research projects. Scientific information is born-digital or only available in digital form. At the moment libraries, archive and scientific institutions are primarily dealing with the challenge of long term preservation[2].

The main idea of this research is related the management of heritage. It could be useful as a cultural case study related directly in terms of preservation. It exposes the experimental methodology and a valid analysis of the results. The reliability and accuracy of data are very strategic points in this article because the knowledge base is wide and complete.

On the other hand is important to differenciate the terms “data”, “information”, “knowledge” and “knowledge management” is almost as large as a number of authors contributing to the field . 'Data is a set of discrete, objective facts'. 'Information is data that has been organized or given structure that is, placed in context and thus endowed with meaning'. 'Knowledge is information combined with experience, context, interpretation and reflection that is ready to apply to decisions and actions'[3].

III. Considerations

- The strategies for commercialization have to define: the niche of the market, the receptor, the affections to the community and the results. The deadlines are also important, because the best solution is not only the best investigation, also could be the solution that take place in the appropriate time.
- Definition of the rights and property laws. We have to be aware to know the implications of the Intellectual Property. Furthermore, the alliances and if some enterprise want to buy our idea, is good to have arguments.
- The long-term following. It means that the solutions and the business point of view have to have sustainable duration. The innovation never ends. In general, I also believe that to focus in the new markets could be a right way to learn. Sometimes, to mix the fields in the knowledge could have better results than only focus

in one solution. For example, technology and health, mechanics and biotechnology will give us an idea. Finally I strongly believe that the relation between research and commercial field are undisputed.

IV. Conclusion

- The proposed Meta-model allow to develop an alternative for the understanding of Heritage Preservation
- The dimensions around the data is treated, related with the basic considerations about the origins of the information.
- The different dimensions of the Digital Heritage Preservation develops a real significance of Heritage. The non-physical world shows the opportunity to do Information Management.
- The context, relation and situation of Heritage are impressive relevant in the research because it gives the sense of the future of the Knowledge in the World. Through process of Preservation will constitute a contribution for society advance.
- Information management through the uses of Technology help us to develop a real proposal.
- The use of tools like Hadoop, Softwarch, Archimate and Bonitasoft, the concepts of Software Architecture will have a real approach and meaninful characteristics for the relevance of the investigation.
- The Business Process Management give us a good approach to the development of Performance and Data Preservation. Through process the increase of data can be justified.

V. Future Projects

Data preservation: Digitalization of the Heritage, the result of proposal is to have like a result of the experimental work, a reliable Framework for measure the digital age of the information and patterns that qualified usability and accessibility of the data. The best pathway for commercialization could be some of them.

- Commercial Business Structure like a Partnership assuming the cost of the investment and the taxes that generate the buying of the equipment for implementation of the scanning in the digitalization.
- Initial Public Offering IPO, because the application of the data preservation could be focus on Entities from Government and Historical materials and artifacts that sometimes have to be preserved with a public responsibility.
- This research could have through the market with POCs proof of concepts, showing the advantages and challenges of the new solution. In this case the relationship between the process and the final patterns there is a model. The final product is a open source software through that, I can advance and model the real situation of the information and the result will be the digital age of the data. The services of consultancy will be the product and service for the end user. About the Funding, the budget from Universities and Entities from the Government could cover the IP registration and we can manage a specific amount for the legal issues. To share the capital and the risks is optional, but in research the vision of the future of the project is the main concern for knowing how to sell the solution. For instance, the consultancy services could be mixed with the implementation of the concrete result in the systems development. Finally, the awareness and the knowledge have to be register, and the experimentation could be shared with public and private investment for using in legal, financial and technological issues.

References

[1] S. Thalmann, I. Seeber, R. Maier, Ren, #233, Peinl, J. M. Pawlowski, L. Hetmank, P. Kruse, and M. Bick, “Ontology-based standardization on knowledge exchange in social knowledge management environments,” Proceedings of the 12th International Conference on Knowledge Management and Knowledge Technologies. ACM, Graz, Austria, pp. 1–8, 2012.

[2] S. Strodl, C. Becker, R. Neumayer, and A. Rauber, “How to choose a digital preservation strategy: evaluating a preservation planning procedure,” Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries. ACM, Vancouver, BC, Canada, pp. 29–38, 2007.

[3] M. Kaner and R. Karni, “A capability maturity model for knowledge-based decisionmaking,” Information, Knowledge, Syst. Manag., vol. 4, no. 4, pp. 225–252, 2004.

[4] C. Becker, Gon, #231, alo Antunes, Jos, #233, Barateiro, R. Vieira, and Borbinha, “Modeling digital preservation capabilities in enterprise architecture,” Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times. ACM, College Park, Maryland, pp. 84–93, 2011.

[5] S. Challa, T. Gulrez, Z. Chaczko, and T. N. Paranesha, “Opportunistic information fusion: a new paradigm for next generation networked sensing systems,” in Information Fusion, 2005 8th International Conference on, 2005, vol. 1, p. 8 pp.

[6] L. McCay-Peet, “INVESTIGATING WORK-RELATED SERENDIPITY, WHAT INFLUENCES IT, AND HOW IT MAY BE FACILITATED IN DIGITAL ENVIRONMENTS,” 2014.

[7] Olson, J. M., & Janes, L. M. (2002). Asymmetrical impact: Vigilance for differences and self-relevant stimuli. European Journal of Social Psychology.

[8] Gulden, J. 2013, ’Methodical support for model-driven software engineering with enterprise models’, Universität Duisburg-Essen

[9] Strodl, S., Becker, C., Neumayer, R. & Rauber, A. 2007, ’How to choose a digital preservation strategy: evaluating a preservation planning procedure’, paper presented to the Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries, Vancouver, BC, Canada.

[10] [5] Neumann, A., Miri, H., Thomson, J., Antunes, G., Mayer, R. & Beigl, M. 2012, ’Towards a decision support architecture for digital preservation of business processes’, Proceedings of the 9th Int. Conf. on Digital Preservation (iPres 2012), Citeseer. . Harlow, England: Addison-Wesley, 1999.

[11] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1

[12] Gordon, L. & Chaczko, Z. 2015, ’Digital Patterns for Heritage and Data Preservation Standards’, in G. Borowik, Z. Chaczko, W. Jacak & T. Łuba (eds), Computational Intelligence and Efficiency in Engineering Systems, vol. 595, Springer International Publishing, pp. 47-59.

[13] Engelsman, W., Jonkers, H. & Quartel, D. 2011, ’ArchiMate® Extension for Modeling and Managing Motivation, Principles, and Requirements in TOGAF®’, White Paper, The Open Group.

[14] Bathurst 2014, Bathurst Council, <http://www.bathurst.nsw.gov.au/building-and/heritage/what-is-heritage.html

Architectural Framework to Preserve Information of Cardiac Valve Control

Authors : Lucia Carrion Gordon, Jaime Santiago Sanchez Reinoso

Abstract : According to the relation of Digital Preservation and the Health field as a case of study, the architectural model help us to explain that definitions. .The principal goal of Data Preservation is to keep information for a long term. Regarding of Mediacal information, in order to perform a heart transplant, physicians need to preserve this organ in an adequate way. This approach between the two perspectives, the medical and the technological allow checking the similarities about the concepts of preservation. Digital preservation and medical advances are related in the same level as knowledge improvement.

Keywords : medical management, digital, data, heritage, preservation
Conference Title : ICICSP 2015 : 17th International Conference on Information, Communications and Signal Processing
Conference Location : Venice, Italy
Conference Dates : November 09-10, 2015

International Science Index, Information and Communication Engineering Vol:9, No:11, 2015 waset.org/abstracts/39341

The Metamodel of Heritage Preservation for Medical Big Data

Zenon Chaczko[1], Lucia Carrion Gordon[1], Wojciech Bozejko[2]
[1] University of Technology, Sydney Australia, Sydney
[2]Institute of Computer Engineering, Control and Robotics, University of Technology, Wroclaw, Poland
Zenon.Chaczko@uts.edu.au[1], Lucia.CarrionGordon@uts.edu.au[1],
Wojciech.Bozejko@pwr.wroc.pl[2]

Abstract. At present the real challenge of Digital Data Preservation concerns methods of keeping all important attributes of the data and preserving their originality. The key is to keep the living part of the data. It is the essence of the Heritage concept. The Heritage is about the concrete data, the concept gives the interconnection to other aspects of the reality. Nowadays the physical value and the aspects of items complete the relevance of information. The question of what is the heritage and what parameters define the artifacts' heritage? The context and the interpretation of data is the answer. The heritage term is defining as the crucial and central part of the presented research, and we can refer the heritage term as: '…those items and places that are valued by the community and is conserved and preserved for future use or reference by the future generations'.

Keywords: Metamodel, Preservation, Digital, Heritage, DHP, Ontology, Management,

1. Introduction

There are two tendencies around the understanding of the management of the ideas. The ontology and the Epistemology of this study, centralized the future use and the Serendipity tendency of the item. However, in the perspective of the nonphysical items there is a World of Physical and Logical and how the Preservation need to look items and how will be the manifestation. Digital Preservation has evolved into a specialized, interdisciplinary research. Through the time the challenge in to jointly develop solutions. As the patterns and alternative solutions there are Information Retrieval and, Machine Learning or Software Engineering. The Digital Preservation [1] show us the reality of the understanding of the World about the facility to have digital expressions rather than just physical. The Heritage of the collected information define the quality of the Data. This is specifically important in medical field. At this stage, the definition of Heritage involved the presence not only the content. It is the express by itself the real meaning of the data. The perception of the importance and relevance of the information is measured through the definitions and the proposed metamodel. Digital Data and Heritage Preservation concepts are related to medical data management, contextualization and storage. There are many related issues. This research explores the definition, context and the need of patterns of heritage specifically in medicine. The relations, interpretation and context give us the appropriate methods to keep information for a long term use. The management of massive amounts of medical data involves designing, modeling and processing.

2. Patterns, the Metamodel and the Ontology of DHP

The Metadata Model explores dynamic data representations and specifically the new relations, their origin and the mechanism(s) that generate these relations. The formatting of information provides the unique result as a digital age of the information.

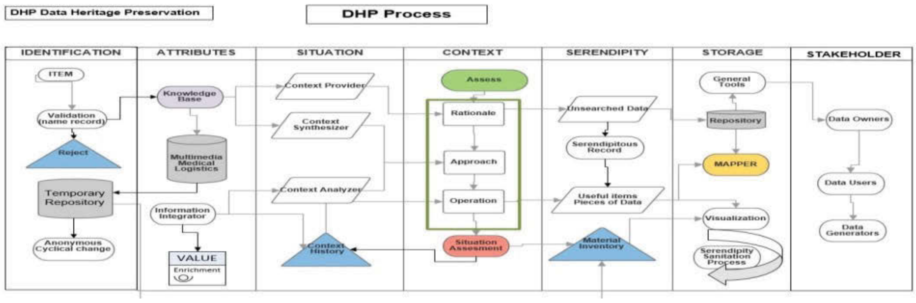


Figure 1. DHP Process

Other objective is the knowledge management and ontology [2, 3] as techniques for analyzing information. One concern of digitalization would be the formatting, standards and migration of the data. It should be solved with the use of Architectural Methodology and with the development of a fast prototype. This requires the definition of the sequential process. First, we consider a Framework front-end and for reception of the information in a basic way. Using the Open Group Architectural Framework is found to be suitable for the dissemination of the data, as it is associated with an Ontology and Knowledge Management terms and to be more specific with a deeper sense of the definition. Secondly, the Methodology with an architectural vision using concepts of Architectural Development Method. It has been identified in terms of enterprise description for validating information of several types of data. Thirdly, the conceptualization of patterns for a centralization of the preservation knowledge providing a unique result: the digital age of the data. The connection is also with the artefacts and the correct use of them. The recovering of the information is another issue should consider between the techniques used for this purpose like: migration and emulation. The challenges and constraints shown by the type of data classified as structured and unstructured information are reflected in the requirements of each field. The evaluation is based on PRIST model defining by Privacy, Rights, Integrity, Security and Trust (TOGAF +ADM+PRIST) across PC considerations related with Physical and Cognitive characteristic of data [4]. The authenticity of the information and the reliability of the same is the principal challenge of the study. The concepts of e-infrastructure are useful for the evaluation of requirements in specific cultural matters like Libraries. The proposed metamodel aims to provide an alternative for the understanding of the Heritage Preservation concept that relates to important dimensions around the processed data and its origins. The different dimensions of the Digital Heritage Preservation capture the real significance of Data Heritage.

3. References

[1] S., Becker, C., Neumayer, R. & Rauber, A. 2007, 'How to choose a digital preservation strategy: evaluating a preservation planning procedure', Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries, Vancouver, BC, Canada.

[2] Syerina Azlin Md, N. & Noor, N.L.M. 2010, 'Integrating ontology-based approach in Knowledge Management System (KMS): construction of Batik Heritage Ontology', Science and Social Research (CSSR), 2010 l Conference, pp. 674-9.

[3] Thalmann, S., Seeber, I., Maier, R., Ren, #233, Peinl, Pawlowski, J.M., Hetmank, L., Kruse, P. & Bick, M. 2012, 'Ontology-based standardization on knowledge exchange in social knowledge management environments'. Proceedings of the 12th International Conference on Knowledge Management and Knowledge Technologies, Graz, Austria. pp. 1–8, 2012.

[4] Gordon, L. & Chaczko, Z. 2015, 'Digital Patterns for Heritage and Data Preservation Standards', in G. Borowik, Z. Chaczko, W. Jacak & T. Łuba (eds), Computational Intelligence and Efficiency in Engineering Systems, vol. 595, Springer International Publishing, pp. 47-59.

Development of an Expert System to assist in Resource Management

Zenon Chaczko[1], Lucia Carrion[1], Wael Alenazy[2], Mikayla Mu[1]
[1]Faculty of Engineering and Information Technology
University of technology, Sydney (UTS), Sydney, NSW, Australia
E-mail: {Zenon.chaczko, Lucia.C.Carrion}@uts.edu.au, Mikayla.Mu@alumni.uts.edu

[2]Self-Development Skills Department, Preparatory Year Deanship
King Saud University (KSU)
Riyadh, The Kingdom of Saudi Arabia
E-mail: Walenazy@ksu.edu.sa

***Abstract*—This paper aims to demonstrate an idea of utilizing Kohonen Maps as a tool to portray and study resource allocations when constructing an expert system to assist in Resource Management. The context of work encompasses resource allocation and management tasks related to design of courses, as well as, various teaching and learning projects. The key aspect is to show the design of an expert system for resource allocation with the use of Kohonen Maps as an alternative way to visualise the demand and availability of project resources.**

Keywords—Kohonen Maps; Self-Organizing Maps (SOM); Resource Allocation; Resource Demand

I. Introduction

In today's day and age of information technology it is important to various aspects of project management in academic environment. In this work, a particular focus is to show that a balanced management of project resources related to teaching and learning, as well as course planning activities can be achieved using a computer aided resource management tool(s) which can be made available to users at various levels. Inadequately calculated and allocated resources can cause serious delays in the academic (or student) project schedules and potentially affect project costs. Often, this happens due to having to source scarcely available resources or, at times, too many resources may have been allocated which can contribute to chaotic actions of project stakeholders.

The development of the Kohonen Maps based resource allocation system can enable a different approach of portraying the demand and availability of resources.

Kohonen Maps also known as Self-Organizing Maps (SOM) has a topological structure and is a computational model. Due to the Kohonen Maps topological structure it is called a topology-preserving map which is a map that preserves the neighbourhood relations [1] [2]. Based on the neural network Kohonen Maps have the ability to classify data without supervision in a self-organizing way [2]. Kohonen Maps was considered as an idea to be used to display the usage of a resource including how popular that resource is based on how often it is demanded.

The use of Kohonen Maps will enable the user to view the demand for the resource by checking the density of the markers displayed in a generated graph. This graph will determine where the resource is required and used the most and which areas (time) they are more available (have more capacity).

Today project management software provides the user with real-time visibility into the project lifecycle [4].

Resource management is an essential component to ensure the success of a project. Different resource management strategies are used depending on the nature of the resource that is to be managed. Human resources possess certain skill sets required to do a job where as other resources such as equipment and materials are used to assist in completing a job. For resource management purposes, different resource types are considered [5]. When resources are effectively utilised, the goals of the project are achieved. The resources include the human resources and physical items including equipment and tools which are used to achieve the project objectives. Other resources include intangible items such as intellectual property, knowledge and skills. For any project the unavailability or lack of resources is a major obstacle.

This paper will examine the requirements for a resource allocation system, the designs for the resource allocation system portraying the resource demand, capacity and information. Also relates the development of a prototype of Kohonen Maps.

II. Problem Definition

A. Methodology

The method taken was based on the system lifecycle, on a prototype methodology and a research methodology for scoping.

The following steps were taken when completing this project:

1. Planning – identifying objectives of this project, scope and requirements outlined, and deadlines defined.
2. Literature Review – main research component that considered how other Project Management systems function.
3. Designing the System – developed the conceptual design based on required functionality.
4. Implementation – coding the system.
5. Integration and Test – document, build and test the system.
6. Change Management – any changes were documented.

B. *Requirements for Resource Allocation System*

The following requirements for the resource allocation system were implemented:

- Use Kohonen Maps to display the resource allocation
- Display the availability of the resource over periods of time
- Display the tasks the resource is allocated to
- Have the option to display the resource allocation using graphs
- Display the resource information including allocations in a resource profile.

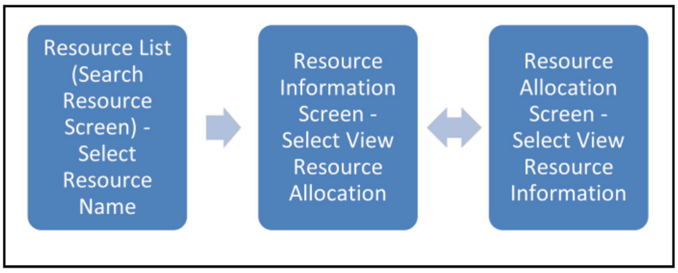

Resource List (Search Resource Screen) - Select Resource Name
Resource Information Screen - Select View Resource Allocation
Resource Allocation Screen - Select View Resource Information

Fig. 1. Example of the Navigational Flow

III. Implementation of Design

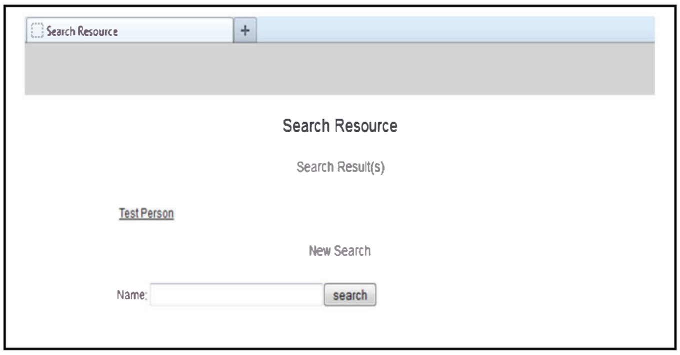
Three screens were designed.

A. *Navigation*

In order to navigate to the Resource Information Screen, the user must select the resource's name (resource's name should be a hyperlink) from a Resource List (a search page) which would then load the Resource Information Screen. From the Resource Information Screen, the user can load the Resource Allocation Screen by clicking a link in this page.

B. *Search Resource Screen*

Searching for specific resources will be done via the Search Resource Screen. Fig. 2 shows an example of the Search Resource Screen where 'Test Person' will be used as the example resource.



Search Resource

Search Result(s)

Test Person

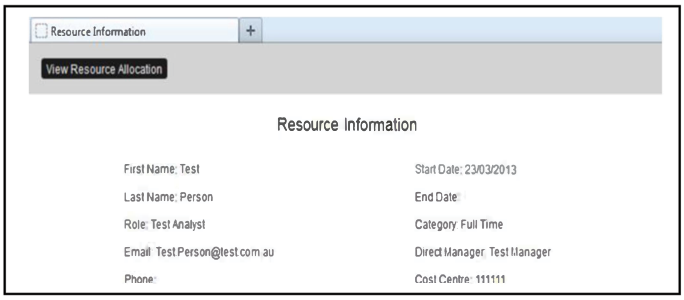
New Search

Name:

Fig. 2. Example of the Search Resource Screen

C. *Resource Information Screen*

Resource Information Screen should display the resource information ranging from their full name, their department, email, role, and location essentially it will act as a resource profile page. Fig. 3 conveys the resource information for 'Test Person'.

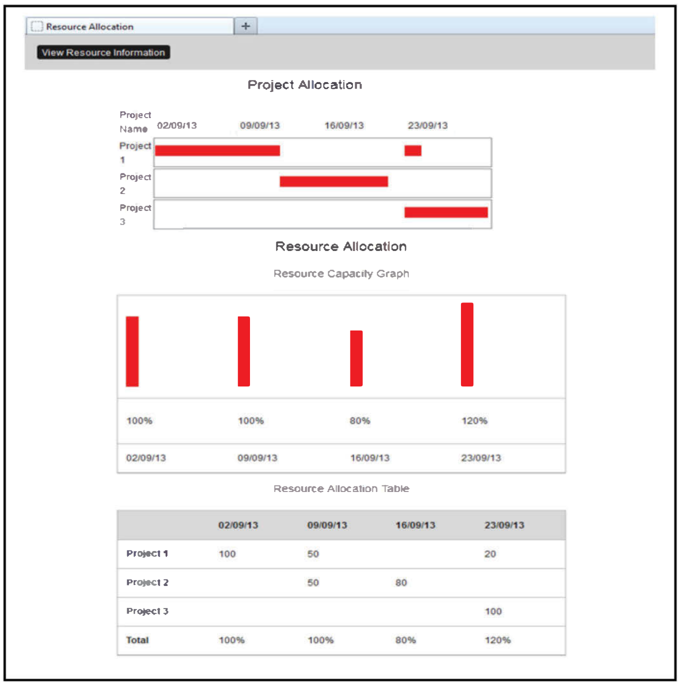


Resource Information

First Name: Test
Last Name: Person
Role: Test Analyst
Email: Test.Person@test.com.au
Phone:

Start Date: 23/03/2013
End Date:
Category: Full Time
Direct Manager: Test Manager
Cost Centre: 111111

Fig. 3. Example of the Resource Information Screen



Resource Allocation Table

	02/09/13	09/09/13	16/09/13	23/09/13
Project 1	100	50		20
Project 2		50	80	
Project 3				100
Total	100%	100%	80%	120%

Fig. 4. Example of the Resource Allocation Screen

D. *Resource Allocation Screen*

The Resource Allocation Screen should display the project allocation (bar graph), the resource capacity (column graph), and a table displaying the resource allocation (refer to Fig. 4).

The main idea will be incorporating Kohonen Maps and enabling it to display the resource allocations.

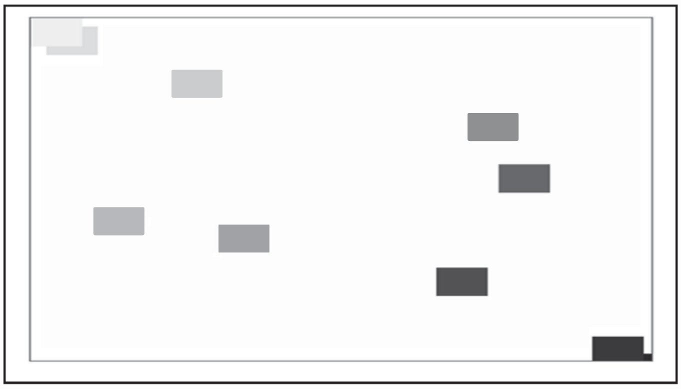


Fig. 5. Example of the Kohonen Map slightly modified from Adam Stirtan's work [3]

From Fig. 5 the Kohonen Map example portrays overlapping square shapes. This indicates a cluster and would convey an increase in resource demand, the greater the overlap the greater the resource demand.

The idea is to have the data for the graphs and the table be gathered from a database in order to generate the display.

The reason for designing a resource allocation system is so that the user can view a user friendly interface that is intuitive and will enable them to view the information they want at a glance. The basic idea is to create a system that is easy to use and relatively simple to implement.

The inclusion of the Kohonen Maps is different and could potentially show the demand of the system in a different way and could in future be used as a predictor of resource demand.

The use of Kohonen Maps could potentially path the way for a different and unique way of viewing resource demand as it can show a pictorial version of what the demand for the resource is. The basic design of the system would contain two main screens, the Resource Allocation Screen and the Resource Information Screen. Both screens will be connected to each other to enable the user to switch between the two screens.

E. *Resource Allocation in Projects*

When using project networks such as the program evaluation review technique (PERT) and the basic critical path method (CPM), both assume the availability of resources are unlimited.

When developing network schedules the resource requirements and the time constraints are considered. Since projects are restricted by three major limitations; resources, time, and performance requirements, trade-offs must be made as these limitations are challenging to satisfy concurrently. Poor resource allocation strategies can lead to quality of work being affected, and longer project schedules are expected with smaller resource bases.

As a guide, the use of the Pareto Principle (Pareto Law/Pareto Theory) can be used for planning, analysis, decision making and change management. The Pareto Principle is the phenomenon of an 80:20 rule discovered by Vilfredo Pareto (1848-1923) where the idea is that for example 20% of causes will produce a result of 80%. In analysis the use of the Pareto Theory assists in prioritising possible changes. In the analysis you would identify the problems, the cause of the problems, give a score to each problem (the higher the score the higher the priority), group the problems together based on the cause (for example lack of staff is the cause for two problems then group them together), add the scores for each group, then take action to solve the problems. In regards to resource management this can assist in allocations, deciding which resources should be allocated to which tasks in order to produce the most benefit based on prioritisation. Since trade-offs are made in projects, the Pareto method can assist in managing how best to utilise the resources, time, and performance requirements [6].

Using graphical representations to express information about resource assignment, utilisation and availability is part of resource profiling. Two common tools for resource profiling are resource levelling and resource loading graphs. Similarly the critical resource diagram and the resource idleness graph also express resource information.

IV. CODING IMPLEMENTATION

A. *HTML, CSS and C++*

The Resource Allocation system was designed as a website with the idea in mind that users will access the resource allocation information via a network connected to a database.

HyperText Markup Language (HTML) and a Cascading Style Sheet (CSS) were used to design the website. Inspiration for the CSS was from Mack (2011). The CSS influences the design of the table, graph, general layout and presentation. The code for the website can be found in the appendix. Other sources including from TextFixer.com (2013) and Stanley (2007) provided useful information and inspiration regarding the HTML component.

There are examples of coding in different languages considering the behavior of the resources.

B. *HTML and CSS code*

```
/*General*/
body
{
background: #ffffff;
color: #1e1e1e;
font: 12px/20px Arial, sans-serif;
margin: 0;
```

```
padding: 0;
}
h2
{
font-size: 18px;
font-weight: normal;
line-height: 20px;
margin: 0 0 20px 0;
```

C. HTML – Search Resource Page (Index)

```
<!doctype html>
<html lang="en">
<head>
<meta charset="utf-8">
```

```
<meta name="viewport" content="width=1024">
<title>Search Resource</title>
<link rel="stylesheet" href="css.css">
</head>
<body>
<div id="wrapper">
<p>
<h2>Search Resource</h2>
<table id="data-table" border="1" cellpadding="10" cellspacing="0"
summary="Search Results">
<caption>Search Result(s)</caption>
</table>
<table style="width: 600px;">
<tr>
<td height="23" style="text-align: left; width: 100px;"><a
href="Resource Information.html">Test Person</a></td>
<td style="text-align: left; width: 100px;">&nbsp;</td>
</tr>
</table>
<p>
<table id="data-table" border="1" cellpadding="10" cellspacing="0"
summary="New Search">
<caption>New Search</caption>
</table>
<form id="search box" form method="get" action="">
Name:
```

D. HTML – Information Page

```
<!doctype html>
<html lang="en">
<head>
<meta charset="utf-8">
<meta name="viewport" content="width=1024">
<title>Resource Information</title>
<link rel="stylesheet" href="css.css">
</head>
<body>
<div id="wrapper">
<p>
<h2>Resource Information</h2>
<table style="width: 600px;">
<tr>
<td height="23" style="text-align: left; width: 100px;">First Name:
Test</td>
```

E. HTML – Resource Allocation Page

```
<!doctype html>
<html lang="en">
<head>
<meta charset="utf-8">
<meta name="viewport" content="width=1024">
<title>Resource Allocation</title>
<link rel="stylesheet" href="css.css">
</head>
<body>
<div id="wrapper">
<table style="width: 400px;">
<tr>
<td colspan="5">
<div align="center"><h2>Project Allocation</h2></div>
</td>
</tr>
<tr>
<td style="width: 75px; vertical-align: bottom;">Project Name</td>
<td style="width: 200px; vertical-align: bottom;">
<table style="width: 405px;">
<tr>
<td height="23" style="text-align: left; width: 100px;">02/09/13</td>
<td style="text-align: left; width: 100px;">09/09/13</td>
<td style="text-align: left; width: 100px;">16/09/13</td>
<td style="text-align: left; width: 100px;">23/09/13</td>
</tr>
</table>
</td>
</tr>
```

F. C++ – Kohonen Maps

The following code has been slightly modified from Adam Stirtan's (2010) work in orderincorporate a new item in the dropdown menu. In order to initialise the images for the Kohonen Maps the following code was used to initialise the settings to create a new image list and directory:

```
newImages = new List<Bitmap>();
directory = new DirectoryInfo(Environment.CurrentDirectory + "\\..\\..\\" +
"New");
files = directory.GetFiles();
foreach (FileInfo file in files)
{
Bitmap b = new Bitmap(file.FullName);
Page | 51
newImages.Add(b);
}
```

Based on how the other methods were set out, a new method for the new images had to be

created using the same structure:

```
protected List<List<double>> CreateNewImageTrainingSet()
{
List<List<double>> trainingSet = new List<List<double>>();
txtStatus.Text += "Reading new images\r\n";
foreach (Bitmap b in newImages)
{
List<double> example = ImageReader.ReadImage(b);
trainingSet.Add(example);
txtStatus.Text += ".";
Application.DoEvents();
}
txtStatus.Text += " Done\r\n";
progressbar.Maximum = Settings.GetNumIterations();
return trainingSet;
}
```

In order to enable the newImages to appear and function as a selection in the dropdown it has

been given the following condition:

```
if (cboxImageTrainingSet.SelectedIndex == 2)
{
trainingSet = CreateNewImageTrainingSet();
images = newImages;
scale = 3;
}
```

Starting from zero, there is a total of three if statement options for this particular menu dropdown with newImages being the last option (option two). Hence for newImages to have been selected the third option must be selected. The images have been scaled by 3 in order to reduce the size of the image when displayed in the application (images are originally 100px by 100px).

V. FUTURE DEVELOPMENT

Future improvements for Kohonen Maps could include:

- Integration of the executable file into the web page as an embedded interactive file
- Utilised as a predictor of future resource demand by basing it on past trends
- Potential expansion or integration into a project management system
- Analyse the resource demand for projects based on a particular period
- Analyse the web traffic each page in the system receives.
- Enable adaptability to other systems. This idea can be improved to analyse the resource and how often a page in the system is viewed.

VI. CONCLUSIONS

The development of a web based resource allocation system is something that will be useful for projects in order to manage a resource's capacity. The utilisation of Kohonen Maps will be different and is something that could assist projects in the future.

Labour resources need to be effectively managed in order to be utilised efficiently minimise redundant costs in training design and development of a resource allocation system assists users in managing, controlling, and allocating resources.

Resource allocation is a difficult task, tool through the verification and validation of existing project management tools create a web based system utilising Kohonen Maps as a unique feature to enable a different way of viewing the resource demand.

Important measurements are applied to all projects by corporations that perform exceedingly well in project delivery. These measurements include return on investment, creating value and accomplishing business benefits.

The complexity of resource management is constrained by limitations such as resource interdependencies, limitations on resource availability and substitutions, resource priorities and mutual exclusivity of resources.

The constraints determine the techniques and tools that are used for resource management. If the human resource does not have the proper equipment and other resources, the work won't be completed.

REFERENCES

[1] Orr, G., Schraudolph, N. & Cummins, F. 2013, Kohonen's Self-Organizing Map (SOM), Willamette University, Oregon, viewed 5 May 2013, <http://www.willamette.edu/~gorr/classes/cs449/Unsupervised/SOM.html>.

[2] Rojas, R. 1996, Neural Networks, Freie Universität Berlin, Berlin, viewed 5 May 2013 <http://page.mi.fu-berlin.de/rojas/neural/chapter/K15.pdf>.

[3] Adam Stirtan 2010, Kohonen Network Self Organizing Map for Image Clustering, video recording, viewed 18 November 2013, <http://www.youtube.com/watch?v=2YTco1z_BGU>.

[4] Hewlett-Packard Development Company, L.P. 2010, HP Project and Portfolio Management Center, December 2010, cat. no. 4AA0-9340ENW, Hewlett-Packard Development Company, L.P., Palo Alto, viewed 17 March 2013, <http://www.ts.avnet.com/uk/vendors/hp/assets/hp_saas_project_portfolio_management_centre_brochure.pdf>.

[5] Badiru, A. B., Badiru, A. & Badiru, A. 2008, Industrial Project Management: Concepts, Tools, and Techniques, Taylor & Francis Group, Boca Raton.

[6] Mind Tools Ltd 2013, Pareto Analysis, viewed 16 June 2013, <http://www.mindtools.com/pages/article/newTED_01.htm>.

Design of Unit Testing using xUnit.net

Zenon Chaczko, Robin Braun, Lucia Carrion and Julian Dagher

Faculty of Engineering and Information Technology
University of technology, Sydney (UTS), Sydney, NSW, Australia
E-mail: {Zenon.chaczko, Robin Braun, Lucia.C.Carrion}@uts.edu.au, Julian.Dagher@alumni.uts.edu

***Abstract*—This paper presents an in-depth study of designing, implementing and executing unit test cases using the *xUnit.net* testing tool in general and in the context of the TeleMedicine Cluster System project within the ICT Design subject delivered at UTS, Australia. The case studies are based on the utilisation of the tool in Visual Basic 2012 using the .NET framework for C#. The paper elucidates on how and why the *xUnit* framework can be applied in the context of the TMC system, and how it can be tailored to meet the testing ad integration needs of the delivery of TMC system.**

Keywords—Unit Testing, Automated Testing, Software Development Process

I. Introduction

In development of software intensive system, the main goal of test automation is to help improve the efficiency of production and development of software. It is targeted at giving the developers engaged in software projects the tools and process to be more efficient, agile and precise. This is able to be achieved by providing the developer with instant feedback due to any changes or new code implemented. The benefits of this are that it reduces the stress felt by the developers, having this instant feedback, which allows them to focus more closely on their task at hand. For test code to be effective however, it is expected that there is about as much code used solely for testing as there is code used for the actual production and development of the software. The challenge in this scenario is now to provide that test code without inhibiting the development process and increasing the effort needed to maintain the software being developed.

A. The Need for Automation

Test automation needs to be implemented at many phases throughout the development process. This can start before any development code is written. These tests are written to test according to specifications, therefore when a test programmer is writing the development code he is given instant feedback on how the code meets the requirements, or breaks unexpectedly. After the code is written, test programmers are required to run tests as documentation, as well as, to discover any bugs and defects in the code. All of this can be automated as part of the testing process and if the tests are designed correctly, made fully automated, repeatable and robust, and the cost of running these tests throughout the whole development process can be minimised. As a result, it is possible to minimise the total cost of the development process itself, as one can gain the rewards of automated tests. Test code may be as numerous as production code, as production code, but it must also be maintained along with the production code. The aim however, is to make the test code easier to maintain. If this is done incorrectly it will cause more problems than benefits and be a source of delay, eventually becoming redundant. In other words, if test code is not easy to maintain, it will get left behind and lose all its value, eventually forcing the programmer to turn away from it and go to another approach such as manual testing. To avoid this it must be kept in mind that tests need to be written in a maintainable format. The following figures below show how automating tests can improve productivity and help to reduce effort, or if written in an un-maintainable style, lose all their value, forcing the test programmers to turn back to the original model of manual testing. Here the original effort placed into the development over time is demonstrated, while no extra efforts were added into automating test at any other stage of the development process. This approach requires consistent work throughout the whole development process.

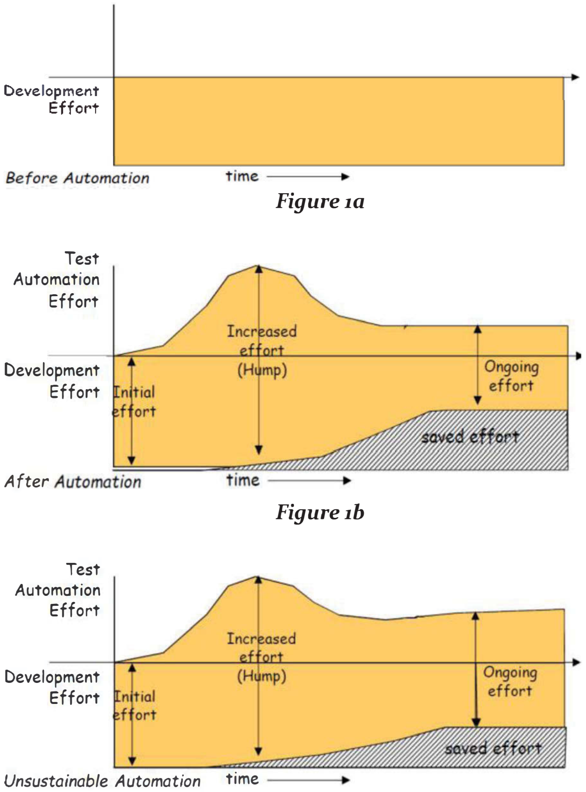


Figure 1a

Figure 1b

Figure 1c

Figure 1 – Development effort before (a) and after automation (b), Unmaintainable automation (c); adopted from Meszaros [2].

Figure 1b shows the effort needed to implement test automation. In this process it can be seen there is a large initial increase in effort the write and maintain test automation code. This at first seems very unappealing, but as demonstrated, if the unit tests are implemented correctly and in a maintainable fashion, the effort required to maintain the tests is very minimal. The effects of having these tests in place can be seen on the development side of project. It shows as the tests are developed and become automated, the development effort is greatly reduced as the automation of tests work their magic. This is because the automation instantly allows the developer to see the flaws in their code and makes the rest of the development process flow easier due to more peace of mind from the developer making the coding much efficient and effective. The benefits gained from test automation, however, might be lost, if the tests produced are not easy to maintain, and therefore unsustainable. Here the same initial increase in effort can when attempting to automate the testing process can be seen. However, this is not greatly reduced after the initial increase, as the tests made are not always easy to maintain, as a result, a doubling effect in the effort might be needed to maintain both the development and testing. The effort saved in the development is more than replicated in the maintenance of the tests, thus eventually causing the developer to turn away from automation and back to the original testing methods.

B. *Test Smells*

Test Smells are underlying problems in the code which arise due to the automation of testing. As soon as test developers begin to write their unit tests, some problems in the written code become to be noticeable. The symptoms underlying this problem are referred to as test smells. These are not necessarily the actual cause of the problem, but rather just a set of symptoms which may be defined by several causes. There are several different types of test smells [2] known as the following:

- *Code smells* – These are problems in test code which are visible in the actual code itself.
- *Behaviour smells* – These are problems caused by incorrectly written test code, which are not obvious until they result in tests performing unexpectedly or in an incorrect manner.
- *Project smells* – These are testing problems related to the entire project as a whole.

Code smells are the cause of behaviour smells, which are then the cause of project smells. Code smells can also be directly the cause of project smells. Basic types of code smells can be simple issues such as hard coding values into the tests. This can lead to fragile tests which are not robust as need or intended by the developer. An example is shown below [2]:

```
assertEquals(new BigDecimal("30"),
actualLineItem.getPercentDiscount())
```

Figure 2 – Code Smell Fragile Test

Another common smell could be testing each individual method of an object in a single test; which can lead to a verbose and difficult to read test (see Fig 3 below).

```
assertEquals(expectedLineItem.getInvoice(),
actualLineItem.getInvoice());
assertEquals(expectedLineItem.getProduct(),
actualLineItem.getProduct());
assertEquals(expectedLineItem.getQuantity(),
actualLineItem.getQuantity());
assertEquals(expectedLineItem.getPercentDiscou
nt(), actualLineItem.getPercentDiscount());
assertEquals(expectedLineItem.getUnitPrice(),
actualLineItem.getUnitPrice());
assertEquals(expectedLineItem.getExtendedPrice
(), actualLineItem.getExtendedPrice());
```

Figure 3 – Code Smell Verbose Test [2]

C. *Test Patterns*

A test pattern is referred to as a "recurring solution to a recurring problem" [2]. The problems arise from test automation and are called test smells as discussed above. Test patterns are simply solutions to problems which one may keep replicating due to the fact that the problem appears several times, and needs the same solution to solve the issue. There may be some problems which can be solved with a single pattern, while others may need more than just once pattern to solve.

There are three general categories of test patterns which are at different levels of abstraction. These levels [2] are defined as follows:

- Strategy level
- Test design level
- Test coding idioms level

In order to implement test patterns first the test code need to be written, starting with the simple tests first, then doing a review of the code and identify the test smells; test programmer is able to find. Once these are identified, then test patterns are used to solve these issues. As a result, rewriting the code in a more effective and maintainable manner. The test patterns can be applied to solve the above code smells. For the first code smell an expected line item is defined with the chosen variable value set to it. This allows for robust and repeatable coding, which then can include assertions defined as the variable values [2] as shown below:

```
LineItem expectedLineItem =
newLineItem(invoice, product, QUANTITY);
assertEquals(expectedLineItem.getPercentDis
count(),
actualLineItem.getPercentDiscount())
```

Figure 4 – Test Pattern Robust Test

For the second code smell the pattern which can be used to solve the issue is the use of expected objects rather than expected methods. In this a whole collection of assertEquals is replaced with a single assertion which includes the expected object only [2]:

```
assertLineItemsEqual(expectedLineItem,
actualLineItem)
```

Figure 5 – Test Pattern Expected Object

II. Case Study

A. Overview

The following case study describes design and development methodology of *xUnit.net* based unit tests for C# using Visual Basic (VB) 2012 and the .NET framework. The paper discusses the *xUnit* framework and its application to the TMC. It will explain why *xUnit* test are required for the TMC, and discuss and demonstrate how this framework will be applied and tailored specifically to the TMC. It will then provide users with a quick set up procedure of how to install all the related components and prepare test programmers to get started. It will then proceed to provide a framework for building unit test cases, and show how to execute these third party unit tests within the existing Visual Basic test explorer. Following on from this, several examples of relevant unit tests are demonstrated. These test examples utilise the *xUnit.*net testing tool and were developed to use as a guide for creating all unit tests during the development of the TMC system in ICTD [13] in Autumn 2013.This paper explains the need for the use of the *xUnit* framework on the TMC project, and how it was used to benefit the project over the course of the development and system integration.

B. Scope

This case study will assume the following:

- User has basic knowledge of VB 2012
- User has basic knowledge of C#
- Use has installed VB 2012
- User has installed the .NET framework

The case study will try to address the following issues:

- What is *xUnit* unit testing
- The need for *xUnit* in the TMC
- Downloading and installing NuGet Package Manager
- Downloading and installing *xUnit.*net runner
- Downloading and installing *xUnit.*net
- Creating a class library for the *xUnit.*net unit tests
- Creating a class which will comprise the unit tests for this tutorial
- Giving samples of unit test cases based on the TMC as developed by the Blue Team
- Executing unit tests within the VB test explorer

C. xUnit.net Framework

The *xUnit* facility is a collection of test automation frameworks, it is available in most languages and its end goal is to help developers automate their tests. It does this by attempting to make it easier for developers to write their tests using the same language they are developing in. This allows the developer to focus on the important tasks at hand rather than attempt to code tests in an unknown language. The aim is to make unit testing simpler, by allowing tests to be implements at a class or object level, without the need of any of the remaining code being written. Therefore as long as tests are designed correctly, it enables developers to start testing from the minute the coding phase gets started. The *xUnit* tool aims to improve the way tests are executed. This should be a simple process which allows the developer to run a single test, a collection of tests or all the tests with the single click of a button. This provides instantaneous feedback allowing the developer to instantly see where there is a break in the code. This enables the developers the reduce the costs involved with constant testing, encouraging them to run test more frequently, and as a result improving the overall quality and execution of the software. Unit testing is used to test code and make sure that it performs as expected. Unit tests are able to:

- Discover vulnerabilities in the code to see might break
- Highlight where changes to the code, even simple changes, may unexpectedly break the code
- Discover any design flaws during the code development
- Allow for a greater understanding of the functionality of the code

The *xUnit.*net framework is a third party testing tool which can be integrated into Visual Studio (VS) to provide all the above benefits and many more to help discover all the bugs imbedded in the code, helping to ensure more effective solutions. Some features available to *xUnit* include automation features such as AutoFixture (Evans 2013), this extension can be used to generate random variables at the beginning of each test, this enables the automation of the first phase of unit testing discussed below, the Arrange phase. This phase is used to define all the variables to be tested, and through this feature programmers are now able to automate that part of the testing. This makes for more efficient tests which are more flexible, independent and repeatable. The AutoFixture feature can also be very useful when developing unit tests in *boundary cases*. This can help the user define a range of arbitrary values for the inputs based on boundary cases in the code to help analyse at which points they may break the code [1]. By automatically generating the inputs from the other units and projects programmers are able to test just the unit under test at several different boundary cases with just one repeatable test. This allows the developer to analyse weaknesses in the code which may be incorrectly defined, and help them gain a clearer understanding of the code and how to properly define the necessary boundaries, and avoid any unplanned for or undesired breaks in the code.

As far as the boundary cases are concerned, there are also other helpful tools that can be used such as the PEX tool. This tool, which is an add-on to VS, can allow for automated white box testing [3]. This will automatically generate the input values into the unit, thus allowing programmers to test without having the actual inputs into the code. This allows once again for easier automation of the code when it comes to testing boundary cases. The *xUnit* functionality is also integrate-able into Visual Studio, thus allowing for the tests to be run repeatedly through the test explorer in Visual Studio [10]. The tests can be automatically run whenever required, at any stage of the development. This feature saves a lot of time and helps with continual troubleshooting and debugging of the code, and allows the developer to remain on top of any issues that may arise due to changes, even minor changes, which may unexpectedly break the code.

1) *Attributes*

Listed below (Table 1) are the attributes and their definitions specific to the *xUnit.net* framework [5, 6]. These attributes can be used to set or define certain parameters throughout the test code and create the tests to the exact specifications needed to achieve the desired testing scenario. Through these attributes one is able to test things such as whether or not the code throws and Exceptions, and even define which type of *exception* is expected the code to throw. This allows a thorough analysis of the code in order to ensure it executes as expected and breaks where expected.

Table 1 *xUnit* Attribute. Adapted from [5, 6]

***xUnit.net* Attributes**	**Comments**
[Fact]	Marks a test method.
Assert.Throws or Record.Exception	*xUnit.net* has done away with the ExpectedException attribute in favor of Assert.Throws. See Note 1.
Constructor	It is believed that use of [SetUp] is generally bad. However, one can implement a parameterless constructor as a direct replacement.
IDisposable.Dispose	There is a consensus that the use of [TearDown] is generally bad. However, one can implementIDisposable.Dispose as a direct replacement.
IUseFixture<T>	To get per-fixture implement setup, IUseFixture<T> on the test class.
IUseFixture<T>	To get per-fixture teardown, implement IUseFixture<T> on the test class.
[Fact(Skip="reason")]	Set the Skip parameter on the [Fact] attribute to temporarily skip a test.
[Fact(Timeout=n)]	Set the Timeout parameter on the [Fact] attribute to cause a test to fail if it takes too long to run. Note that the timeout value for *xUnit.net* is in ms
[Trait]	Set arbitrary metadata on a test
[Theory],[XxxData]	Theory (data-driven test).

2) *Assertions*

In the code assertions can be made at the end of the code to ensure the desired test scenario is met. For example if the test is to ensure that a certain double value generated by calling a certain method is the same as the expected double value, one would define the expected value and then Assert.Equal() using the correct parameters to ensure that the right output is generated. These assertions are specific to the *xUnit* framework and used as the final stage of a unit test method. The methods of creating a unit test stages [8, 9, 11] are discussed in the tutorial section of the document. Through the assertions, test developers are also able to test reactions to invalid inputs and how the code behaves or responds in those scenarios.

Table 2 *xUnit* Assertions. Adapted from [5, 6]

***xUnit.net* Assertions**	**Comments**
Equal	MSTest and *xUnit.net* support generic versions of this method
NotEqual	MSTest and *xUnit.net* support generic versions of this method
NotSame	Ensures two values are not the same
Same	Ensures two values are the same
Contains	Ensures a certain value is contained in the code
DoesNotContain	Ensures a certain value is not included in the code
DoesNotThrow	Ensures that the code does not throw any exceptions
InRange	Ensures that a value is in a given inclusive range (note: NUnit and MSTest have limited support for InRange on their AreEqual methods)
IsAssignableFrom	Ensures a value is assignable from a part of the code
Empty	Ensure an empty value is returned
FALSE	Ensures a certain Boolean returns false
IsType	Ensures code return is a certain type
NotEmpty	Ensures a non-empty value is returned
IsNotType	Ensures code return is not a certain type
NotNull	Ensures a Null is not returned
Null	Ensures Null is returned
TRUE	Ensures a certain Boolean returns true
NotInRange	Ensures that a value is not in a given inclusive range
Throws	Ensures that the code throws an exact exception

III. UNIT TESTING USING *XUNIT.NET* IN THE TMC

In the TMC system development project, during its implementation and test phases the *xUnit* framework was used for unit testing. The developers and testers were able to continually debug and update the test code in order to ensure it is not vulnerable to any unexpected changes in the source code which may cause it to break. This is seen to be very beneficial to the quality and efficiency of the of the code development as it would allow for continual automated testing through the test explorer at any stage of the development. Also, it was expected, the *xUnit* framework would allow for the code developers to have instant debugging with any changes they make to the code, ensuring that it does not break, and being able to debug when it actually does.

There are some drawbacks to this approach, as it can be very time consuming and requires a lot of effort which could have been solely focused into the development of the code. On the other hand though, the effort spent developing the unit tests can be very beneficial throughout the development, as identifying issues would become simpler and could save time throughout the process.

A. *Background to the TMC*

What is the significance of unit testing? In general, the developed Tele-Medicine Cluster (TMC) system is a solution to automate and simplify the ordering of medicine in medical institutions. It consists of several modules which define the overall system and make up the final product. Unit testing involves the testing of these modules throughout the development of the TMC. This will allow for the TMC developers to progressively validate and ensure the functionality each individual module. This procedure is very important in the TMC as every module is a key aspect to the overall operation of the system, and to be able to integrate this solution, one must be able to ensure each module first functions as desired.

The TMC is designed to be a scalable solution where one is able to continually add functional units to the supervisor and allow the functionality to continue as normal. For this to be achievable each unit must be correctly developed and coded to allow for seamless integration with other units. This is where *xUnit* unit tests come in to allow for continual monitoring throughout the development process, ensuring the critical functions of each unit are able to perform as specified. In order to tailor the functions of the *xUnit* to the TMC, there is a need to incorporate an additional software, called the *xUnit* runner, for Visual Studio. This add on will allow for easy, and repeatable automation and running of the design unit tests whenever deemed necessary to assist with the continual monitoring, and allow the Blue team to save its limited resources for the development of the TMC itself. Through this process, and by correctly implementing the *xUnit* framework, developers are then able to save time in other areas of the development by this automation and ease of debugging.

1) *Advantages*

Advantages of implementing unit test using *xUnit* for the TMC are as follows:

- Automated testing through the test explorer
- Automated variable generation through AutoFixture
- Instant debugging
- Identifying issues due to changes
- Testing code reliability (if and where it breaks)
- Saves time down the track after tests are written

2) *Disadvantages*

Disadvantages of implementing unit test using *xUnit* for the TMC are as follows:

- Time consuming
- Limited resources in the Blue team would become even less
- Time could be spent developing code
- Incorrectly coding the tests could lead to misleading results

B. *Setting up xUnit*

1) *Scope*

This section presents as a procedure to simplify the structure and act as a quick start set by step guide in setting up the system to be ready to start writing and executing test cases. The paper will not show any samples of unit tests, rather just the required format the tests need to be in and how they are to be referenced in Visual Basic to represent *xUnit* test methods. Actual samples relating to the TMC will be discussed in the following section of the document.

The below listed quick set-up steps covers the activities needed to get started using the *xUnit* testing tool. It will just cover the basic software which needs to be added on to Visual Studio in order to get started, as well show how to set up a class in Visual Studio which will be used to hold the unit test created. It will also cover a basic outline and format which is the recommended format the test methods will be created in. Then finally this guide will show how to build and run the unit tests created through Visual Studio's in built test explorer.

2) *Process Steps*

a) Step 1

The first step is downloading the *xUnit.*net package. The testing tool can be downloaded directly from the following link http://xunit.codeplex.com/downloads/get/423827, then the extract has to be downloaded into the root of the selected project directory.

b) Step 2

The next step is to download the NuGet Package Manager which is just a set of "tools to automate the process of downloading, installing, upgrading, configuring, and removing packages from a VS Project". This can be downloaded from the following link by clicking the download button: http://visualstudiogallery.msdn.microsoft.com/27077b70-9dad-4c64-adcf-c7cf6bc9970c.

Once downloaded, one needs to execute the file and follow the prompts to install it. Visual Basic will need to be restarted for this to take effect.

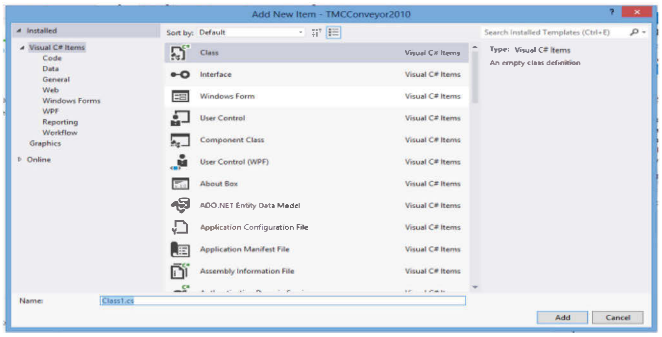
c) Step 3

Once Visual Basic is restarted, users would need to install *xUnit.*net runner for Visual Studio 2012 {VS 2013) . This tool allows running *xUnit* unit tests from inside the Visual Basic test explorer. It can be found using the following link:http://visualstudiogallery.msdn.microsoft.com/463c5987-f82b-46c8-a97e-b1cde42b9099.

Similarly, one must click the download button, execute once downloaded, and follow the prompts. Once again users must restart Visual Basic after this process is completed.

d) Step 4

The next step in this process is to create a class for the *xUnit.*net tests. To do this one must click on the class library holding the code that is to be tested right click and add Class. A class can name as required. In this tutorial the tests will be based on the TMCConveyor so the class will be named TMCConveyorTests for reference.



Add New Item - TMCConveyor2010

Sort by: Default

Type: Visual C# Items

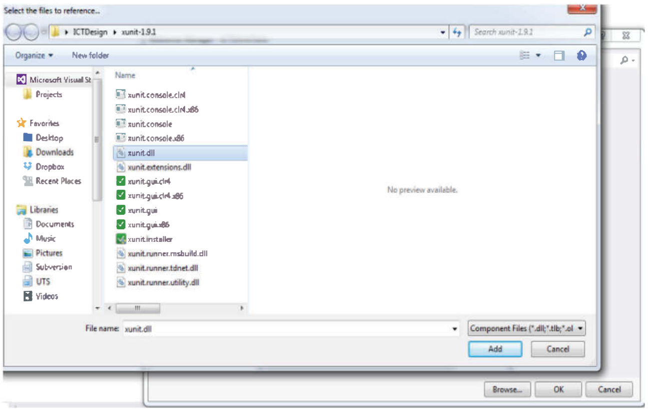
An empty class definition

Name: Class1.cs

Figure 6 – Add Class

e) Step 5

Once this is completed programmers must add a reference from that class library, TMCConveyor, to *xunit.dll* (Fig. 7). This can be achieved by right clicking the library>>Add reference>>Browse. This file will be located in the *xUnit*.net package which was downloaded in the first step.



Select the files to reference...

File name: xunit.dll

Component Files (*.dll;*.tlb;*.ol

Figure 7 – Add *xUnit.dll* reference

f) Step 6

One must now edit the class holding the tests for this tutorial. To set up the class to use *xUnit* test programmers must refer to *using Xunit;* The following format will be used for the tests which will be run using this tutorial (Wilson 2013).

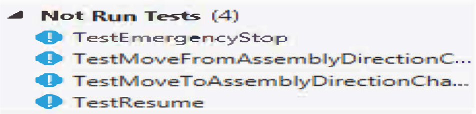
```
namespace TMCConveyor
{
    public class TMCConveyorTests
    {
        [Fact]
        public void EnterTestMethodNameHere()
        {
            Enter test data here; // Arrange
            //Act
            Call the required method to implement
what one would like
            to test;
            //Assert  the  required  assertion  is
met.
            Assert.EnterAssertionFromAboveHere
        }
    }
}
```

Figure 8 – *xUnit.net* unit test format

For each new unit test method created, a new name will be assigned and new steps relevant to the class being tested will be added. As discussed above, the [*Fact*] attribute defines it as a new test method, allowing it to be picked up by the test explorer to be run as a test. After this, a new test method has to be declared, named according to the test which is being performed. In this method, the steps necessary to complete the test are entered. The above format of *Arrange*, *Act* and *Assert* is the recommended format to structure each test method. *Arrange* is just to define the variables and create instances of code for testing. *Act* is acting upon the code selected for testing by calling the relevant method [4]. At the end of each test method there is an *Assert*. These assertions are as discussed above and are called using the Assert method, followed by the type of assertion one would like to make. This is then completed by entering the variables programmers would like to make the assertion based on, based on what is acceptable by the type of assertion being made.

g) Step 7

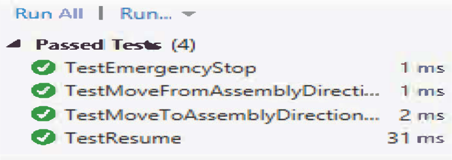
The test programmers then build this solution to ensure that there are no errors. Due to the installed runner in step 3, these tests will now show up in the test explorer as shown (Fig. 9) below.



```
◢ Not Run Tests (4)
   TestEmergencyStop
   TestMoveFromAssemblyDirectionC...
   TestMoveToAssemblyDirectionCha...
   TestResume
```

Figure 9 – Test Explorer

From the test explorer these tests can be run one by one or all at once using the run all button. If the tests are successful, it will result in the following output.



```
◢ Passed Tests (4)
   TestEmergencyStop                    1 ms
   TestMoveFromAssemblyDirecti...       1 ms
   TestMoveToAssemblyDirection...       2 ms
   TestResume                          31 ms
```

Figure 10 – Successful Tests

h) Step 8

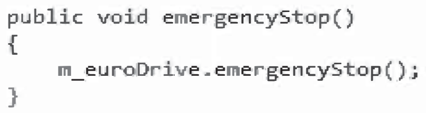
Once the test code is written, once again one needs to build the solution to ensure that there are no errors. When this is confirmed, programmers need to execute all the tests using the run all method discussed in Step 7. This allows us to see if there are any errors in the code and then change the code as needed to ensure it is operational.

C. TMC xUnit Test Cases

Now, when all the basics are out of the way it is time to select a few classes which will be run unit tests on. At first, the code for test must be selected, and then it needs to be analysed it to see what the expected output is. After this task is completed one can write some code to test the functionality to see if it performs as planned, and then finally execute the test and make adjustments as necessary to fix the code.

1) Case 1: Emergency Stop

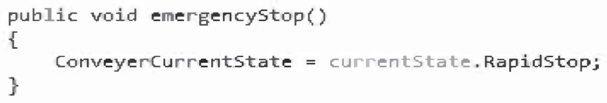
The first test that is run should be a simple test to ensure the *emergencyStop* function of the *TMCConveyor* is functioning correctly. The reference to the code will be tesed in the *FullConveyor.cs* class can be found below:



```
public void emergencyStop()
{
    m_euroDrive.emergencyStop();
}
```

Figure 1 – *emergencyStop* code

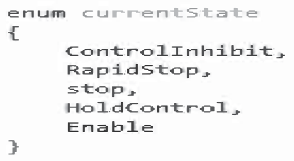
Then the *emergencyStop* procedure is referred to in the *RS485Controller* class file (see Fig 12).



```
public void emergencyStop()
{
    ConveyerCurrentState = currentState.RapidStop;
}
```

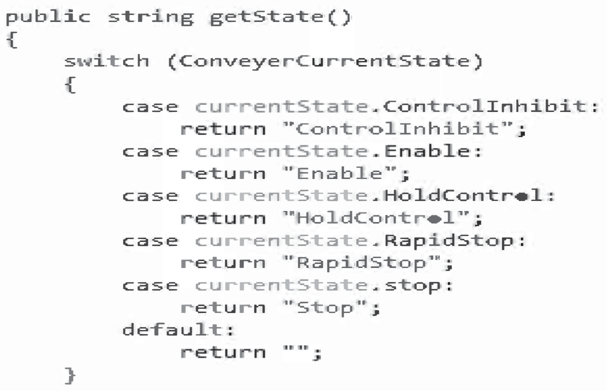
Figure 2 – *RapidStop* code

This also leads us to the following code relating to the *currentState* under the class.



```
enum currentState
{
     ControlInhibit,
     RapidStop,
     stop,
     HoldControl,
     Enable
}
```

Figure 3 – currentState code



```
public string getState()
{
    switch (ConveyerCurrentState)
    {
        case currentState.ControlInhibit:
            return "ControlInhibit";
        case currentState.Enable:
            return "Enable";
        case currentState.HoldControl:
            return "HoldControl";
        case currentState.RapidStop:
            return "RapidStop";
        case currentState.stop:
            return "Stop";
        default:
            return "";
    }
```

Figure 4 – getState code

As can be seen from the above code, calling the emergencyStop sets the currentState of the conveyor into the RapidStop state. Programmers then can be able to get this state using the getState method which converts the code to strings. In this scenario, one can set an expected state which is expected the conveyor to be in, call the command, and then by using the Assert method used by *xUnit*, one can compare, if the state is as expected. One must first ensure though that the conveyor was not already in this state. This leads to the following code:

```
[Fact]
//declares method as an xunit test method
public void TestEmergencyStop()
{
  RS485Controller m_euroDrive;
  m_euroDrive = new RS485Controller();
  //create a new instance of RS485Controller
  string RapidStop = "RapidStop";
  // define a string with the expected value
  // of currentState after calling emergencyStop
 Assert.False(m_euroDrive.getState().Equals(Rapi
 dStop));
  // Test if the conveyor is not in emergencyStop
  // state
  m_euroDrive.emergencyStop();

 //Call the emergencyStop method
 Assert.True(m_euroDrive.getState(.Equals(RapidS
 top));
 //Test  to  ensure  that  the  state  correctly
changed
 // to the emergencyStop state.
}
```

Figure 5 – *TestEmergencyStop* code sample

2) Case 2: Resume from Emergency Stop

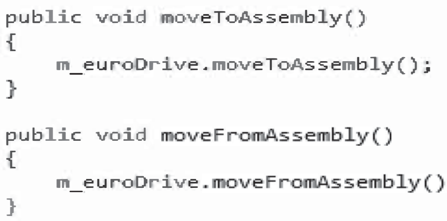
Using a similar method to the first test, it is possible to make a test in order to ensure that the conveyor is able to resume after being in an emergency stop state, the code for this is as shown below. Here, the conveyor is put in the emergency stop state and then test to ensure it is in fact not enabled. Then, the operation can be resumed and test executed to see, if the operation resumes correctly.

```
[Fact]
public void TestResume()
{
  RS485Controller m_euroDrive;
  m_euroDrive = new RS485Controller();
  // create a new instance of RS485Controller
  string Enable = "Enable";
  //define a string with the expected value of
  // the currentState after the Resume is called
  m_euroDrive.emergencyStop();
  // Put the conveyor into emergencyStop state

 Assert.False(m_euroDrive.getState().Equals(Enab
 le));
  //Test to check the conveyor is not enabled
  m_euroDrive.startDrive();
  // Resume the operation of the conveyor

 Assert.True(m_euroDrive.getState().Equals(Enabl
 e));
  // Test to ensure the conveyor correctly
resumed
  // and changed state to enabled
}
```

Figure 6 – TestResume

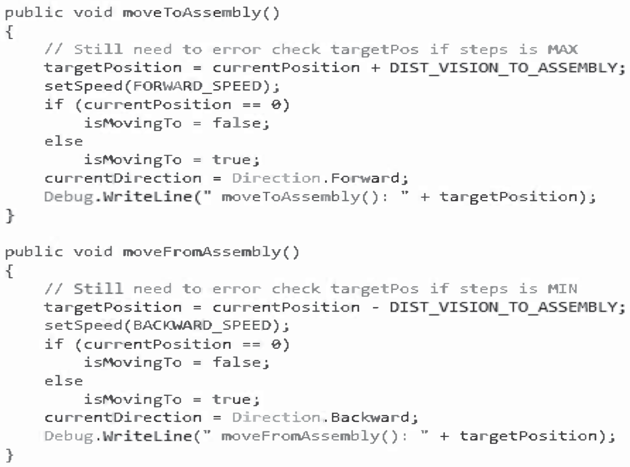
D. Case 3/4: Change Direction

This case will involve running two tests to confirm the full functionality of the requirement. Once again, one needs to look through the classes and find the following sets code relating to the direction of the conveyor and to where it is moving.



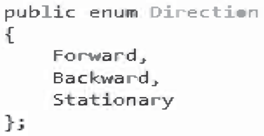
```
public void moveToAssembly()
{
    m_euroDrive.moveToAssembly();
}

public void moveFromAssembly()
{
    m_euroDrive.moveFromAssembly();
}
```

Figure 7 – Move To and From Assembly Methods



```
public void moveToAssembly()
{
    // Still need to error check targetPos if steps is MAX
    targetPosition = currentPosition + DIST_VISION_TO_ASSEMBLY;
    setSpeed(FORWARD_SPEED);
    if (currentPosition == 0)
        isMovingTo = false;
    else
        isMovingTo = true;
    currentDirection = Direction.Forward;
    Debug.WriteLine(" moveToAssembly(): " + targetPosition);
}

public void moveFromAssembly()
{
    // Still need to error check targetPos if steps is MIN
    targetPosition = currentPosition - DIST_VISION_TO_ASSEMBLY;
    setSpeed(BACKWARD_SPEED);
    if (currentPosition == 0)
        isMovingTo = false;
    else
        isMovingTo = true;
    currentDirection = Direction.Backward;
    Debug.WriteLine(" moveFromAssembly(): " + targetPosition);
}
```

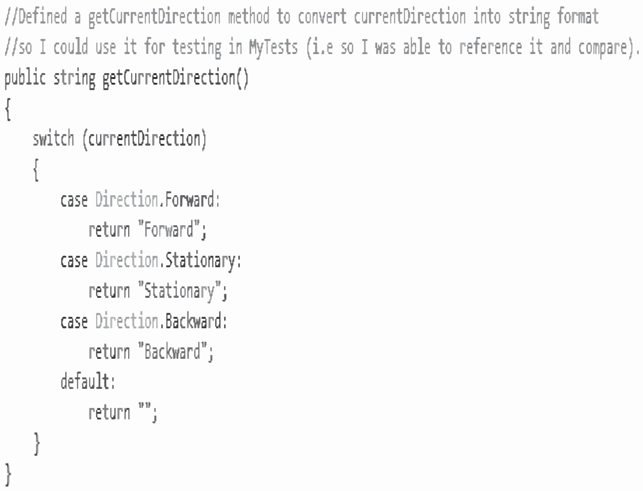
Figure 8 – Move To/From Methods in RS485Controller



```
public enum Direction
{
    Forward,
    Backward,
    Stationary
};
```

Figure 9 – Direction enum

In the existing code, there was no get method to convert the private value currentDirection into an exportable string. Such a get method can be added to the RS485Controller code (Fig. 20) to facilitate the string export.



```
//Defined a getCurrentDirection method to convert currentDirection into string format
//so I could use it for testing in MyTests (i.e so I was able to reference it and compare).
public string getCurrentDirection()
{
    switch (currentDirection)
    {
        case Direction.Forward:
            return "Forward";
        case Direction.Stationary:
            return "Stationary";
        case Direction.Backward:
            return "Backward";
        default:
            return "";
    }
}
```

Figure 20 – Get currentDirection code

Using the following sets of code one is able to design a test to check whether the direction of the conveyor changes as defined in the code, when the move to and from assembly methods are called. Samples of the code developed are shown below.

```
[Fact]
public void TestMoveToAssemblyDirectionChange() {
   RS485Controller m_euroDrive;
   m_euroDrive = new RS485Controller();
   //create a new instance of RS485Controller
   string expectedDirection = "Forward";
   //create a string containing an expected
   direction
   m_euroDrive.moveFromAssembly();
// call the method moveFromAssembly which sets
// the conveyor in the Backward direction

Assert.False(m_euroDrive.getCurrentDirection().Eq
uals(expectedDirection));

//Test to ensure that the current direction does
not match
// the expected forward direction
 m_euroDrive.moveToAssembly();
//call the method moveToAssembly to set the
conveyor is the
//expected forward direction

Assert.True(m_euroDrive.getCurrentDirection().Equ
als(expectedDirection));
// test to ensure the current direction equals
the expected direction
}
[Fact]
public void TestMoveFromAssemblyDirectionChange()
{
   RS485Controller m_euroDrive;
   m_euroDrive = new RS485Controller();
   //create a new instance of RS485Controller
   string expectedDirection = "Backward";
   //create a string containing the expected
   direction
   m_euroDrive.moveToAssembly();
   //call the method moveToAssembly which sets
     the conveyor in the Forward direction

   Assert.False(m_euroDrive.getCurrentDirection()
   .Equals(expectedDirection));
   //Test to ensure that the current direction
     does not match the
   //expected backward direction
   m_euroDrive.moveFromAssembly()
   //call the method moveFromAssembly to set the
     conveyor is the expected backward direction

   Assert.True(m_euroDrive.getCurrentDirection().
   Equals(expectedDirection));
   //test to ensure the current direction equals
     the expected direction
  }
}
```

Figure 10 – Change direction test code

IV. Conclusion

It is apparent that there was a need for unit testing to be implemented throughout the development of the TMC. There were several reasons for this, and the main reasons being:

- Continual debugging of the TMC throughout the development process.
- Automated testing through the test explorer
- Automated variable generation
- Identifying issues due to changes
- Testing code reliability (if and where it breaks)
- Saves time down the track after tests are written

There are several important notes to remember when attempting to implement these unit tests. This mainly refers to the structure of the test methods. The general structure includes such steps as: *Arrange, Act, Assert.* The *Arrange* step can be automated, if designed correctly, but it is, in simple terms, the arranging of the variables needed for the test to be performed. *Act*, is where one calls the method under test to put the code in action. *Assert* is the key element where programmers ensure that the code was achieved the desired result based on the inputs given to it.

Some pitfalls to avoid while implementing unit test are to ensure that the code is well understood, and that one is able to implement the correct procedures to test the code, otherwise this may lead to test results which report false positives, and thus misleading testers to believe the code is functioning correctly. Other pitfalls one may want to avoid include spending too much time on developing the unit test cases, taking away the time from developers by implement the unit tests right the first time, and therefore be able to continually run them in an automated fashion throughout the remainder of the development process. Therefore, if implemented correctly early on, the hard effort put it at this stage will make it easier through the remainder of the project.

Unit testing using the *xUnit* framework is a very effective way of developing and automating unit tests throughout the development of the TMC project. It enables developers and testers to gain a greater understanding of their code while developing a test method(s), which stretches code boundaries and thus ensures the code to behave as desired. This work is a good lesson to take in, especially for inexperienced developers, as inheriting these habits now will lead to improving their ability to code and debug issue that may arise.

Test automation is a very important task through the whole software development process. In particular, it is important to developers, as it helps reducing costs of software development throughout the entire software development cycle. If tests correctly automated, it was demonstrated here how test automation enables the reduction of effort required throughout the development process. Test automation is also important in increasing the efficiency and the effectiveness of development and thus contributing to improvement in the quality of the final product. The *xUnit* testing framework enables test developers to use an integrate-able platform which allows for automation of their code tests in an efficient and effective manner. Test automation, however, may lead to several problems which are here referred to as *test smells* which are due to errors in the test code, which then may eventually branch out and cause problems, such as unexpected behaviour in test code. A remedy to this particular problem is to apply *test patterns*. These are a recurring solution to a recurring *test smell* problem, which arise due to automation. Solving these problems increases the quality and effectiveness of the test code and as a result the implementation of test patterns, through refactoring code, allows the test automation to become easily maintainable. Consequently, this leads to a reduction in effort spent maintaining the test code, which could greatly reduce the effort spent in developing code.

V. Acknowledgment

We would like to acknowledge work and help received from 48481 ICTD students of the Blue Team in A2013.

VI. References

1. Evans, B. 2013, *AutoFixture,* Microsoft Corporation, viewed June 27th 2013, <http://autofixture.codeplex.com/>.
2. Meszaros, G. 2007, *xUnit Test Patterns: Refactoring Test Code,* Addison Wesley Professional.
3. *Pex and Moles - Isolation and White box Unit Testing for .NET,* 2013, Microsoft Corporation, viewed June 27th 2013, <http://research.microsoft.com/en-us/projects/pex/>.
4. Wills, A. & Hilliker, H. 2012, '2: Unit Testing: Testing the Inside', in R. Corbisier & N. Michell (eds), *Testing for Continuous Delivery with Visual Studio 2012,* Microsoft Corporation.
5. Wilson, B. 2013, *Comparisons,* Microsoft Corporation, viewed June 25th 2013, <http://xunit.codeplex.com/wikipage?title=Comparisons&referringTitle=Home>.
6. Wilson, B. 2013, *How do I use xUnit.net?,* Microsoft Corporation, viewed June 25th 2013, <http://xunit.codeplex.com/wikipage?title=HowToUse&referringTitle=Home
7. About xUnit.net, 2013, Microsoft, viewed June 25th 2013, <http://xunit.codeplex.com/>.
8. *Pragmatic Unit Testing: Summary,* 2004, The Pragmatic Programmers, viewed June 27th 2013, <http://media.pragprog.com/titles/utj/StandaloneSummary.pdf>.
9. Unit Testing, 2013, Joomla, viewed June 27th 2013, <http://docs.joomla.org/Unit_Testing>.
10. *Visual Studio Gallery,* 2013, Microsoft Corporation, viewed June 25th 2013, <http://visualstudiogallery.msdn.microsoft.com/>.
11. Unit Testing, 2013, Joomla, viewed June 27th 2013, <http://docs.joomla.org/Unit_Testing>.
12. *Visual Studio Gallery,* 2013, Microsoft Corporation, viewed June 25th 2013, <http://visualstudiogallery.msdn.microsoft.com/>.
13. 48481 ICTD, http://handbook.uts.edu.au/subjects/48481.html

Objective Quality Metrics in Correlation with Subjective Quality Metrics for Steganography

Raniyah Wazirali*, Shaher Slehat*, Zenon Chaczko* , Grzegorz Borowik† and Lucía Carrión *
*Faculty of Engineering and Information Technology
University of Technology, Sydney(UTS)
Australia, NSW, Sydney
Email: raniyah.a.wazirali@student.uts.edu.au and Zenon.chaczko@uts.edu.au
†Warsaw University of Technology
Nowowiejska 15/19
Email: g.borowik@tele.pw.edu.pl

***Abstract*—The main goal of hiding data is to conceal the very existence of the hidden information, therefore there is a significant demand for steganographic approaches that can ensure imperceptibility of such infromation. However, there is a limited corresponding eval- uation parameters available. Most of the studies use the Peak Signal to Noise Ratio (PSNR) as a metric for imperceptibility evaluation, although it could provide less accurate results than the Human Visual System (HVS) evaluation. This paper provides a review of the existent evaluation metrics that are used to assess the quality of steganography. The examination of the correlation between the existing objective and subjective metrics is also conducted. Pixel differences metrics have a poor correlation with the subjective metrics, hence the HSV based metrics have better correlation than pixel metrics.**

I. Introduction

Hiding data has become a key issue in the information security field. Steganography is a science of hiding information within another information and basically work as embed secret file in appropriate cover file without arising suspicion of attackers and produce stegoed file which is the file that have the hidden information. This resulted stegoed file must be identical to the cover file. The success of any steganographic system depends on the following three main factors which are: undetectability, imperceptibility and payload capacity.The hidden information must not arise any suspicion of an attacker as hiding a secret message in the host image may introduce some noise to the carrier; however, the introduced noise must remain invisible and can not be detected by any statical means or human visual system. Therefore, the imperceptibility of the hidden information is an important aspect to develop any stenographic approach [9].

Measuring the imperceptibility of the stegoed file is essential for most approaches dealing with image or video steganography. The most accurate and reliable way to determine the visual quality of such stegoed file would be by human visual evaluation (subjective evaluation). However, this type of evaluation is, time consuming, expensive, and can not be part of an automatic system. For these reasons, in recent year, many objective evaluation metrics were developed that can work in a similar way as the basic process of human perception.

The simplest and most extensively utilized quality assessment parameter is the Mean Squared Error (MSE), calculated by averaging the squared concentration alterations of cover and stegoed image pixels, beside it's relation to amount of Peak Signal-to-Noise Ratio (PSNR). MSE and PSNR are attractive parameters since they are modest to compute, have pure physical connotations, and are statistically suitable in the perspective of optimization. However, PSNR measure the mathematical differences between the cover image and the stego image and does not take into account the characteristic of human visual system (HVS). Therefore, they have

978-1-4799-7588-4/15 $31.00 © 2015 IEEE
DOI 10.1109/APCASE.2015.49

poor correlation with the perceived quality by the human visual system (HVS) Wang et al. [6, 7]. In the last few decades, an excessive work has gone into developing advance quality assessment methods that take power of the features of (HVS).

This paper will review the existing image quality metrics that can be used to evaluate the imperceptibility of steganography in section 2 represents the objective evaluation metrics in two main categories which are pixel differences measurement based evaluation and human visual system based evaluation. Section 3 reviews the subjective evaluation approach. Section 4 analyze the correlation between subjective and objective evaluation. Section 5 provides the processing time for some objective metrics.

II. OBJECTIVE EVALUATION

Image quality assessments can be classified as objective and subjective evaluation. Objective evaluation evaluates the image quality based mathematical algorithms. On the other hand, subjective evaluation assess the image quality based on the human visual capability and characteristic.

Objective evaluation metrics based on mathematical process to measure the imperceptibility between the original file and the stegoed file. It can be classified as Pixel Differences Measurements and Human Visual Based Measurements.

A. Pixel Difference Measurement

1) Mean Square Error (MSE): The distortion of an image is calculated by averaging the squared intensity of the cover image (carrier) and the resultant stegoed image pixels as in 1.

$$MSE = \frac{1}{m \times n}\sum_{i-1}^{m}\sum_{j-}^{n}(C_{ij} - S_{ij})^2 \tag{1}$$

Where (m, n) donates the amount of error differences between the original and the stego images. C is the original image and S is the stegoed image.

2) Peak Signal-to-Noise Ratio (PSNR): PSNR computes the ratio of peak to signal to noise for the cover image with the stegoed image as given in 2 where MAX is the value of maximum number of pixel of an image. . The cover and stego images must be the same size and type. The higher PSNR value indicate higher quality and match between the two input images.

$$PSNR = 10\log_{10}\frac{(I_{max})^2}{MSE} \tag{2}$$

3) Weighted Peak Signal to Noise Ratio ($wPSNR$) : $wPSNR$ is a quality metric developed by Voloshynovskiy et al. which mainly based on the value of Noise Visibility Function (NVF) as a correction factor that classified the image property based on textured and edge region as embedding data in this region can hide more data effectively with less distortion Voloshynovskiy et al. [4]. For smooth region, NVF is almost1 while for edge regions is close to 0 which indicates that for flat region $wPSNR$ is almost equal to $PSNR$ but for textured image, $wPSNR$ is higher than $PSNR$.

$$wPSNR = 10\log_{10}\frac{MAX^2}{||(C_{ij} - S_{ij})||^2_{NVF}} \tag{3}$$

$$wPSNR = 10\log_{10}\frac{MAX^2}{||(C_{ij} - S_{ij}).NVF||^2}$$

$$NVF(i,j) = \frac{1}{1 + \theta\sigma_x^2(i,j)} \tag{4}$$

Where $\sigma_x^2(i,j)$ represents the local alteration of the image in a window sprinted on the pixel with coordinates (i,j) and θ is a modification parameter consistent to the specific image which can be calculated as given in 5 and 6.

$$\overline{x}(i,j) = \frac{1}{(2L+1)^2}\sum_{k=-L}^{L}\sum_{l=-L}^{L}x(i+k, j+L) \tag{5}$$

$$\sigma_x^2(i,j) = \frac{1}{(2L+1)^2}\sum_{k=-L}^{L}\sum_{l=-L}^{L}(x(i+k, j+L))^2 \tag{6}$$

The image based on modification parameter is given as $\theta = \frac{D}{\sigma^2_{x(max)}}$ Where σ_x^2 is the maximum local variance for a particular image and $D \in [50, 100]$ is a range of determined value.

B. Human Visual Based Measurement

HVS is another important factor to measure the imperceptibility between the cover image and the stegoed image. It uses human perspective as a reference of the evaluation. As humans are

The main idea is that humans are involved in various characteristics of specific image other than taking it as a whole that include orientation, contrast, brightness, texture, ...etc. Although HVS measurement is very difficult to be assumed with psychophysical means, it is the instrument for anthropoid creature to recognize the contiguous world and the tool that detects brain secrets.

Cover image and stegoed image must be transformed into frequency domain using Discrete Wavelet Transform (DWT) or Discrete Fourier Transform (DFT). After that, Contrast Sensitivity Function (CSF) is utilized to both cover and the stegoed images. The CSF has a band-pass for distinguishing which associates with low human eye measure an image in the frequency domain. CSF can be calculated as given in 7.

$$H(w) = \begin{cases} 0.05e^{w^{0.554}} & w < 7 \\ e^{-9[|\log_{10} w - \log_{10} 9|]^{2.3}} & w \geqslant 7 \end{cases} \tag{7}$$

Where $w = (u^2 + v^2)^{1/2}$ and u and v are the spatial frequencies. After CSF filter is utilized, various techniques could be used to compute image quality include the metrics that given in section II-A. Moreover, various HVS based models have been proposed to compute image quality. In fact, HVS based quality metrics are the most close metrics to the subjective measurements. Below are two of HVS based metrics that have been used in measuring the imperceptibility of steganographic techniques.

$a = th$

1) Universal Image Quality Index (UQI) : UQI is a Universal image quality index for measure the distortion between two input images Wang and Bovik [5]. This approach based on configuring three factors which is luminance distortion, contrast distortion and structural comparisons. Despite this index is mathematically defined without considering the HVS, experimental results show that it reveals amazing reliability with subjective quality measurement. UQI performs better quality evaluation comparing with MSO and PSNR. Based on the above three comparasion, UQI can be described as given in 8 and 9.

$$UQI(c, s) = L(c, s), C(c, s), S(c, s) \tag{8}$$

$$UIQI = \frac{4\mu_c\mu_s\mu_{cs}}{(\mu_c^2 + \mu_s^2)(\sigma_c^2 + \sigma_s^2)} \tag{9}$$

where

$L(c, s) = \frac{2\mu_c\mu_s}{\mu_c^2+\mu_s^2}$ (Luminance Distortion)

$C(c, s) = \frac{2\sigma_c\sigma_s}{\sigma_c^2+\sigma_s^2}$ (Constrat Distortion)

$S(c, s) = \frac{2\sigma_{cs}}{\sigma_c+\sigma_s}$(Structural Comparisons)

Where $\mu_{c,}\mu_s$indicates the mean values of cover and stegoed images. And $\sigma_c\sigma_s$indicates the standard deviation of cover and stegoed images, and σ_{cs} is the covariance of both images.

UQI is a straightforward parameter that counts only on first and second order statistic of the original and stego images. UQI is considered unstable measure and doesn't correlate will with subjective assessment that is why Wang et. al proposed structural similarity index metric.

2) Structural Similarity Index (SSIM): In 2004, Wang and his team developed SSIM as an improvement of UQI Wang et al. [8]. SSIM measures the image quality based on original initial image which is free of compassion or distortion. SSIM index estimates perceived errors which mean that is consider image distortion as perceived alteration in structural information. It is based on estimating when the pixels have inter dependencies particularly when theses pixels are spatially close. Inter dependencies provides significant structure information of the objects in visual scene. SSIM can be mathematically calculated as given in 10

$$SSIM(c, s) = \frac{(2\mu_c\mu_s + C_1)(2\sigma_{cs} + C_2)}{(\mu_c^2 + \mu_s^2 + c_1)(\sigma_c^2 + \sigma_s^2 + C_2)} \tag{10}$$

where

$\mu_x$the average of c;

μ_y the average of$s$;

σ_x^2 the variance of c;

σ_y^2 the variance of s;

σ_{xy} the covariance of c and s;

$C_1{=}(k_1L)^2$,$C_2{=}(k_2L)^2$ two non consistent varibles to stable down the division with weak divisor;

L the dynamic domain of the pixel-values (usually this is $2^{\#bits\ per\ pixel}{-}1$);

$k_1{=}0.01$and $k_2{=}0.03$ by default.

SSIM is a decimal value and $SSIM \in [1, -1]$ and the value 1 refers to identical set of data.

On the other hands, some researchers claims that SSIM does not provide a correlation with HVS better that MSE valuesDosselmann and Yang [1]. Despite SSIM declares to generate quality measurements based on human perception, its equations does not consist any involved of HVS modeling and its totally relay on non perceptual parameters.

3) PSNR-HVS : It is an adaptive quality metrics of PSNR while taking into account HVS Egiazarian et al. [2]. It is based on taking away the mean shifting and contrast stretching utilizing a scanning window as shown in figure 2.

The adaptive version of PSNR is given in equation 11 while $MSE_H$is computing the mean square error between the cover and stegoed image with respect of HVS. PSNR-HVS also consider the high human sensitivity in low frequency range.

$$PSNR - HVS = 10\log(\frac{MAX^2}{MSE_H}) \qquad (11)$$

$$MSE_H = K\sum_{i=1}^{I-7}\sum_{j=1}^{J-7}\sum_{m=1}^{8}\sum_{n=1}^{8}((X[m,n]_{ij} - X[m,n]_{ij}^{e})T_c[m,n])^2 \qquad (12)$$

where i, j show image size,

$K = 1/[(I-7)(J-7)64]$

$X_{ij}$are DCT coefficient of 8*8 image block which the coordinates of its left upper corner are equal to i,j

$X_{ij}^{e}$are the DCT coefficient of the corresponding block in the cover image

T_c is the matrix of correcting factors. In PSNR-HVS the authors use quantization table shown in table I

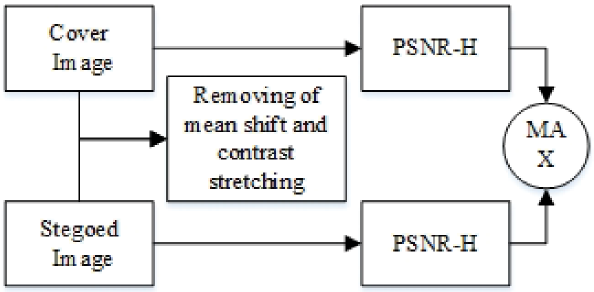
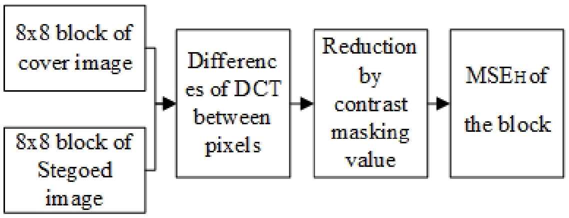


Figure 2. PSNR-HVS Flowchart

1.6084	2.3396	2.5735	1.6284	1.0723	0.6434	0.5046	0.4219
2.1446	2.1446	1.8382	1.3545	0.9898	0.4437	0.4289	0.4679
1.8382	1.9796	1.6084	1.0723	0.6434	0.4515	0.3730	0.4596
1.8382	1.5138	1.1698	0.8874	0.5046	0.2958	0.3217	0.4151
1.4297	1.1698	0.6955	0.4596	0.3785	0.2361	0.2475	0.3342
1.0723	0.7353	0.4679	0.4021	0.3177	0.2277	0.2277	0.2797
0.5252	0.4021	0.3299	0.2958	0.2499	0.2145	0.2145	0.2548
0.3574	0.2797	0.2709	0.2626	0.2298	0.2574	0.2499	0.2600

Table I
T_cMATRIX USED IN PSNR-HVS

4) PSNR-HVS-M : This metric designed to enhance the performance of PSNR and MSE Ponomarenko et al. [3]. It divided the input image into $8 \text{X} 8$ non-overlapping blocks follow by applying Discrete Cosine Transform DCT $\delta(i, j)$. Then, the differences of $\delta(i, j)$ for the cover and stegoed image is weighted for every $8 \text{ X } 8$ non-overlapping blocks by using Contrast Sensitivity Function (CSF). Figure 3 shows the basic flowchart for PSNR-HVS-M.



8x8 block of cover image
8x8 block of Stegoed image
Differenc es of DCT between pixels
Reduction by contrast masking value
MSEH of the block

Figure 3. PSNR-HVS-M Flowchart

III. SUBJECTIVE EVALUATION

The subjective quality of the stego images has been assessed by evaluating mean opinion score. A series

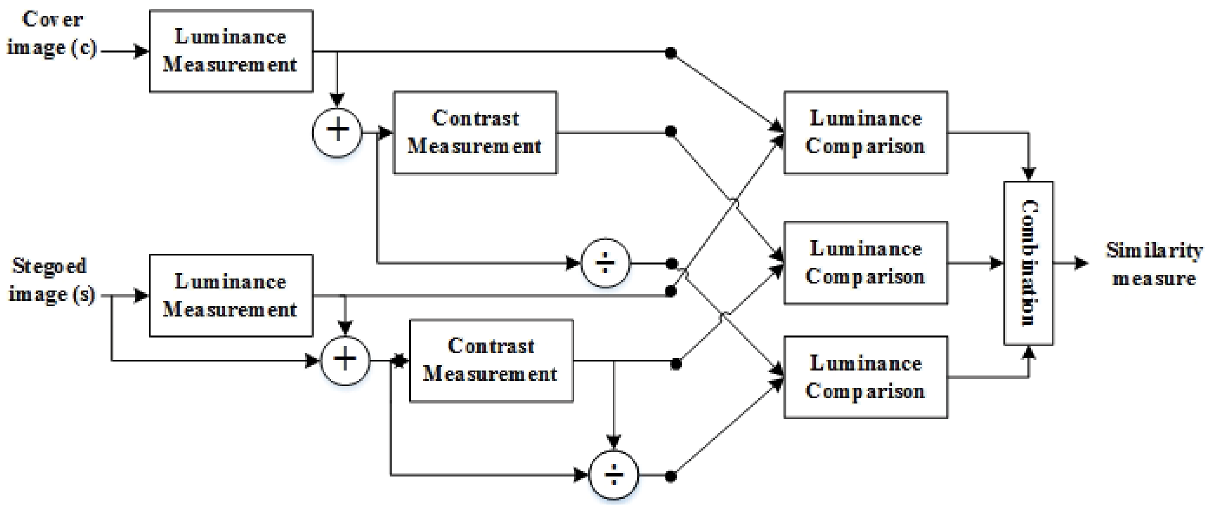


Figure 1. Flowchart diagram of SSIM for steganography

of stego images has been presented to 20 subjects in a Double Stimulus arrangement. The inspection environment has been organized as described in [11]; however, the image quality score has been assessed on a scale of 1 to 5, which is then normalized to 1. The Difference Mean Opinion Score (DMOS) is calculated and normalized to one for comparison with the objective quality metrics. The difference score is calculated as 13

$$d_{ij} = s_{ij_{ref}} - s_{ij} \tag{13}$$

The observers are asked to clarify which one is the cover and which one is the stegoed image and asked them to rate based on MOS rating which can be range from Excellent (5) to bad (1) as given in Table II.

Table II
MOS RATING

MOS	Quality	Impairment
5	Excellent	Unnoticeable
4	Good	Noticeable but not irritating
3	Fair	Slightly irritating
2	Poor	Irritating
1	Bad	Very irritating

The difference scores that are zero in the case of reference image have been removed for further processing. These difference scores are converted to the Z-scores [18] by using the following equations 14, 15 and 16 when N donates the total number of images seen by the observer. Then, the resulted Z-scores are scaled from [0,1] in order to create a suitable evaluation model.

$$\mu = \frac{1}{N}\sum_{j=1}^{N} d_{ij} \tag{14}$$

$$\sigma_i = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N}(d_{ij} - \mu_i)^2} \tag{15}$$

$$z_{ij} = \frac{d_{ij} - \mu_i}{\sigma_i} \tag{16}$$

Then, the Difference Mean Opinion Score (DMOS) can be computed as the average of all tested Z-scores for specific image as given in equation 17 where M is number of observers (20 observers in this study).

$$DMOS = \frac{1}{M}\sum_{i=1}^{M} z'_{ij} \tag{17}$$

DMOS is the most popular subjective metric and its provide a good quality evaluation based on the perceptual imperceptibility.

IV. Objective Vs. Subjective Correlation Analysis

In order to analyze the correlation between subjective evaluation and subjective evaluation mainly "PSNR", the experiment involves hiding the same secret message into red, green and blue level of Lenna and Pepper images correspondingly using LSB method. The PSNR value of embedding the same amount in red, green and blue stego image was similar in both Lenna and Pepper images as shown in Table III. However, the result of the three stego images were judge by human eyes for Lenna and Pepper by comparing the quality of the original image and each steged image. Mindful inspections were done to evaluate the degradation of each stego image using 20 observers. In Figure 5 and 6, the most close stego image to the original image is the one with embedding the data in blue level. However, the worse stego image were obtained from embedding the data in the green level. This evaluation allow us to decide that based on human vision sensitivity, human eyes are more sensitive to green color and less sensitive to blue color.

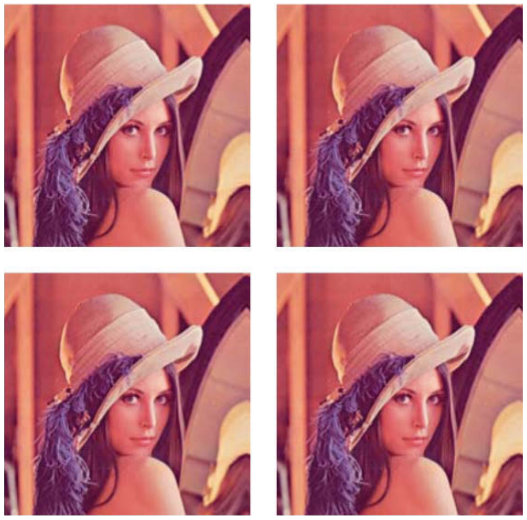
In addition, the objective and subjective image quality scores have been normalized to present a comparison. The plots in Figure 4 and 5 show the image quality scores for the test images (Lenna and Pepper) respectively. The difference mean opinion score is shown as MOS in the plots. As the mean squared error increases with distortion while the other metrics have opposite behavior, the mean squared error is adjusted to show similar behavior for better comparison; the MSE* in the plots is given as

$$MSE* = 1 - MSE^{normalized}$$

In general, it is found that the objective quality metrics that are driven from the HVS features have better correlation with the subjective assessment compared to standard MSE and PSNR. The universal image quality index has shown highest correlation with the subjective score due to its ability to detect luminance distortion and loss in the correlation of cover and stego images. The UIQI score for the second test sample is closer to the subjective evaluation. It is perhaps due to low luminance in the cover image that has been severely affected by the steganography.Therefore, the HVS PSNR follows the MOS trajectory when the data encoding is more than four bits.

Color	PSNR (Lenna)	PSNR (Pepper)
Red	45.3087	44.8547
Green	45.3057	44.8658
Blue	45.3158	44.8789

Table III
PSNR for embedding data in different level



The left top image is the original image of Lenna, the right top image is embedding data in Red level, the left down image is embedding data in Green level and the right down image is embedding data in Blue level.

Figure 5. Secret message hidden in different color level of Lenna

The objective and subjective image quality scores have been normalized to present a comparison. The plots in Figure 4 and 5 show the image quality scores for the test images (Lenna and Pepper) respectively. The difference mean opinion score is shown as MOS in the plots. As the mean squared error increases with distortion while the other metrics have opposite behavior, the mean squared error is adjusted to show similar behavior for better comparison; the MSE* in the plots is given as $MSE* = 1 - MSE^{normalized}$

In general, it is found that the objective quality metrics that are driven from the HVS features have better correlation with the subjective assessment compared to standard MSE and PSNR. The universal image quality index has shown highest correlation with the subjective score due to its ability to detect luminance distortion and loss in the correlation of cover and stego images. The UIQI score for the second test sample is closer to

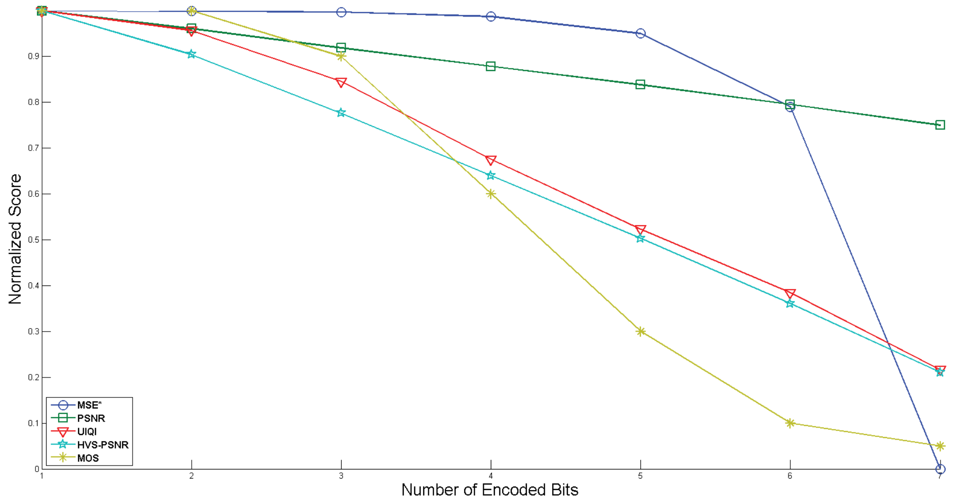
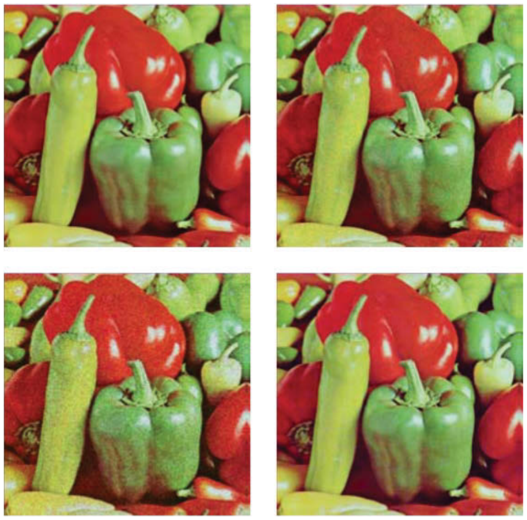


Figure 4. Normalized image quality score



The left top image is the original image of Pepper, the right top image is embedding data in Red level, the left down image is embedding data in Green level and the right down image is embedding data in Blue level.

Figure 6. Secret message hidden in different color level of Pepper

the subjective evaluation. It is perhaps due to low luminance in the cover image that has been severely affected by the steganography.Therefore, the HVS PSNR follows the MOS trajectory when the data encoding is more than four bits.

A. *Processing Time Evaluation*

The assessment fastness is an important standard in some cases. This section intends to provide speed median of some evaluation metrics. The test were execute on Intel Core i7 Q720 CPU streaming at 1.60 GHz with 8 GB RAM using Matlab R2014a. As demonstrate in Table IV, PSNR, UQI and SSIM are faster than PSNR-HVS and PSNR-HVS-M.

Quality Assessment Metric	Processing Time (second)
PSNR	0.2548
PSNR-HVS	11.6482
PSNR-HVS-M	11.9854
SSIM	0.3845
UQI	0.2578

Table IV
PROCESSING TIME FOR DIFFERENT QUALITY METRIC

V. CONCLUSION

A comparative study of the existing image quality metrics is performed for the steganographic images. The image quality score for commonly used objective quality metrics in the field of steganography has been compared with the subjective assessment performed by 20 observers. It is found that the selected objective metrics has a poor correlation with the subjective assessment, and may fail to accurately evaluate the performance of a steganography algorithm. The HVS based metrics have better correlation compared to the standard pixel based metrics such as MSE and PSNR; this shows the effectiveness of using features of HVS in the quality assessment metrics.

References

[1] Richard Dosselmann and Xue Dong Yang. A comprehensive assessment of the structural similarity index. *Signal, Image and Video Processing*, 5(1):81–91, 2011.

[2] Karen Egiazarian, Jaakko Astola, Nikolay Ponomarenko, Vladimir Lukin, Federica Battisti, and Marco Carli. A new full-reference quality metrics based on hvs. In *CD-ROM proceedings of the second international workshop on video processing and quality metrics*, pages 2–5, 2006.

[3] Nikolay Ponomarenko, Flavia Silvestri, Karen Egiazaria, Carli Marco, Jaakko Astola, and Vladimir Lukin. On between-coefficient contrast masking of DCT basis functions. *CD-ROM Proc. of the Third International Workshop on Video Processing and Quality Metrics*, 4(1):1–4, 2007.

[4] Sviatoslav Voloshynovskiy, Alexander Herrigel, Nazanin Baumgaertner, and Thierry Pun. A stochastic approach to content adaptive digital image watermarking. *In Information Hiding*, pages 211–236, 2000.

[5] Zhou Wang and Alan Bovik. A universal image quality index. *Signal Processing Letters*, 4:81–84, 2002.

[6] Zhou Wang, Hamid R. Sheikh, and Alan C. Bovik. Objective video quality assessment. *The handbook of video databases: design and applications*, 2003.

[7] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image Quality Assessment: From Error Visibility to Structural Similarity. *Image Processing, IEEE Transactions*, 13(600-612), 2004.

[8] Zhou Wang, Eero P. Simoncelli, and Alan C. Bovik. Multiscale structural similarity for image quality assessment. In *Conference Record of the Thirty-Seventh Asilomar Conference on, Signals, Systems and Computers,*, pages 1398 – 1402 Vol.2, 2004.

[9] Raniyah Abdullah Wazirali, Zenon Chaczko, and Anup Kale. Digital multimedia archiving based on optimization steganography system. In *2014 Asia-Pacific Conference on Computer Aided System Engineering (APCASE)*, pages 82–86. IEEE, February 2014.

2015 Asia-Pacific Conference on Computer Aided System Engineering

Haptic Middleware Based Software Architecture for Smart Learning

Zenon Chaczko, Cheuk Yan Chan, Lucia Carrion
Faculty of Engineering and Information Technology
University of Technology, Sydney (UTS)
Sydney, New South Wales, Australia
Email: zenon.chaczko@uts.edu.au,
cheukyan.chan, lucia.carrion{@student.uts.edu.au}

Wael Mohammad G. Alenazy
Self-Development Skills Department, Preparatory Year
Deanship
King Saud University (KSU)
Riyadh, the Kingdom of Saudi Arabia
Email: walenazy@ksu.edu.au

Abstract— **The software architecture of smart learning environment can be perceived as an environment that is equipped with various audio-visual objects to capture human motion, utterance and gesture; allowing the teacher to deliver lectures to both local and remote audience through the Internet. The interactive objects in such architectural environment are interfaced with simple navigation, depending on operation characteristics, voice, tactile and visual interaction with the aim to improve Human-Machine Interaction. This facilitates effective data acquisition and statistical analysis, in order to assist in decision making by the participants, as well as, apply them in the process of self-assessment. This paper discusses the design and implementation of integrating haptic technologies into the architecture of smart learning environment by designing components of service oriented software middleware that defines a common gesture framework, and thus allowing multiple haptic and gesture peripherals to share a within common protocol, as well as, enabling individual device to work as stand-alone entity.**

Keywords— Software architecture, Smart learning environment, Haptic Middleware

I. Introduction

Smart learning environment can maximize a student's potential and their cognitive retention, and smart classrooms support various learning and teaching methods, such as student-centered learning method [1]. Therefore, the main objective of implementing haptic middleware technologies in classrooms to assist students and to enable teachers to facilitate better achievement of students; and therefore increase the interactivity. The smart classroom can entertain students through its visualization facility by providing highly engaging visuals and animations which will make the learning environment become more enjoyable for both teachers and students and also improve students' "overall academic performance" [2].

Smart learning comprises of smart pedagogies, smart content, learning methodology and smart equipment. This creates a complex system characterized by many connections and various sections. An interactive classroom design involves various complexities and the complete list of parts and connections may be may not be possible. Creating an innovative and active learning environment to achieve some goals requires concerted effort by the stakeholders, which can be achieved through proper planning [3].

There is a growing gap between technology and education. The research area relates to develop the future generation classroom through the development of smart classrooms system by designing smart system architecture.

This paper presents the analysis and development on a haptic middleware system the aim to replace traditional inputs on the system by integrating haptic body/hand motion control middleware to simplify the overall teaching process within a smart classroom context.

II. Smart Learning Environment

The smart classroom relates to the optimization of the interaction between the teacher and the learner, development of teaching presentation, accessibility of learning and teaching resources at own convenience, detection and awareness of the context, classroom management and others. The overall vision of an intelligent classroom would include a collaboration of sensors, passive computational devices and lower power networks which provide infrastructure for context-aware smart classroom that sense and respond to the human activities.

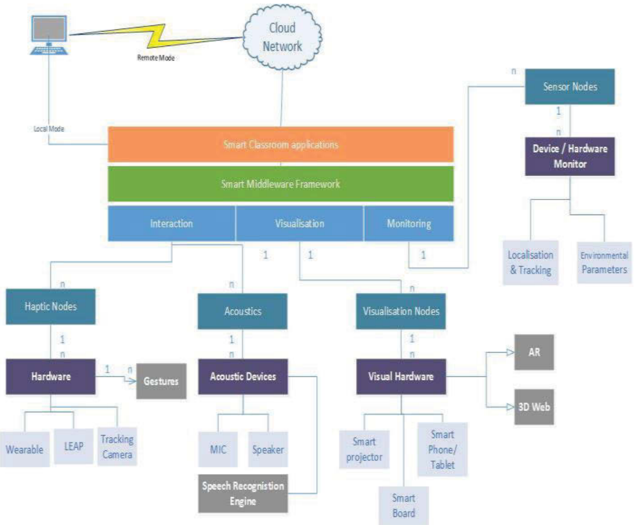


Fig. 1. Conceptual View of Smart Learning Environment

978-1-4799-7588-4/15 $31.00 © 2015 IEEE
DOI 10.1109/APCASE.2015.52

The proposed solution (Fig. 1) requires the development and integration of various components in areas such as: distributed computing, networking, acquisition of sensor data, speech recognition system, signal processing and human identification [4].

Based on the conceptual View, the new system can be accessed in real-time enabling interaction between the teacher and the learners through media tools. The interactive equipment is interfaced with simple, clear navigation, reliable operation characteristics, voice, touch and visual interaction which improve Human Machine Interaction (HMI). The hardware in the future generation classroom meets the requirements of the multi-terminal points. Multi-terminal means that a single active class can conduct multiple classrooms spanning geographically. The smart classroom can also record and store basic data in the computer, in order to assist decision making of the participants as well as use them for self-conducted activity [5][6]. Moreover, the new idea of the system architecture is a construction of a robust and scalable middleware system that supports the level of interactions required within the context of smart classroom. Software components of the smart classroom include:

- Smart classroom applications - High level implementation of the smart classroom logic through utilizing the middleware framework within a cloud environment
- Smart middleware framework - Bridge and collaborate a suite of peripherals via IOT framework, implementing a set of common library between these devices. It also will represent the underlying connectivity and data distribution bus which will interlink various hardware and software components and include critical services for arbitration and haptic ranging. This component involves different APIs for machine vision, actuator and sensor devices.

This includes capturing of important information in a classroom to achieve high level interaction, dynamic learning environment and effective delivery to teaching contents. The information is sensed through systematically arranged microphones and tracking cameras. The sensing quality mainly depends on the performance of the audio and video devices, appropriate positioning of devices etc. [5, 6]. Attentive teaching can further be enhanced through context-awareness sensing such as speech recognition system that interpret voice into command, tracking the speaker‟ signal strength, automatic zooming and focus from emphasis of vital images, sensor fusion which extract behavioral information in the classroom and gaze tracking to predict some actions. The sensed information is then rendered.

Haptic devices and sensors installed in convenient places in the classroom can detect automatically parameters such as noise, temperature, odor, light and others, as well as adjust lamps, air conditioning so as to maintain temperature, light, sound and fresh air which are suitable for mental and physical status in smart classroom [7, 8].

III. System Analysis and Design

A. Requirements

New haptic middleware system is defined through gathering raw requirements from multiple phases of prototyping as well as via discussion with the stakeholder. These raw requirements (Table 1) are then refined by extracting individual requirements on the core functionality of the middleware.

TABLE I. General Requirements of Haptic Middleware

Requirement ID	Requirements Description	Priority
REQ0001	Ability to handle a common library set of gestures	Mandatory
REQ0002	Ability to interface additional devices	Mandatory
REQ0003	Ability to be used by a third party software	Mandatory
REQ0004	Allow user to interface with individual device using the same function calls	Mandatory
REQ0005	Ability to track distance if detected	Mandatory
REQ0006	Ability to determine velocity if detected	Mandatory
REQ0007	Ability to choose between devices when user is in range	Mandatory
REQ0008	Have the capability to add new common libraries	Mandatory
REQ0009	Providing an API that can be implemented and used	Mandatory
REQ0010	Allow any sensors to be added and implemented	Mandatory
REQ0011	Ability to send out gestures via network	Mandatory
REQ0012	Ability to capture finger coordinates if possible	Mandatory
REQ0013	Ability to capture hand coordinates if possible	Mandatory
REQ0014	Ability to capture live video feed of the device if possible	Mandatory
REQ0015	Ability to switch devices when in range	Mandatory
REQ0016	Ability to guess the device a user is moving towards	Mandatory

B. Architectural Analysis

Architectural patterns are analyzed and reviewed to meet the design of the middleware system to fit to the Smart Learning Environment context. A high level design will be developed using SysML which will then be decomposed into lower level design, which determines the functionalities of which component block. Service Oriented Architecture (SOA) made up from a collection of discrete software modules, known as services that collectively provide the complete functionality of a large software application. This allow users to combine and reuse them in the production of applications. Services may be implemented using traditional languages like Java, C, C++, C#, Visual Basic, COBOL, or PHP. The prototype in development for this paper is chosen to be C#.

Services adhere to a communications agreement, as defined collectively by one or more service-description documents. This maintain a relationship that minimizes dependencies and only requires that they maintain an awareness of each other. Beyond descriptions in the service contract, services hide logic from the outside world. Benefits include ability to leverage existing infrastructure and Applications, reusable logic and better control on encapsulation of logic. Limitations include lower security on services, reusability on service level rather than micro class level as well as time performance of service oriented applications is significantly lower than that of applications using more traditional means of communication.

Service oriented architecture would be beneficial for the haptic control middleware as the middleware framework will need to provide communication between integrations between new system developments that are going to add the library to. On the other hand service-oriented communication has its toll on the time performance of the system as it is possible to have large chunks of data on different devices to be transmitted to the middleware and at the same time compute and interpret before output to another system. This may cause the performance to be significantly degraded.

C. Architectural Design

The Sensory Hardware service (Fig. 2) defines a common set of functions and data requirements for any device and defines a common set of gestures utilizing a common gestures library. This service is handled by the Core Middleware service and its logic interfaced to select exclusive data from the haptic controllers LEAP motion manager manages all connections to the LEAP motion device and implements a defined set of methods that can return consistent data to the sensory hardware service. It implements a common gesture library. Microsoft Kinect manager manages all connections to the Microsoft Kinect device and implements a defined set of methods that can return consistent data to the sensory hardware service. It implements a common gesture library. Thalmic Labs MYO manager manages all connections to the Thalmic Labs MYO arm band and implements a defined set of methods that can return consistent data to the sensory hardware service. It implements a common gesture library. The common gesture library contains a common set of gestures that can be implemented by different haptic devices. This library is implemented within the sensory hardware service. The Distance Resource service (Fig. 3) handles the resource selection logic based on the output from the sensory hardware service. This is achieved by assessing the distance of the user, the velocity as well as the confidence ratio of the device. The Core Middleware service (Fig. 3) is the main service which is used to implement the haptic control middleware.

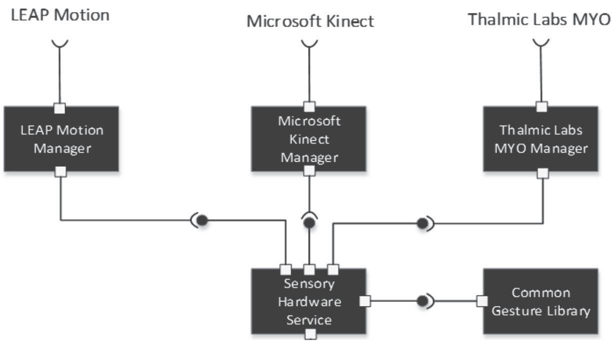


Fig. 2. High level Design (Sensory Hardware Service)

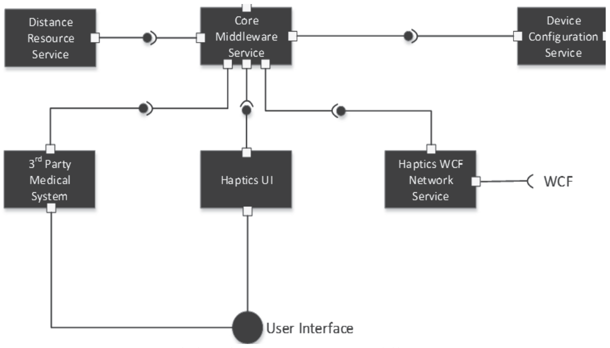

Distance Resource Service
Core Middleware Service
Device Configuration Service
3rd Party Medical System
Haptics UI
Haptics WCF Network Service
WCF
User Interface

Fig. 3. High level Design (Core Middleware Service)

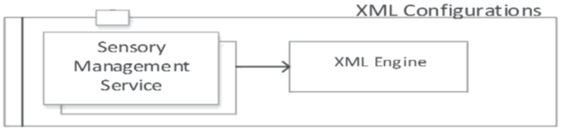


Fig. 4. Implementation Architecture (Lower Configuration Layer)

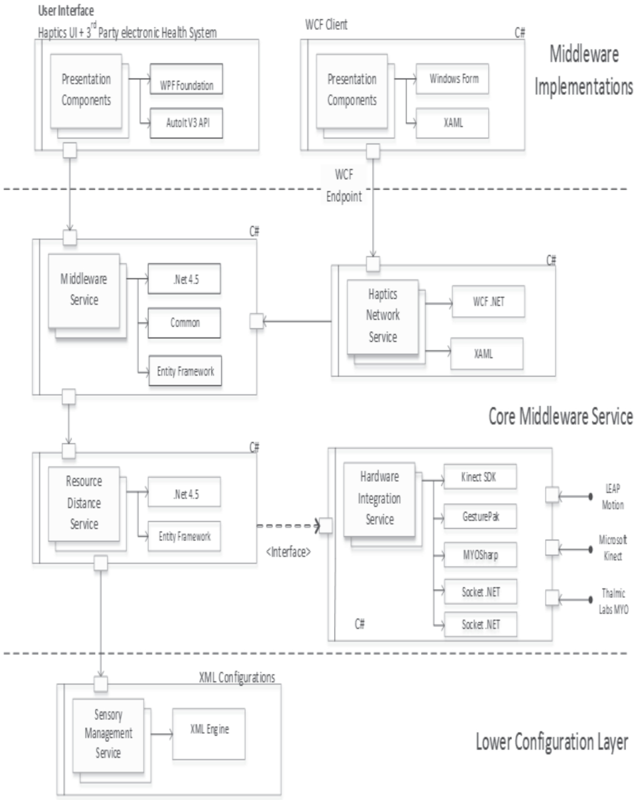


Fig. 5. Implementation Architecture

This is the service point which is used for other third party systems to interface and interact. The device configuration service is an XML configuration of all devices that are defined. This will contain the name of the device, the connectivity method as well as some basic capabilities of the item. This is a network service which outputs gestures data out to the network space. This is aim at interacting systems across the internet using haptic controllers on the workstation.

D. *Implementation Architecture*

The middleware framework is implemented and displayed utilizing a Layered Service oriented architecture would be beneficial for the haptic control middleware as the middleware framework will need to provide communication between integrations between new system developments that are going to add the library to. The middleware is defined into two main service layers- Lower Configuration Layer and Core Middleware Service Layer.

The Lower Configuration Layer (Fig. 4) implements the XML configuration of all devices that are defined. This will contain the name of the device, the connectivity method as well as some basic capabilities of the item. This is achieved by utilizing an XML Engine to create and parse XML files. The Core Middleware Service (Fig. 5) implements the haptic controller framework. The service composed of four main services:

- Middleware Service – This is the outbound service which implements the middleware API for third party systems.
- Haptics Network Service – This is a network service which outputs gesture data across the internet via TCP/IP or HTTP endpoints. This utilizes Windows Communication Foundation framework.
- Resource Distance Service – The resource distance service controls the logic in switching hardware devices based on availability and other parameters such as distance and velocity.
- Hardware Integration Service – Hardware Integration service handles all the devices connected to the middleware system. These devices are implemented using their corresponding libraries and uses a common interface for ease of interaction.

E. *Low Level Design*

Core low level design shown will include the haptic middleware API, common hardware interface, common gestures library, individual hardware classes. Haptics Middleware API (Fig. 6) implements the following classes and methods:

- *Haptics Device* - This is a device enumerator representing the devices implemented within the project. There are three devices in total: Microsoft Kinect camera, Thalmic Labs MYO arm band and LEAP motion haptic controller
- *IHaptics* - This is an interface which hooks onto the main hardware interface to obtain data from all devices.
- *Haptics* - This implements the IHaptics interface. This class handles the resource distance service on selecting devices depending on distance as well as containing the main methods for the middleware API:
  - o *deviceSelected()* – Returns information on which device is currently selected to read its output data.
  - o *eulerAngle()* – Returns additional information of the MYO reading on the gyroscope. The data returned will contain the roll, yaw and pitch of the Thalmic Labs MYO arm band.
  - o *getCoordinate()* – For camera devices, provides the live coordinates of the hand in x, y and z axis.
  - o *getDistance()* – Returns the distance of the user's hand from the camera devices. The data return is in centimeters.
  - o *getGesture()* – Returns the gesture detected from the selected listening device. The gesture is defined within the common gestures library
  - o *Initialise()* – Initializes all connected devices;
  - o *Shutdown()* – Shutdown all connected devices;

Common Hardware Interface (Fig. 7) implements the following classes and methods:

- *Listening Mode* - This is a part of the common hardware library which defines all haptic controllers and sensors. This enumerator contains all supported connection methods for devices implemented within this project. The supported connection methods are Bluetooth connectivity, connectivity via HTTP endpoints, connectivity via TCP/IP endpoints and connectivity via USB.
- *IHardware* - Defines the common hardware interface for this project, this implements the hardware service to handle all devices and sensors for the middleware:
  - o *getCoordinate*() – If the device is defined as a camera device, then it will obtain the coordinates of the user's hand within the project.
  - o *getGesture*() – Obtains the gesture of the device. The gesture is defined within the common gesture library.
  - o *Initialise*() – Initializes the haptic controller
  - o *SetParameters*() – Obtains the configuration data from the XML configuration files and setup the device for connectivity and communication.
  - o *Shutdown*() – Shutdown the device

Common Gestures class (Fig. 8) defines the gesture library for the middleware framework. Within the project, the following five common gestures were created for the devices:

- *Circular Anti-clockwise gesture* – hand moving anti-clockwise
- *Circular clockwise gesture* – hand moving clockwise
- *Okay gesture* – hand showing OK or an upper O shape in the air with both hands
- *Swipe Left gesture* – Hand swiping left
- *Swipe Right gesture* – Hand swiping right
- *Unknown* – An unknown gesture as a safety fall back if the device is unable to recognize the user's gesture

The class implementation of the LEAP motion [9] device (Fig. 9) connects to the main hardware and gesture library to maintain consistency between communicating devices

- *ILeapMotion* – Interface created to hook onto the common hardware library
- *LeapMotion* – Implements the common hardware interface with the methods defined below.

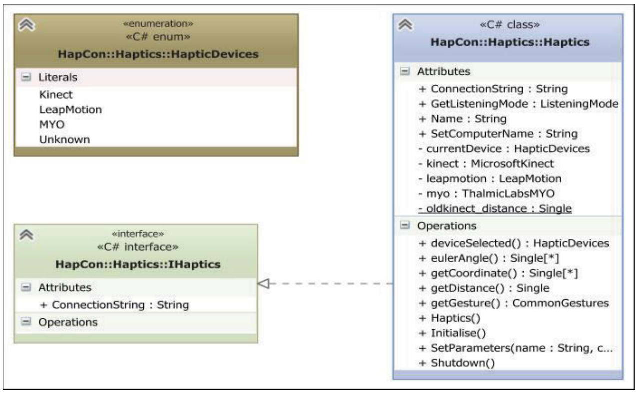

«enumeration»
«C# enum»
HapCon::Haptics::HapticDevices
Literals
Kinect
LeapMotion
MYO
Unknown
«interface»
«C# interface»
HapCon::Haptics::IHaptics
Attributes
+ ConnectionString : String
Operations
«C# class»
HapCon::Haptics::Haptics
Attributes
+ ConnectionString : String
+ GetListeningMode : ListeningMode
+ Name : String
+ SetComputerName : String
- currentDevice : HapticDevices
- kinect : MicrosoftKinect
- leapmotion : LeapMotion
- myo : ThalmicLabsMYO
- oldkinect_distance : Single
Operations
+ deviceSelected() : HapticDevices
+ eulerAngle() : Single[*]
+ getCoordinate() : Single[*]
+ getDistance() : Single
+ getGesture() : CommonGestures
+ Haptics()
+ Initialise()
+ SetParameters(name : String, c...
+ Shutdown()

Fig. 6. Low Level Design (Haptics Middleware API)

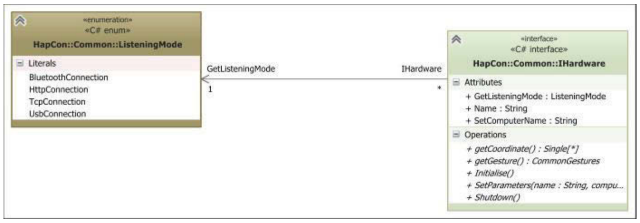

«enumeration»
«C# enum»
HapCon::Common::ListeningMode
Literals
BluetoothConnection
HttpConnection
TcpConnection
UsbConnection
GetListeningMode
1
IHardware
*
«interface»
«C# interface»
HapCon::Common::IHardware
Attributes
+ GetListeningMode : ListeningMode
+ Name : String
+ SetComputerName : String
Operations
+ getCoordinate() : Single[*]
+ getGesture() : CommonGestures
+ Initialise()
+ SetParameters(name : String, compu...
+ Shutdown()

Fig. 7. Low Level Design (Common Hardware Interface)

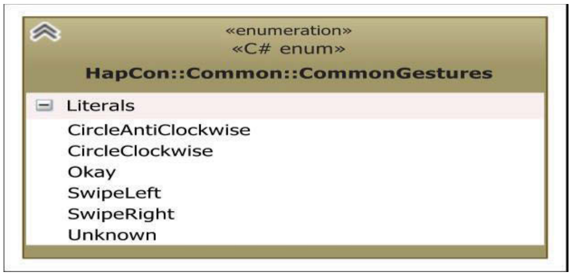


Fig. 8. Low Level Design (Common Gestures)

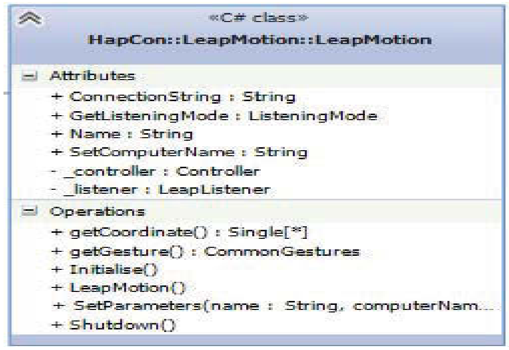

«C# class»
HapCon::LeapMotion::LeapMotion
Attributes
+ ConnectionString : String
+ GetListeningMode : ListeningMode
+ Name : String
+ SetComputerName : String
- _controller : Controller
- _listener : LeapListener
Operations
+ getCoordinate() : Single[*]
+ getGesture() : CommonGestures
+ Initialise()
+ LeapMotion()
+ SetParameters(name : String, computerNam...
+ Shutdown()

Fig. 9. Low Level Design (Leap Motion)

The Microsoft Kinect (Fig. 10) device connects to the main common hardware and gesture library to maintain consistency between device communications using:

- *IMicrosoftKinect* – Interface created to hook onto the common hardware library.
- *MicrosoftKinect* – Implements the common hardware interface by utilising Microsoft Kinect SDK and Kinect Gesture library.

Implementation of the Thalmic Labs MYO armband device (Fig. 11) connects to the main common hardware and gesture library to maintain consistency between device communications.

- *IThalmicLabsMYO* – Interface created to hook onto the common hardware library.
- *ThalmicLabsMYO* – Implements the common hardware interface by utilizing Thalmic Labs class library and MYOSharp wrapper interops.

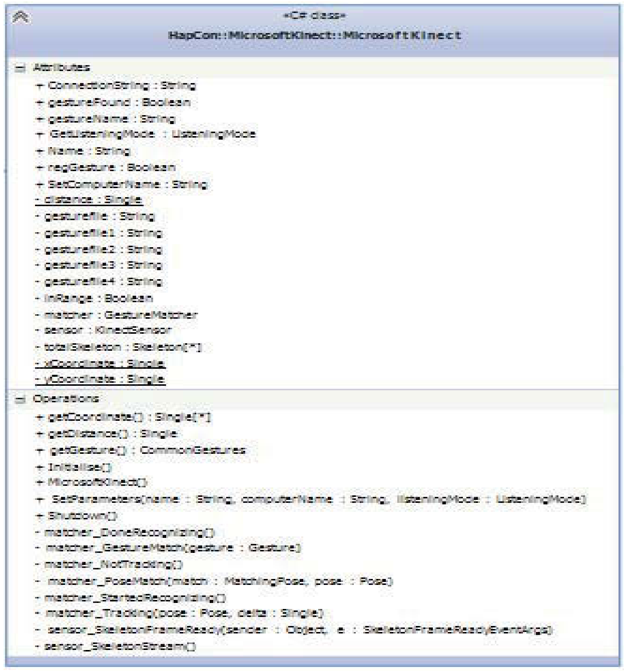

«C# class»
HapCon::MicrosoftKinect::MicrosoftKinect
Attributes
+ ConnectionString : String
+ gestureFound : Boolean
+ gestureName : String
+ GetListeningMode : ListeningMode
+ Name : String
+ regGesture : Boolean
+ SetComputerName : String
- distance : Single
- gesturefile : String
- gesturefile1 : String
- gesturefile2 : String
- gesturefile3 : String
- gesturefile4 : String
- inRange : Boolean
- matcher : GestureMatcher
- sensor : KinectSensor
- totalSkeleton : Skeleton[*]
- xCoordinate : Single
- yCoordinate : Single
Operations
+ getCoordinate() : Single[*]
+ getDistance() : Single
+ getGesture() : CommonGestures
+ Initialise()
+ MicrosoftKinect()
+ SetParameters(name : String, computerName : String, listeningMode : ListeningMode)
+ Shutdown()
- matcher_DoneRecognizing()
- matcher_GestureMatch(gesture : Gesture)
- matcher_NotTracking()
- matcher_PoseMatch(match : MatchingPose, pose : Pose)
- matcher_StartedRecognizing()
- matcher_Tracking(pose : Pose, delta : Single)
- sensor_SkeletonFrameReady(sender : Object, e : SkeletonFrameReadyEventArgs)
- sensor_SkeletonStream()

Fig. 10. Low Level Design (Microsoft Kinect)

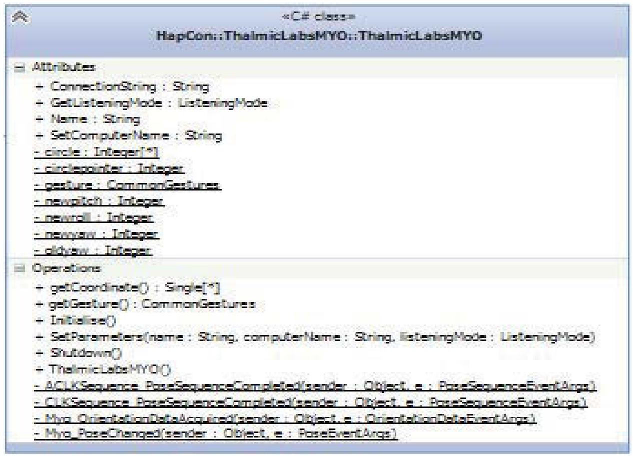

«C# class»
HapCon::ThalmicLabsMYO::ThalmicLabsMYO
Attributes
+ ConnectionString : String
+ GetListeningMode : ListeningMode
+ Name : String
+ SetComputerName : String
- circle : Integer[*]
- circlepointer : Integer
- gesture : CommonGestures
- newpitch : Integer
- newroll : Integer
- newyaw : Integer
- oldyaw : Integer
Operations
+ getCoordinate() : Single[*]
+ getGesture() : CommonGestures
+ Initialise()
+ SetParameters(name : String, computerName : String, listeningMode : ListeningMode)
+ Shutdown()
+ ThalmicLabsMYO()
- ACLKSequence_PoseSequenceCompleted(sender : Object, e : PoseSequenceEventArgs)
- CLKSequence_PoseSequenceCompleted(sender : Object, e : PoseSequenceEventArgs)
- Myo_OrientationDataAcquired(sender : Object, e : OrientationDataEventArgs)
- Myo_PoseChanged(sender : Object, e : PoseEventArgs)

Fig. 11. Low Level Design (Thalmic Labs MYO)

IV. Application Development

A. Development Technologies and Tools

For the development and testing of the haptic middleware design, the following hardware devices were integrated and implemented to the middleware:

- LEAP Motion (Fig. 12)
- Microsoft Kinect 360 (Xbox edition) (Fig. 13)
- Thalmic Labs MYO gesture armband (Fig. 14)

For developing the software application, the following software tools and libraries were used:

- Visual Studio 2013. This is the main program for developing the middleware system in C# and C++. It also allowed doe integrations to all the software SDK in the middleware system.
- Microsoft .Net 4.5 framework - As part of implementation towards C# and utilizing Microsoft visual studio development environment, a high majority of prototypes were created using this framework. The communication between the server and the client through HTTP & TCP/IP also utilizes the .NET framework via the Windows communication foundation (WCF) framework.
- Microsoft (MS) Kinect Software Development Library – This software library is used to initialize connectivity and communicate with MS Kinect. The components used within this library would include basic video streaming and skeleton tracking.
- Kinect Gesture Pak Software Development Library – This software library provides customizable gestures and store the tracking data into XML files and interpreted within the project. The aim is to assist in the integration of MS Kinect into the middleware system.
- LEAP Motion [9] Software Development Library – A C# library that handles the connectivity with the LEAP motion device. It allows the developers to define the listening space within the camera (i.e. from a circular cone viewpoint to a restricted box area) as well as provide hand and finger tracking on the device [8].
- Thalmic Lab MYO Software Development Library – This is a beta library in C++ which provides connection to the MYO armband. The library contains some basic data from the gyroscope, the accelerometer as well as a few general hand gestures. It has a built in Bluetooth connection functionality which simplifies the connectivity process when integrating Bluetooth connection with the middleware.
- MYOSharp C# Wrapper API - This is a C# library which wraps the MYO beta development library via interop with the C++ class library. This software is open-sourced and is available for personal and business use.

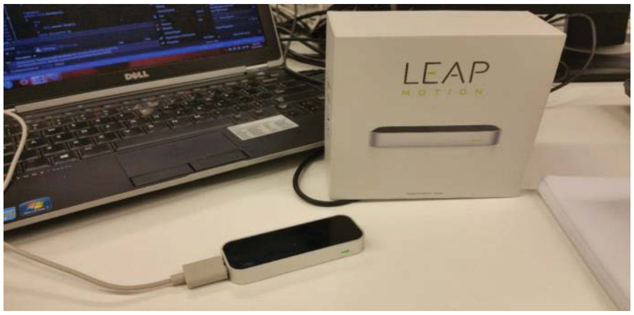


Fig. 12. LEAP Motion Haptics Controller

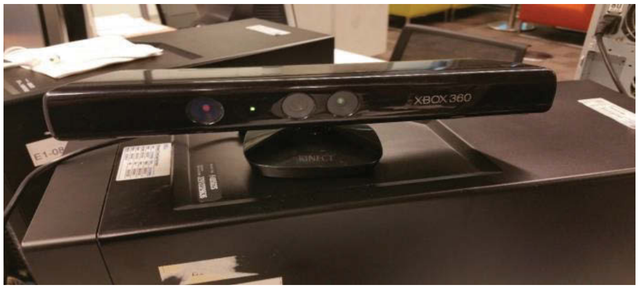


Fig. 13. Microsoft Kinect Camera (Xbox Edition)

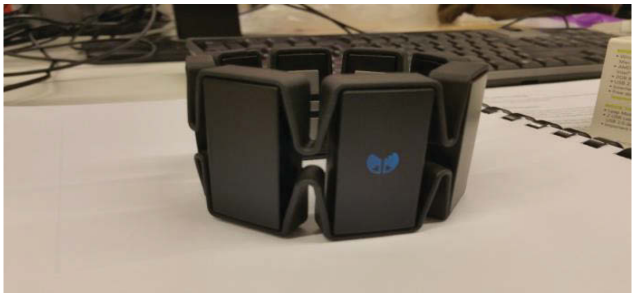


Fig. 14. Thalmic Lab MYO Gesture Armband

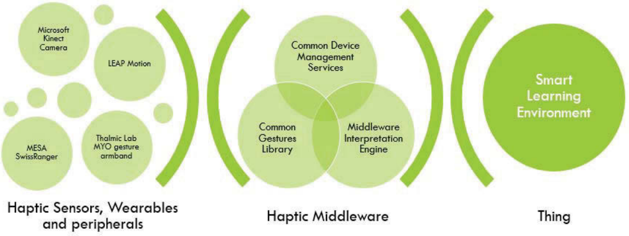
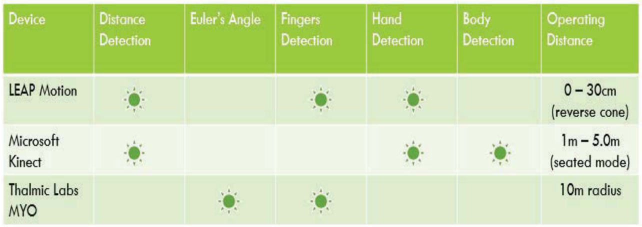
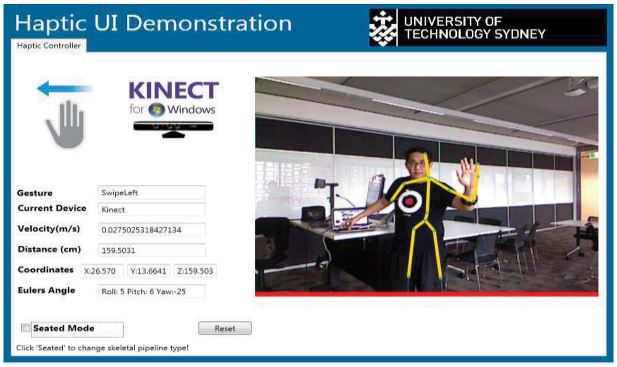

Microsoft Kinect Camera
LEAP Motion
MESA SwissRanger
Thalmic Lab MYO gesture armband
Common Device Management Services
Common Gestures Library
Middleware Interpretation Engine
Smart Learning Environment
Haptic Sensors, Wearables and peripherals
Haptic Middleware
Thing

Fig. 15. Haptic Middleware vision



Device	Distance Detection	Euler's Angle	Fingers Detection	Hand Detection	Body Detection	Operating Distance
LEAP Motion	●		●	●		0 – 30cm (reverse cone)
Microsoft Kinect	●			●	●	1m – 5.0m (seated mode)
Thalmic Labs MYO		●	●			10m radius

Fig. 16. Haptic controllers complementing their strengths over other controller's weaknesses



Gesture: SwipeLeft
Current Device: Kinect
Velocity(m/s): 0.0275025318427134
Distance (cm): 159.5031
Coordinates: X:26.570 Y:13.6641 Z:159.503
Eulers Angle: Roll: 5 Pitch: 6 Yaw:-25
□ Seated Mode
Click 'Seated' to change skeletal pipeline type!

Fig. 17. Final prototype for Haptic Middleware system

B. Prototype Development

The prototype of the haptic middleware system (Fig. 15) gives developers to utilize different haptic devices to interact with the system using each of the haptic controller's strength to cover for other haptic device's weaknesses (Fig. 16). The framework would provide useful data of such as gesture interactions, user or hand distances, and velocity of human motion as well as general skeleton statistics. In consideration of the above, the developed functional prototype demonstrates:

- Ability to integrate multiple peripheral and sensors into the framework.
- Ability for the sensors to interact based on a common gesture library.
- Ability for the middleware to sense user distance and velocity and calculate as a factor to determine the next listening haptic controller

V. Final System

A working prototype (Fig. 17) for the Haptic Middleware Framework designed and developed. Below are the functions tested for the API with all three hardware components implemented into the haptic middleware service:

- *deviceSelected*() – Returns information on which device is currently selected to read its output data.
- *eulerAngle*() – Returns additional information of the MYO reading on the gyroscope. The data returned will contain the roll, yaw and pitch of the Thalmic Labs MYO arm band.
- *getCoordinate*() – For camera devices, provides the live coordinates of the hand in x, y and z axis.
- *getDistance*() – Returns the distance of the user's hand from the camera devices. The data return is in centimeters.
- *getGesture*() – Returns the gesture detected from the selected listening device. The gesture is defined within the common gestures library

VI. Conclusion

Effective integration of haptic technologies into the architecture of smart learning environment by designing components of service oriented middleware is achievable through well executed architectural development and design.

The core of this research have been focused towards designing and creating the flexible architectural framework for the haptic middleware, providing the ability to allow each haptic controller's strength to be able to mitigate weaknesses of the other haptic devices.

Through a functional prototype, it is demonstrated that it is possible to replace traditional keyboard and mouse input with gesturing devices within the Smart Learning Environment context. The presented middleware enables advanced services for multiple gesture device handling to control components of the software system. Further development of the middleware will focus on improving usability of its services.

References

[1] Bouslama, F. & Kalota, F. 2013, 'Creating smart classrooms to benefit from innovative technologies and learning space design', *Current Trends in Information Technology (CTIT), 2013 International Conference on*, IEEE, pp. 102-6.

[2] Jena, P.C. 2013, 'Effect of Smart Classroom Learning Environment on Academic Achievement of Rural High Achievers and Low Achievers in Science', *International Letters of Social and Humanistic Sciences*, no. 03, pp. 1-9.

[3] Kossiakoff, A., Sweet, W.N., Seymour, S. & Biemer, S.M. 2011, *Systems engineering principles and practice*, vol. 83, John Wiley & Sons. K. Elissa, "Title of paper if known," unpublished.

[4] Sailor, W. 2009, *Making RTI work: How smart schools are reforming education through schoolwide response-to-intervention*, John Wiley & Sons.

[5] Harrison, R. 2013, *TOGAF® 9 Foundation Study Guide*, Van Haren.

[6] Sobh, T.M. & Elleithy, K. 2013, *Emerging Trends in Computing, Informatics, Systems Sciences, and Engineering*, Springer.

[7] Augusto, J.C., Nakashima, H. & Aghajan, H. 2010, 'Ambient intelligence and smart environments: A state of the art', *Handbook of Ambient Intelligence and Smart Environments*, Springer, pp.3-31.

[8] D Zuehlke 2010, *Smart Factory – Towards a factory of things*, Annual Reviews in Control, Germany

[9] Leap Motion 2014, *API Overview*, <https://developer.leapmotion.com/documentation/csharp/devguide/Leap_Overview.html>

Augmented Reality Based Monitoring of the Remote-Lab

Zenon Chaczko[1], Wael Alenazy[2], Lucia Carrion[1], Amy Tran[1]
[1]Faculty of Engineering and Information Technology
University of technology, Sydney (UTS), Sydney, NSW, Australia
E-mail: {Zenon.chaczko, Lucia.C.Carrion}@uts.edu.au, Amy.Tran@alumni.uts.edu

[2]Self-Development Skills Department, Preparatory Year Deanship
King Saud University (KSU)
Riyadh, The Kingdom of Saudi Arabia
E-mail: Walenazy@ksu.edu.sa

***Abstract*— Augmented Reality technology approach has been being adopted within the education sector. The advanced technology tools in many classes have the potential changed of users' attitudes toward the pedagogical and psychological objectives and goals. Moreover, augmented reality has not elicited so much attention within the corridors of education sector as it is now. In order to improve the interactive effectiveness in the smart classroom environment, there is a demand to tailor the innovation technology and align it with every changing requirements and capabilities of various users. Consequently, the educators are increasingly finding augmented reality suitable for deployment in education. In this paper, a project shows how Augmented Reality utilised with overlay Smart-Grid can support the learning process in attractive methods for monitoring events of captured scenes in remote-lab such as video stream, Web-link from smart devices' camera.**

Keywords—augmented reality; smart grid; remote-lab.

I. Introduction

Augmented Reality (AR) allows for the establishment of collaborative environments through teacher's and students' interaction within virtual objects leading to the creation of various interactive scenarios in the classroom. Virtual reality refers to the combination of various display and devices' interface which results from immersive overlay interactive computer-generated environment. In addition, the Mixed of reality refers to the merger of virtual overlay objects with on the real scene [1]. Although in terms of acceptance technology, a recent model has been appeared and showed that because of some smart technologies features, users are capable to function those devices easily [2]. Hence, AR as a smart innovation system is being incorporated in the learning and teaching processes to achieve the pedagogical objectives and goals demand. The AR system in classes allows the students to locate the students to discover and explore the virtual materials as they are in real through the use of overlaid scenes[3].

In this paper, an experiment has applied for utilising augmented reality (AR) elements within UTS Remote-Labe. The aims of the study are: firstly is to monitor captured (selected) of a scene for studying purposes, such as a video stream from an iPad camera, within the novel idea 'a smart grid layout'. The process of the experiment has toolkits to leverage functionalities. Secondly, the significant objective is also to investigate the capabilities AR frameworks and Software Development Kits (SDK) for determining and selecting particular cell(s). Furthermore, the main focus of toolkits is to provide a simple way to overlay an image or animation over a video stream. Hence, a number of prototypes have been utilised in different AR toolkits, and ultimately the project has optimised an open source image-processing framework 'OpenCV'. The case study has experimented with the use of the Metaio AR SDK, Wikitude SDK and Total Immersion's D'Fusion SDK. Moreover, due to image-processing tools, an OpenCV is required because of its characteristics. As a result, the function of monitoring and selecting of this project have succeeded and meet the study's gaols. The first section, the paper will show the use of AR into education process and its advantages. The second section introduces the case study that reveals the mixture of AR and the novel smart-grid feature. The final section includes conclusion and recommendation.

II. AR in The Classroom

Augmented reality (AR) shifts the meaning of converting the virtual materials into genuine classroom environment by using spectacular technological techniques. In other words, it is the technology that schemes digital pedagogical materials onto real objects. The use of AR and its associated technologies in education permits both instructors and students to experience the virtual interaction in real time [4]. Real time learning, sharing of information and pedagogical interactions is a new concept that is in line with the level of civilisation in the world today. It is flexible system application that can be used and deployed already in available technological platforms and devices such as personal computers, laptops, tablets and smart phones. These devices are commonly used by students and educators alike. Moreover, they are easy to use and therefore suitable for suitable for students from various age groups and education level. Billinghurst (2002) pointed out that a smooth transmission between reality and virtuality will enhance and create a new experience. However, the integration into any system needs to be through an intermediate interface to allow for collaboration. Hence, many of AR and other tangible user interface system have been deployed and introduced in classes equipped specially for visualisation purpose. In contrast, in ordinarily equipped classrooms with no special instruments. Instructors are responsible for conducting lectures and activities by placing course books in digital form and using computers facilities. Learners follow instructor's guidelines while receiving prompts through slides shows, handouts or whiteboard. However, the equipped AR classrooms are relatively expensive making it difficult for their integration into the high collaborative environment that they must to be. Pre-existing teaching using both virtual and real time materials or re-planned activities in class is still fairly structured and limited for use with AR [5]. In the education sector, many educators attempt to teach and help their students to acquire the curriculum in an enthusiasm style. By using advanced interactive technology in education, the classrooms are likely to be the favourite place for teaching and learning process. Nowadays, Augmented reality has shaped a new interaction method for many sectors other than education. Consequently, a new interaction level has been rapidly increased by introducing AR. The new advanced technologies within the classroom facilities play a vital role in elevating the motivation levels amongst student while they are taught which is vital when it comes to attaining learning outcomes and other pedagogical outcomes objectives.

A. Advantages of AR to Education

With its three dimensions and real time interactivity capabilities, AR has key advantages over the traditional classroom which has seen it revolutionise the learning process. It uses some of the commonest technologies in today's world: computing and mobile telephony technologies. Most of the devices suing these technologies are very portable and therefore permits learning anywhere. That is, AR permits learning on the go. Such convenience enables students to learn with relative ease without the need to be confined within the walls of a traditional classroom setting. It allows for educators to disseminate content to the students from the remote location. It is an important cog in the wheel of e-learning whose popularity has increased with the advent computing technology and the internet, such as Remote-Lab. The educators and students have a platform where they can interact with the Remote-Lab in real time; doing the experiments.

Shelton and Hedley (2002) in their study of the impact of AR in the education sector found that there was a positive impact among undergraduate student using AR in geography lesson. The study targeted over thirty students and found that there was a significant improvement in student understanding when AR was used during the learning process. They also established that there was a reduction in student misunderstanding of the various lesson concepts [3]. This is because not only does AR improves the motivational levels amongst learners it also improves the creativity and innovativeness amongst learners. Moreover, AR superimposes and enhances information on the real world elements. Such contextual information which combines visual and sensory information results in improved cognitive skills and enhances associative learning [5]. AR provides learners with a friendly interactive interface which is rich in knowledge leading to high motivation. It is modern and most learners and educators in today's world can easily identify with it leading to high learning performance [5].

B. AR Based Visualisation and Interaction

Visualisation system opens new dimensions of human computer interaction by providing accessibility and user interface to all areas. AR provides interaction between the devices and users that leads to a decrease tangible interaction. On the other hand, in ordinary classroom, classrooms have limitations and allow less functionality in the workplace to be accomplished. However, technologies that provide and advanced technique, such as gesture and speech interaction, help our daily activities to interact with real world objects in more natural way [6]. Thus, the functionalities of reality-based-interface provide the next generation interface to interact simultaneously with real and augmented environment [6]. To address all obstacles in traditional classes or equipped classes, the augmented reality technique plays a crucial role in enriching and boosting learners to achieve their goals efficiently.

Visualising things and creation of a big loophole in understanding are caused by teaching situation and demands for many concept, such as physics, chemistry and biology [6]. AR tries to solve these issues by combining real and overlaying information and visualising scenes in a proper way to understand with the interaction function. Nevertheless, there is still need to improve the interaction method in a way that will provide the reality of interaction meaning with objects [7]. Supporting the AR technique needs to be proved by applications and tangible novel devices, such as smart overlay grid. Most of the applications do not support gesture-base direct manipulating of the augmented scene that is responsible for user interaction with objects for more real interaction efficiency [7]. Therefore, our study has come up with solutions that enable AR interaction which includes new methodology, deploying user's smart selection tool in both virtually and the possibility of physical interaction with objects that will be shown to the end-user.

According to Dunleavy (2009), a simulation study showed that technology-mediation within AR technique helps with interaction and collaboration within the highly engaging environment in which teachers and students operate. Although the AR simulation system provided prospective added value to the learning and teaching process, it showed some technical, manageability and cognitive challenges [8]. It is expected that AR technology will continue to progress to deliver high quality multimedia-interaction towards more powerful for mix-reality.

III. AR – *SmartGrid* for Educational Remote-Lab Monitoring (Case Study)

The case study utilises augmented reality (AR) elements to select and monitor captured of scenes for studying purposes, such as a video stream from an iPad camera, within the novel idea a smart grid layout. The process of the experiment that has taken is to develop application and investigate existing AR toolkits to leverage functionalities. Moreover, the significant objective is to investigate the capabilities AR frameworks and Software Development Kits (SDK). The main focus of toolkits is to provide a simple way to overlay an image or animation over a video stream. So, after a number of prototypes utilising different AR toolkits; ultimately the project has optimised an open source image-processing framework 'OpenCV'. The case study has been experimented with the use of the Metaio AR SDK, Wikitude SDK and Total Immersion's D'Fusion SDK. As a result, the monitoring functionality of this application required image-processing tools and because of that OpenCV seems to be sufficient due to its characteristics.

A. The Study's Rationale

In an educational context, there are many situations where AR can be applied. In studies by Montero et al. [10] AR technology was leveraged as a means to overcome the current limitations of Classroom Communication Systems (CCS), and thus allowing teachers to obtain a personal feedback from students in the class [10]. The communication system was utilised to allow participating students to signal whether they understand the material being presented to them by the lecturer in a non-intrusive manner. The flow of the lecture was not hindered and this method also overcame the issue where students are too shy to speak up or interrupt in the classroom [10]. Typically, a classroom communication system like this also allows the lecturer to gauge which students are not paying attention if they are not providing regular feedback. Another approach to determine attentiveness is to monitor the classroom for facial expressions and body language. This is a possible extension to the monitoring application, which would require advanced image processing and machine learning functionalities.

A monitoring application can also be applied to analysing data such as graphs and gauges. For example, a researcher is required to observe an experiment with graph and meter outputs from sensor data. He needs to record events where there are spikes in the output. The application can be employed to automate this process and all that is required of the researcher is some initial configuration.

B. Architecture and Design of the AR SmartGrid

In case-study, it has been chosen iPad running iOS 7.1.1 for testing environment for the iOS application. In addition, Xcode 5.1 which is the Integrated Development Environment (IDE) required developing on the iOS platform. The IDE is provided free from the App Store by Apple and comes with all the essential framework and libraries required to start developing applications. Having one central IDE makes it easy for third party frameworks, such as AR Software Development Kits (SDK), to import to iOS projects. Moreover, Objective-C, Objective-C++ and C++ have also been selected for programming languages that used to develop AR *SmartGrid* because Objective-C is the default language in the iOS framework. Objective-C++ combines C++ elements with Objective-C to extend the language allowing the use of C++ features. C++ has been utilised for the complex image processing functionality because of the better performance and also allows easy interfacing with the OpenCV SDK, which is written in C++[11].

OpenCV is an open source library for computational image processing written in optimized C/C++[12]. The library has a very active support community and extensive documentation. The source code is freely available on GitHub and supports Windows, Linux/Mac, Android and iOS.

Architectural Model of AR SmartGrid System

The solution that has been developed is a prototype application that utilises monitoring grid overlay functions. The AR application, called *SmartGrid*, has been built on the iOS platform for iPad devices and interfaces with a corresponding also from Pebble Smart Watch application. Figure 1 presents the high level architecture of the AR *SmartGrid* application. The Input Handler takes either a video stream feed or a website as the scene to be displayed. So, the Smart Grid provides an overlay of the smart Cells of the grid. If a Cell is selected, the Motion Detector monitors the Cell. The Motion Detector utilises the image processing functionalities of the OpenCV SDK. If motion is detected, the Cell generates an Event, which is displayed on the screen. Setting and event logs are persisted by the local application database.

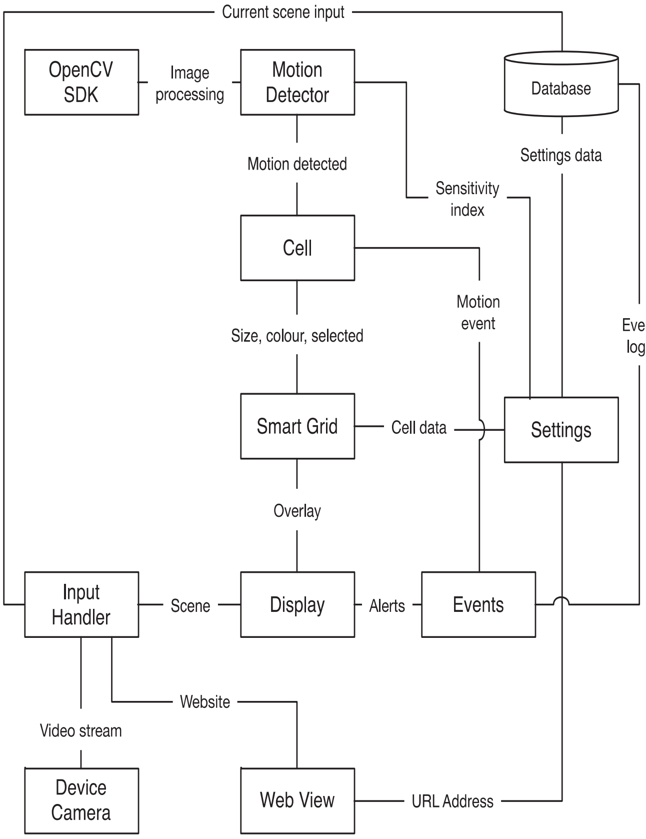


Fig. 1. High level architecture of AR Smart Grid application

C. Implementation of AR SmartGrid System

The functionality of the AR *SmartGrid* app is that it monitors selected cells for movements or deviations that are selected by students or instructor while they are in the session remotely. The monitored scene can either be configured for a video stream from the iPad device's camera or a web view of any website (i.e. UTS *Remote Labs* facility [13]). The grid for monitoring activities or events can be formed as free (Fig. 2) or fixed (Fig. 3 and Fig. 4). In a free grid placement mode, monitored cells are organised by dragging and dropping cells freely onto the canvas and in the fixed grid mode cells are laid out evenly according to a configurable number of rows and columns.

When an event or status change is detected in a selected cell by touching screen or through Pebble Smart Watch for selecting, an alert popover will overlay the screen. A log of the event, including a screenshot and timestamp, is registered.

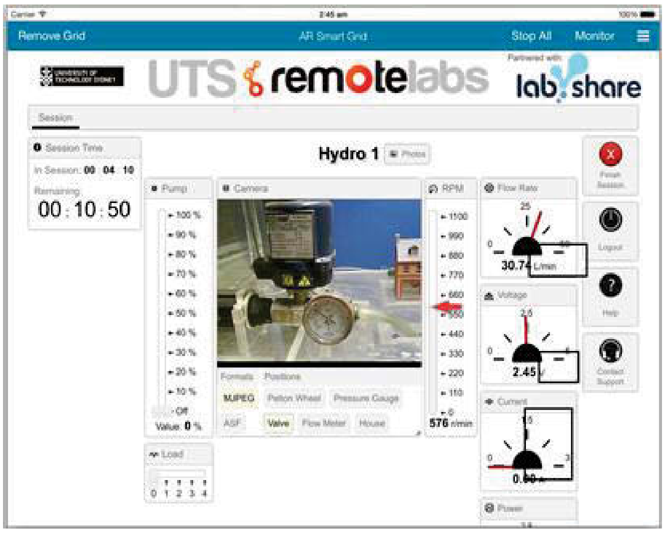


Fig. 2. SmartGrid's free grid mode by dragging and dropping cells running the UTS *Remote Lab's* Hydro experiment [13].

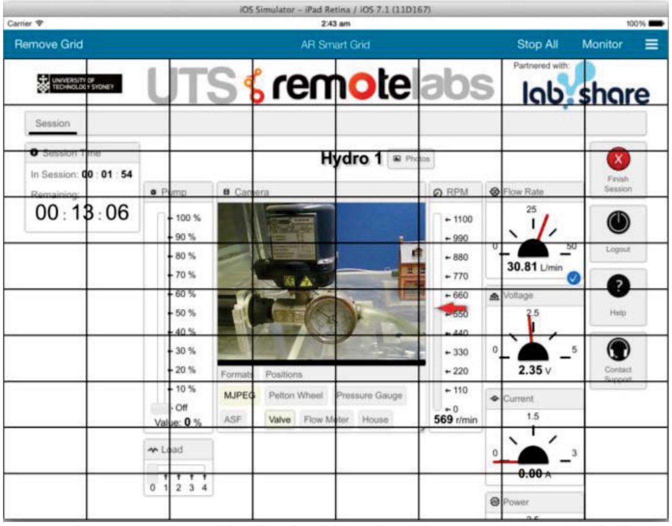
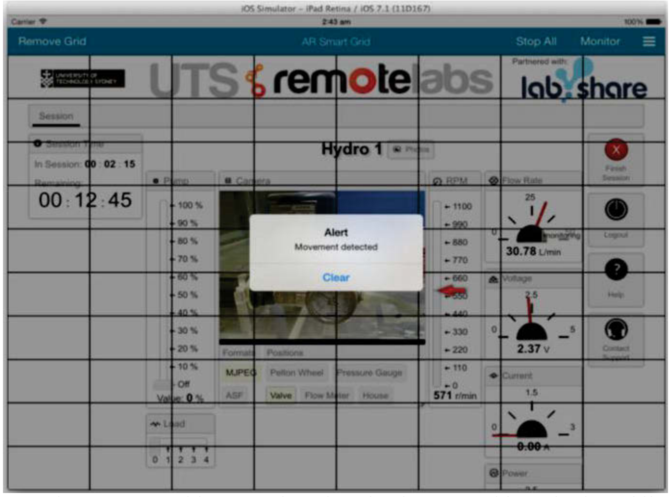


Fig. 3. SmartGrid's fixed grid mode with user positioned cells for threshold running the UTS *Remote Lab's* Hydro experiment [13].



Alert

Movement detected

Clear

Fig. 4. SmartGrid's event detection alert in smart grid monitoring while using the UTS *Remote Lab's* Hydro experiment [13].

CONCLUSION

Augmented Reality as an assistance tool has been increasingly integrated within the learning environment to support learning process and create the meaningful of interactive learning platform. The integration helps to improve knowledge acquisition and retention among learners. Through the advantages, AR can play a key role in raising motivation levels in the education environment by showing pedagogical materials in interactive and enthusiastic style.

For the enhancement of particular study room such as remote-lab, a case study has been taken place to adapt AR within overlay *SmartGrid* into remote-lab for monitoring and selecting scene of events. The development of this prototype application associated with the advantages of using AR-*SmartGrid* in remote-labs has demonstrated the possible to leverage existing technology for the purpose of monitoring and controlling. There is also opportunity to extend the monitoring capabilities of the application to include face detection, for example, and other features that could be possible to enhance the learning process for future demand. The use of AR technology in the context of education is still being explored, but has yielded promising results so far.

REFERENCES

[1] Z. Pan, A. D. Cheok, H. Yang, J. Zhu, and J. Shi, "Virtual reality and mixed reality for virtual learning environments," Computers & Graphics, vol. 30, pp. 20-28, 2006.

[2] Z. Chaczko and W. Alenazy, "The Extended Technology Acceptance Model and the Design of the 21st Century Classroom," in 2 nd Asia-Pacific Conference on Computer Aided System Engineering–APCASE 2014.

[3] R. Freitas and P. Campos, "SMART: a SysteM of Augmented Reality for Teaching 2 nd grade students," in Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction-Volume 2, 2008, pp. 27-30.

[4] S. Cuendet, Q. Bonnard, S. Do-Lenh, and P. Dillenbourg, "Designing augmented reality for the classroom," Computers & Education, vol. 68, pp. 557-569, 2013.

[5] D.-R. Chen, M.-Y. Chen, T.-C. Huang, and W.-P. Hsu, "Developing a Mobile Learning System in Augmented Reality Context," International Journal of Distributed Sensor Networks, vol. 2013, 2013.

[6] S. Prasad, S. K. Peddoju, and D. Ghosh, "Mobile augmented reality based interactive teaching & learning system with low computation approach," in Computational Intelligence in Control and Automation (CICA), 2013 IEEE Symposium on, 2013, pp. 97-103.

[7] W. H. Chun and T. Höllerer, "Real-time hand interaction for augmented reality on mobile phones," in Proceedings of the 2013 international conference on Intelligent user interfaces, 2013, pp. 307-314.

[8] M. Dunleavy, C. Dede, and R. Mitchell, "Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning," Journal of Science Education and Technology, vol. 18, pp. 7-22, 2009.

[9] C. Dede, "Immersive interfaces for engagement and learning," science, vol. 323, pp. 66-69, 2009.

[10] A. Montero, T. Zarraonandia, I. Aedo, and P. Díaz, "Uses of Augmented Reality for Supporting Educational Presentations," in Advanced Learning Technologies (ICALT), 2013 IEEE 13th International Conference on, 2013, pp. 426-428.

[11] P. Sholtz, "How To Make An Augmented Reality Target Shooter Game With OpenCV: Part 3/4," in RAYWENDERLICH TUTORIALS FOR DEVELOPERS & GAMERS, ed, 2014.

[12] OpenCV. (2014). Open Source Computer Vision. Available: http://www.opencv.org.

[13] Remote Lab at UTS, http://www.uts.edu.au/about/faculty-engineering-and-information-technology/what-we-do/facilities-and-services/remote.