MOH/P/PAK/337.17(BK)

MALAYSIAN HEALTH DATA WAREHOUSE (MyHDW)

2015-2016 START UP: INITIATION

MINISTRY OF HEALTH MALAYSIA

HEALTH INFORMATICS CENTRE, PLANNING DIVISION
PUSAT INFORMATIK KESIHATAN, BAHAGIAN PERANCANGAN

MALAYSIAN HEALTH DATA WAREHOUSE (MyHDW)

2015-2016 START UP: INITIATION

EDITORS:
MARK J FULLER
DR MD KHADZIR SHEIKH AHMAD


TUE
18

TABLE OF CONTENTS

MINISTRY OF HEALTH MALAYSIA

HEALTH INFORMATICS CENTRE
PLANNING DIVISION
MINISTRY OF HEALTH MALAYSIA

ISBN : 978-967-0769-76-9

COPYRIGHT @ MINISTRY OF HEALTH MALAYSIA 2017

PRINTED BY:

GRAPHIC ZONE SDN BHD
NO. 48, JALAN PBS 14/11
TAMAN PERINDUSTRIAN BUKIT SERDANG
43300 SERI KEMBANGAN
SELANGOR DARUL EHSAN
TEL : 03 8943 3855
FAX : 03 8957 5637

FOREWORD BY
SECRETARY GENERAL, MINISTRY OF HEALTH MALAYSIA

The delivery of Malaysian Health Data Warehouse (MyHDW) Project was significant since it was developed based on MIMOS platforms and solutions. That was an aspiration I conveyed to the team when I was first introduced to Mr Mark Fuller, a consultant from Canada, and when they first started with the project.

This is the first Information and Communications Technology (ICT) project in the Ministry of Health (MoH) that is delivered on time. The project also brings other bonuses including saving on the project and future cost saving for Operation and Maintenance.

This project provides an opportunity for MoH to streamline all the data generated and has initiated progression towards Big Data Analytics (BDA) adoption. I believe we will be the first Ministry to clock this achievement. This project has also significantly established and documented successful working collaboration with MIMOS Berhad and will pave the way for many more collaborative engagements.

I would like to acknowledge and congratulate all who were involved in the project and would like to wish you success on the next phases of development.

DATO' SERI DR. CHEN CHAW MIN
SECRETARY GENERAL
MINISTRY OF HEALTH, MALAYSIA

FOREWORD BY
DIRECTOR GENERAL, MINISTRY OF HEALTH MALAYSIA

The Malaysian Health Data Warehouse (MyHDW) project started in 2010. Funding for the project only came in 2015. The gap of two years allowed the Ministry of Health (MoH) to evaluate the selection of technology. The concern had always been on the security layers provided by the technology and the long-term financial obligation to operate and maintain the project.

This is the first time a National Health Data Warehouse is developed based on more than 90 percent local technology. It is also the first time an Information and Communications Technology (ICT) project in MoH is delivered on schedule, a double celebration for all team players.

MoH is now in the second phase of MyHDW project. Some of the focus is to integrate Geographic Information System (GIS) and MyHarmony into the project. MyHarmony is the solution developed in silo to cater to the management of unstructured data using medical terminology standard, which is based on SNOMED CT. The implementation of this project thus prepares MoH towards Big Data Analytics (BDA), a solution towards making more informed decisions in our planning.

I would like to congratulate everyone involved, from the initiation all the way to the delivery of Phase 1 of the MyHDW. I now look forward to the completion of Phase 2 of this project, and MoH becoming the first ministry to have BDA in place.

My sincere thanks to the team for their commitment and contribution towards the project. I wish them all the best in preparing the BDA, which I am positive will be another reality in the future.

DATUK DR. NOOR HISHAM B. ABDULLAH
DIRECTOR GENERAL
MINISTRY OF HEALTH MALAYSIA

FOREWORD BY

DEPUTY DIRECTOR GENERAL (RESEARCH AND TECHNICAL SUPPORT), MINISTRY OF HEALTH MALAYSIA

I was directly involved in the project as Co-chair of the Technical Committee together with representative from MIMOS. It was very important that the project was delivered based on proper methodology. All development and progress of the project was presented to the Technical Committee including carving out the sub-committee work on Security and Testing from the project implementation team activities. This was to provide an independent sub-committee for that purpose.

The Business Intelligence (BI) tools was the new platform and solution developed for the project. To address any reservation on the new tools, the project team introduced Option Analytic Report when selecting any technology and this was applied when selecting the BI tools. Finally, Mi-BIS, the BI tool from MIMOS was adopted.

As MyHDW is to be a trusted source of information, all data generated by the system underwent testing to validate the findings. A group comprising statisticians and medical doctors was appointed to test the validity of the information generated by the system. This was to ensure that the system generated what it should be generating.

The project places emphasis on security and privacy, which is important since we are dealing with health data at the granular level. The Data and Information Governance Committee was established to address the issue of security and privacy. This work was huge but the project team has been pragmatic in the approach, where they focused on immediate delivery, which was access to the data and information.

I would like to congratulate everyone who has been involved in the project. It has not been easy for the team to endure all the pressure and to shoulder this huge responsibility. I would like to thank the team, including MIMOS, for their commitment and contribution towards the project. I look forward to working on the Big Data Analytics (BDA) in the second phase of this project.

DATUK DR. SHAHNAZ BINTI MURAD
DEPUTY DIRECTOR GENERAL (RESEARCH AND TECHNICAL SUPPORT)
MINISTRY OF HEALTH MALAYSIA

INTRODUCTION

The Malaysian Health Data Warehouse (MyHDW) blueprint was proposed and accepted in 2011. MyHDW is a trusted source of truth of comprehensive healthcare data structured for query and analysis purposes. More than 90 percent of the MyHDW technology is homegrown, and developed by MIMOS Berhad.

MyHDW has four dimensions and, it is a collection of "Secondary Data Use" from visit to healthcare facilities and services:

The first dimension for "Secondary Data Use" refers to the collection and the use of granular data for purposes other than direct patient care. The data will be collected from all facilities and services that belong to the Ministry of Health (MoH), private healthcare facilities, university hospitals and army hospitals. The healthcare facilities include hospitals, clinics, day care centres, and traditional and complementary medicine centres.

The second dimension of MyHDW is collection of visit data by a person or patient to all services provided in the facilities. Within the hospital, there are services provided via Inpatient Care, Day Care, and Outpatient Care; which comprises Acute and Emergency, General Outpatient, Specialist Outpatient and Patient Assessment Centre. There are also Clinical Support Services which include Physiotherapy, Occupational Therapy, Speech Therapy, Audiology, Diet, Social Support, Laboratory, Diagnostic and Imaging, Forensic, and Nuclear Medicine.

The third dimension of MyHDW is the data collected in granular forms via *Sistem Maklumat Rawatan Pesakit* (SMRP). Data for a very specific disease can be futher collected by a system called Patient Registry Information System (PRIS). There is dependency between these two systems to minimise duplicate data entry and to preserve data integrity. The MyHDW project takes into consideration all requirements in SMRP and in PRIS and harmonises the distribution of the variables.

The fourth dimension in MyHDW is the nature of data collection which is both a Structured and an Unstructured Data. The Unstructured Data will be codified with Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT) using MyHarmony, which is another solution developed together with MIMOS Berhad.

The Data and Information Governance Committee was established to manage the security and privacy process in MyHDW. The committee addresses User Access Control, Data Ownership, Privacy, Use of Data, and Data Quality.

Security system of MyHDW was developed using local technology and is owned by the government. Security is one of the important components of MyHDW, which is to protect health data and information. The technology and security includes Multifactor Authentication and application of standard and guideline provided by Malaysian Administrative Modernisation and Management Planning Unit (MAMPU) and International Standards Organization (ISO). MyHDW subscribes to the Personal Data Protection Act (PDPA) 2010.

MyHDW has six main functions:

1. **Fixed Format Report (FFR)** – pre-defined format for data that has been submitted by the source system.

2. **Dashboard** – visual representation in the form of chart or graphs. This is to cater for Key Performance Indicator (KPI) or Quality Assurance Program (QAP).

3. **Ad-Hoc Query** – allows generation of report not catered in FFR or Dashboard. MyHDW will prepare data marts to allow user with access to conduct analysis of data within MyHDW environment. This will protect integrity of data, security and the system will provide all the relevant analytic tools.

4. **Statistics** – allows analyses of current and historical data to make inference and predictions about the future.

5. **MyHarmony** – harmonises unstructured data together with SNOMED CT for analysis.

6. **Geographic Information System (GIS)** – provides the layer of information in Malaysian map.

MyHDW will enable further development towards Big Data Analytics (BDA). MyHDW provides an opportunity for Malaysia to showcase an application based on local government technology, and this is the first Information and Communication Technology (ICT) project in KKM that is delivered on time.

DATUK DR. ROHAIZAT BIN YON
DIRECTOR, PLANNING DIVISION,
MINISTRY OF HEALTH MALAYSIA

MyHDW Journey

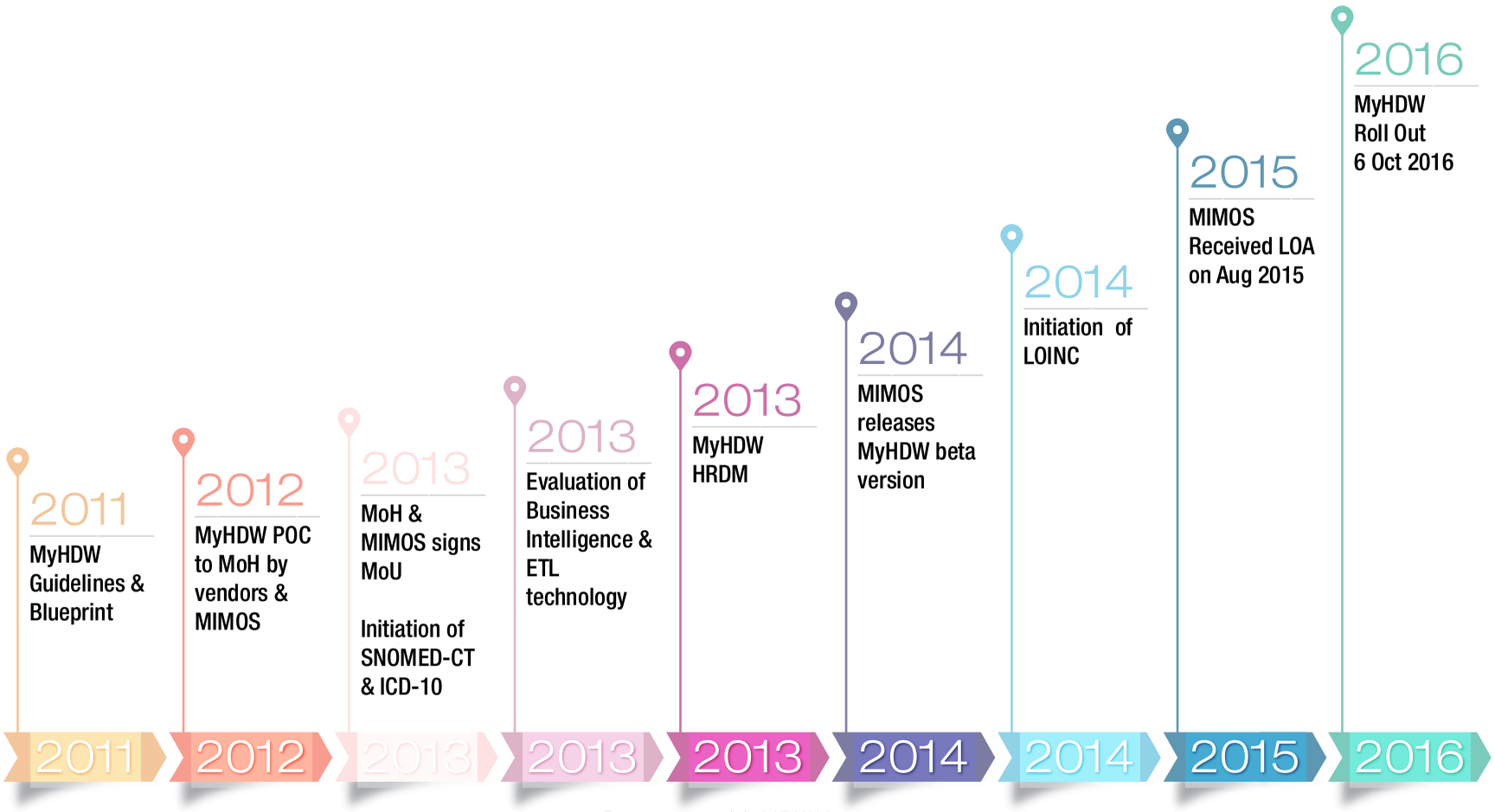


Diagram 1: *MyHDW Journey*

MALAYSIAN HEALTH DATA WAREHOUSE (MyHDW)

2015-2016 START UP: INITIATION

OBJECTIVE

Objective

The objective of this document is to describe the key steps, artefacts, approaches, and techniques used in the development of the Start-Up: Initiation phase of MyHDW. Thus, acting as reference material, and as a record of lessons learned from this implementation, to assist in the future progression of the project and similar initiatives.

MALAYSIAN HEALTH DATA WAREHOUSE (MyHDW)

2015-2016 START UP: INITIATION

BACKGROUND

Background

MALAYSIAN HEALTH DATA WAREHOUSE START-UP: MyHDW-SMRP-PRIS

The Ministry of Health, Malaysia (Kementerian Kesihatan Malaysia, KKM) began the implementation of the Start-up phase of Malaysian Health Data Warehouse MyHDW-SMRP-PRIS in August 2015. This initiative builds on previous work as outlined in the MyHDW Blueprint *(Malaysian Health Data Warehouse (MyHDW) 2011-2013)*[1]. MyHDW-SMRP-PRIS is composed of three sub-projects; Malaysian Health Data Warehouse (MyHDW) Phase 1, Sistem Maklumat Rawatan Pesakit (SMRP) and Patient Registry Information System (PRIS) projects. Its overarching objective is the establishment of a comprehensive national Health Analytics system to service a wide range of health information requirements. The collection of systems associated with this start-up phase is named MyHDW 1.0

MyHDW 1.0 establishes the basic infrastructure and capabilities for this national analytic system, and has been designed to scale and expand to include and integrate additional datasets over a number of phases. The core analytic functionality associated with the system pertains to secondary usage capability, in which data sometimes collected for clinical purposes can be utilised in a secure and privacy sensitive manner after careful anonymization and de-identification. At this stage, MyHDW is not envisaged to support direct clinical functionality. The initial source data for MyHDW 1.0 is associated with a number of acute services and registries; specifically, Inpatient, Daycare and Cancer registry data. To facilitate this collection of the data above, Sistem Maklumat Rawatan Pesakit (SMRP) is being utilised. As part of the current phase, this system (SMRP1.1) is being upgraded. The new version (SMRP 2.0) will support additional data and allow its more comprehensive and timely integration into MyHDW. Furthermore, future phases will utilise this collection platform for additional service type and registry data. SMRP 2.0 will also extend the completeness of datasets to include University, Army and Private hospitals. Updates to registry data are also occurring in MyHDW 1.0. Patient Registry Information System (PRIS 1.0) is a generalised registry collection system designed to extract data from SMRP for utilization in the registry components of MyHDW. During the current phase in addition to this generalised capability, Cancer visit data is added as an initial module. In addition, functionality to automate notification of initial visit of confirmed Cancer cases will be provisioned.

MyHDW 1.0 incorporates traditional Data Warehousing functionality associated with enterprise (health system) wide structured data integrated across the Continuum of Care and other common healthcare subject areas such as Population, Health Human Resources and Costs. Standardise reports and dash-boarding functionality will be included in the initial phase, along with Ad-hoc and Statistical capabilities and dataset export. Moreover, the infrastructure has been designed to allow expansion into Big Data Analytics (BDA) techniques associated with semi and unstructured data and next-generation query tools. In addition to a number of locally developed platform technologies from MIMOS Berhad, functionality will also be included to support Geographic Information System (GIS) analytics. Furthermore, MyHarmony, a platform also developed by

1 MyHDW Blueprint is available online. See: at http://www.moh.gov.my/images/gallery/publications/myhdw2011-2013.pdf

MIMOS Berhad, which utilises Natural Language Processing and SNOMED CT RefSets to provide harmonisation and codification services, will be implemented in the MyHDW 1.0 environment. Census, Geographic and possibly other datasets will also be added and integrated into MyHDW 1.0 to support a broader range of analytic options.

It should be noted that Phase 2 of MyHDW plans to further expand SMRP's capabilities during the second year of this initiative to include visit data associated with; Clinical Support Services (Physiotherapy, Occupational Therapy, Speech Therapy, Audiology, Social Services, Diet, Laboratory, Imaging, Forensic Medicine and Nuclear Medicine); Outpatient Visit (General Outpatient, Specialist Clinic and Acute and Emergency); visit to Procedure Room and visit to Traditional and Complementary Medicine Facilities.

As part of the project inception activities, a project governance and structure was established to recognize legislative constraints particularly as relates to, privacy and security as well as normal project controls and accountabilities. Regarding the former, the Personal Data Protection Act (PDPA) 2010[2], Private Healthcare Facilities and Services Act, 1998[3] and Regulations were considered. MAMPU policy associated with privacy, security and Data Quality were also factored into project governance. During the period 2014/2015, Malaysian government directives required where feasible, investment into local infrastructure and technology as an alternative to commercially available solutions and this factor was incorporated into the planning considerations of this project.

Given its track record of ITC innovations including Big Data Analytics, IT Product Research and Development, security expertise and the above-mentioned directions, MIMOS Berhad was engaged by the Ministry of Health (KKM), to perform the technical development work associated with this initiative in a start-up or incubator capacity. MIMOS Berhad was also selected as the initial data centre for the project. Originally it had been planned that this would be located and provisioned within the federal Data Centre, Pusat Data Sektor Awam (PDSA). In subsequent discussions with MAMPU, it was agreed upon that it would be more optimal to initially provision these services within the MIMOS Data Centre. It is also envisaged that as part of facilitating the long-term handover to KKM a number of capacity building activities should occur. These include; skills development in terms of analysis and statistics, increased allocation and recruitment of additional Data Analysts including Statisticians and the recognition of the new role of Data Scientist. Furthermore, training on the basic use of the MyHDW 1.0 product will be done utilizing a Train-the-Trainer approach and incorporated into roll-out and deployment planning.

2 The Malaysian Personal Data Protection Act 2010 ("the Act") which came into force on 15 November 2013 applies to any person who processes and has control over or authorizes the processing of any "personal data" in respect of commercial transactions ("data user"). See : http://www.pdp.gov.my/images/LAWS_OF_MALAYSIA_PDPA.pdf

3 The act was implemented with effect from 1st May, 2006 and regulates all private healthcare facilities and services. See: http://www.mma.org.my/images/pdfs/Link-LawOfMsiaAct/private-healthcare-facilities-and-services-act-1998.pdf

BLUEPRINT: MALAYSIAN HEALTH DATA WAREHOUSE (MyHDW) 2011 TO 2013

In terms of historical perspectives, the original roadmap and blueprint for this work has been documented as mentioned, in the Malaysian Health Data Warehouse (MyHDW) 2011 to 2013 report. This details a series of national and International consultations and workshops during this period and includes:

- Guidelines and Blueprint
- Portfolio Plan
- Infrastructure, Resources and Initial Deliverables
- Project Initiation and Data and Information Architecture
- Health Reference Data Model
- Technology Selection and General Observations

This report formed the basic portfolio plan and guidelines for MyHDW 1.0 and was compiled jointly by KKM, advised by an international consultant.

LEARNING FROM OTHERS

One strategy established early on was to learn from other similar initiatives, both nationally and internationally. Thereby, fast tracking work and reducing risk. A consultant was assigned to the project, with successful national level experience in building a similar system and used as an adviser and for quality assurance purposes. In addition to contributing to the initial blueprint work, this consultant also provided ongoing advisory and quality assurance services to MyHDW 1.0 during the start-up phase. The underpinning for much of the commentary and narrative in this report has been drawn from that work and its associated reports.

The overarching objective of this work, was to provide guidance, expertise and product quality assurance reviews to MIMOS Berhad and KKM staff to ensure the successful delivery of the project and the setting up of the foundations for future phases. The focus of the consultancy activities which occurred September 2015 to August 2016 included:

- Ensuring the quality of key deliverables.
- Advice on strategies associated with Critical Success Factors (CSF) including: Project Sponsorship Approaches, Data and Information Governance, Data Quality Strategy, Project and Program Management approaches, and optimum design for Healthcare Data Warehousing Architecture.
- Provision of program overarching guidance and advice: Including tactical recommendations, course corrections and sustainability considerations.
- Review and provision of recommendations on DW/BI methodologies.

- Providing high level recommendations on the augmentation of Big Data Analysis technology and approaches within the MyHDW architecture.
- Providing high level technical design guidance on SNOMED CT implementation in MyHDW.
- Providing high level recommendations and oversight for privacy and security considerations.

THE HEALTHCARE DATA WAREHOUSE - HEALTH ANALYTICS ENVIRONMENT

The ability to have timely access to trusted healthcare data and information is necessary to the effective running of the health system as well as being of benefit and necessary for a wide range of clinical, operational and research requirements. In an age where health transformation, sustainment, evidence based decision-making and expectations of a responsive health system are ever present; this fact becomes even more germane. Furthermore, other uses of health related information to support a wide range of areas such as Surveillance, Research and Program Management also necessitate the availability of high-quality data and information, and often the appropriate tools to make use of these.

The Healthcare Data Warehouse is now considered an increasingly necessary underpinning for any Health Analytics System that provisions this type of information. In summary, some of the benefits and drivers of the need for this type of system and its continual evolution are the requirements for:

- Insights about Clinical and Operational performance.
- Using data to drive and sustain change: Healthcare transformation through high quality and timely analytics.
- Patient Safety.
- Sustainability of the Health System.
- Operational reporting, monitoring and alerting.
- Clinical Research – Insights into clinical practice and process.

It has also become apparent over the last decade that to fully realise its value, the Healthcare Data Warehouse in addition to having strong internal data linkage and integration must be considered in the overall design of any Digital Health strategy, including the Electronic Health or Medical Record. Also, supplemental data such as Census and Geographic. This assumes the notion that primary usage data gathered during the operational use of Electronic Health Records is utilised for secondary purposes in the Healthcare Data Warehouse. This further promotes the notion of data capture at every opportunity within the clinical and administrative workflow. Concerning Digital Health generally it could be argued that the Healthcare Data Warehouse is:

- Critical to managing a Healthcare System or an organization's healthcare delivery.
- To be planned as part of a broader Digital Health strategy informed and guided by a Maturity Model.

- A key vehicle for data linkage & integration and thus the associated benefits of a broader range of health information and contexts.
- An underpinning of Self Service approaches including Open Data. This allows appropriate data, information and insight access to the greatest numbers of people and thus increases exponentially potential return on investment and use of data.

Recent years have seen the maturing and advancement of Data Warehousing/Business Intelligence environments, sometimes known as DW/BI to embrace a plethora of new and exciting approaches under the rubric of Big Data Analytics. Big Data Analytics (BDA) while not completely replacing the high precision analytics of commonly deployed DW/BI environments in Relational Database Management Systems (RDBMS), allows significant opportunities to gain insights from data in a manner that was previously difficult and is a paradigm shift in how data might be used for analytical purposes. For example, the analysis of semi-structured, unstructured and streaming data is increasingly becoming a reality, similarly while previously available, Predictive Analytics and similar Data Mining tools have matured further as have our abilities to access data in a highly time sensitive manner. This list only represents a sampling of the possibilities. That said, current industry trends still suggests that the traditional structured DW/BI environment is in many instances still a core requirement for information systems of this type but that now should be augmented by BDA techniques. These have the potential to significantly reduce costs, increase timeliness and allow greater flexibility.

Health Analytics[4] is a term commonly being used to encompass this environment; it's approaches, technology and utilisation in a health context. The field of Data Science and Analytics, including methodologies and architectures associated with designing and building analytic environments is in a current state of rapid evolution. It is assumed that the domain shifts associated with Big Data Analytics as an augmentation to BI/DW is an expected part of these systems. As are newer technologies such as Cognitive Computing and Artificial Intelligence. *See moving-forward: MyHDW Phase 2 and Beyond* section below for further details of BDA approaches, technologies and how these will augment BI/DW structured approaches in MyHDW. In addition, emerging trends such as Precision Medicine and the Internet of Things will be outlined.

4 The systematic use of data, information technology and methods to create insights in context that inform clinical and business decision making around the planning, delivery, management and measurement of health care. Source: HIMSS Clinical and Business Intelligence Community of Interest, 2013

MALAYSIAN HEALTH DATA WAREHOUSE (MyHDW)

2015-2016 START UP: INITIATION

PROJECT MANAGEMENT

Project Management

MyHDW 1.0 (MyHDW-SMRP-PRIS) represents three subprojects. MyHDW Phase 1, which is the start-up phase of the analytic portion of the Health Data Warehouse. SMRP 2.0 which as mentioned above will provide improved data collection services, initially for Inpatient and Daycare. PRIS 1.0 which form the underpinning for a registry collection system. This later system will begin by capturing Cancer visits and leverage is the SMRP 2.0 system itself. In this way, analytical capability is enhanced both by the creation of a Health Data Warehouse environment and improvements to the data collection aspects of this.

The project was approved in early 2014 and subsequently received approval from JTICT MAMPU[5] followed by the Ministry of Finance. This allowed for further approvals by the Tender Board of the Ministry of Health to directly negotiate with MIMOS Berhad who were awarded the project. Final approval was given through a Letter of Acceptance in August 2015. While some start-up activities were already in place, the project formally kicked off at this point. Project Management for the MyHDW-SMRP-PRIS project was administered through the MIMOS Berhad Project management office (PMO) and project managers were assigned. MIMOS Berhad has a standardised approach to Project Management Life-Cycle (PMLC) and System Development Life-Cycle (SDLC), in line with these a Project Management Plan (PMP), sometimes known as a Project Charter or Project Initiation Document produced in September 2015 by the Project Manager assigned to the project.

MyHDW-SMRP-PRIS overall covers a period of 2 years with an additional year as a warranty period. It should be noted that the BI/DW component MyHDW phase 1, is of a single year duration. In terms of project phases associated with the sub-components, these are as follows:

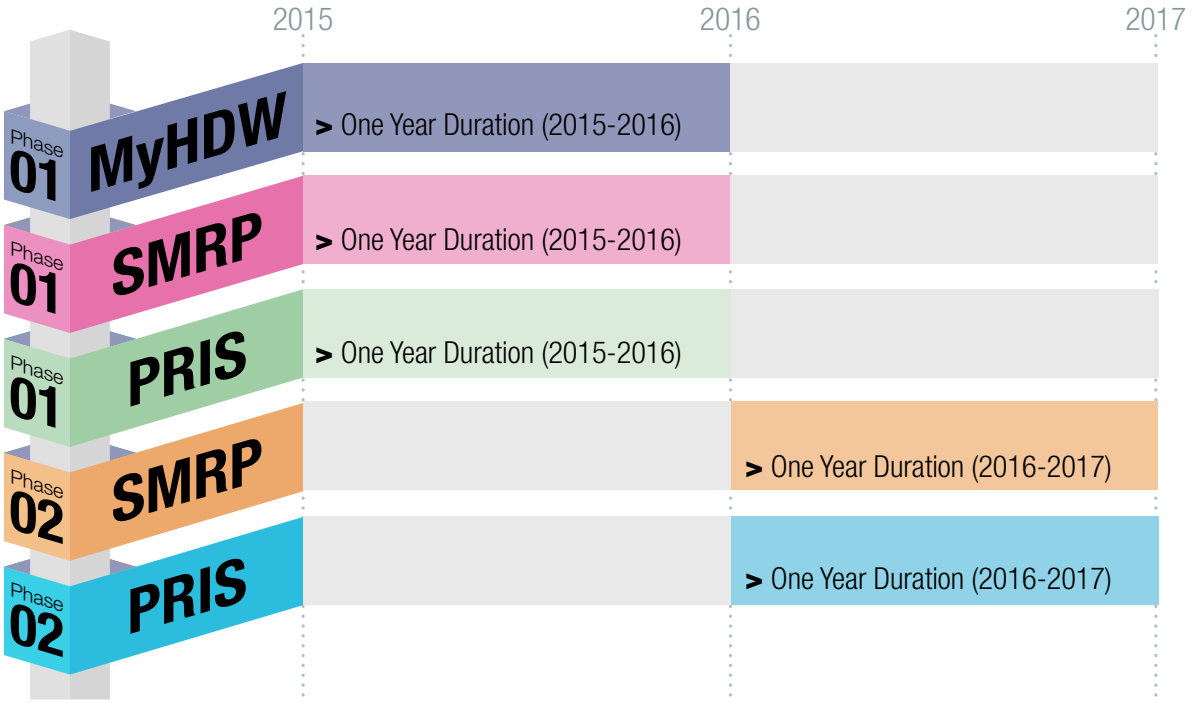


Diagram 2: *MyHDW Project Timeline*

5 Jawatankuasa Teknikal ICT (JTICT) Malaysian Administrative Modernisation and Management Planning Unit (MAMPU)

As mentioned above, the term MyHDW 1.0 is also used as a general name for this collection of projects. In the early planning phases, such as contractual material it was also named MyHDW Phase 1 and Data Source Project. The total budget for MyHDW covers all capital and hardware expenditure, licensing and professional services. It should be noted that a good deal of the hardware and license costs, while mostly allocated to the MyHDW subproject itself are shared across all 3 environments and established the base server infrastructure of the initiative.

A Phase 2 for MyHDW (Health Data Warehouse) sub-component was not included in this project plan and associated budgets. At the time of writing, this is being specified and approval discussions are underway. This will execute in 2016/2017.

As part of the consultant quality review process, the PMP document produced in September 2015 was reviewed and feedback shared with the project team. The PMP was based on a 'waterfall'[6] style project management approach as was the accompanying SDLC. The project timelines could be characterized as accelerated. It should also be recognised that the host organisation, while experienced in large ITC initiatives, had not developed something of this scale associated with BI/DW before. As such, the PMP document required some refinement to ensure it could guide the successful execution of the project. Over the subsequent month, changes were made to the document to allow for this. Following discussions at this time, capacity and skill gaps were identified and over the same period additional resources were allocated to the project or recruited as necessary in line with the updated PMP.

LESSONS LEARNED – PROJECT MANAGEMENT

Bearing in mind the points raised above, the following are the lessons learned associated with the review process of the Project Management of MyHDW 1.0:

- Project management lifecycle and project documentation needs to be tailored to BI/DW/BDA style initiatives to allow successful delivery.
- Scope definitions should be done with the necessary level of precision.
- Resource costs should be broken out at a level of detail to be implementable.
- Critical Success Factors (CSF) need to be articulated and clearly agreed upon by executive sponsors of the project.
- Risk management plans need to be precisely defined.
- Future project plans should recognise and include a Design Phase as part of the SDLC.
- In planning BI/DW projects, it is important that source data stability and availability is considered. An ongoing challenge during MyHDW 1.0 was that data from SMRP 2.0 and PRIS 1.0 was also needed to load new data warehouse and for associated reporting. This interdependency caused complexities throughout the project.

6 Waterfall model is a sequential design process, used in software development processes, in which progress is seen as flowing steadily downwards (like a waterfall) through the phases of conception, initiation, analysis, design, construction, testing, production/implementation and maintenance. (https://en.wikipedia.org/wiki/Agile_software_development)

- During the execution of project plans, it is important to track and monitor issues, risks and the actual utilization of resources associated with the execution of the plan.
- Roll-out, Deployment and Release Planning needs to be considered in project plans. There is a good deal of complexity in these activities, and this needs to be accounted for in terms of resources, cost and timelines.
- As mentioned, the current style of project management was based on a 'waterfall' approach. Current industry practice is finding that Agile[7] project management to be a potentially superior approach and as such it may be worth investigating the feasibility of using this in future initiatives. Iterative project management methodologies such as Rational Unified Process (RUP) or a gated project lifecycle may also be a good alternative, and could act as a stepping stone to Agile techniques.
- While not part of the project plan per se, sustainment activities such as long-term capacity development need to be included into costing and resource considerations, feasibly as part of a broader program initiative.
- Project Management accountability needs to be clear for large-scale initiatives such as this. While it is common for project managers to report to a central PMO and be allocated to projects in a matrixed fashion, consideration should be given as to whether this type of resource reports to the project or some other mechanism to ensure clearer accountability for future initiatives. Similar consideration should also be given to other matrix style functions within the host organisation such as Architects, Technical Leads, Testing, and Security.

While a number of lessons learned have been recognised, it should also be noted that in most instances timely remediation was instigated once issues were identified. The project is essentially on track and delivered on time and on budget.

7 Agile software development is a group of software development methods in which solutions evolve through collaboration between self-organizing,[1] cross-functional teams. It promotes adaptive planning, evolutionary development, early delivery, continuous improvement, and encourages rapid and flexible response to change. (https://en.wikipedia.org/wiki/Waterfall_model)

MALAYSIAN HEALTH DATA WAREHOUSE (MyHDW)

2015-2016 START UP: INITIATION

CAPACITY BUILDING ESTABLISHING A TEAM AND CONDITIONS FOR SUCCESS

Capacity Building: Establishing a Team and Conditions for Success

As identified in the original MyHDW Blueprint Portfolio Plan, capacity, skills, and capabilities, both in terms of technical and analytical, are required for the implementation and sustainment of MyHDW. In addition, infrastructure capability is needed in terms of a high quality, and secure data centre(s). Process and organisational capacity, and maturity is also an important factor in terms of promoting conditions for a successful implementation. One aspect of the original MyHDW Blueprint work was to understand clearly local conditions, capabilities, opportunities for partnerships, vendor landscape, and organisational development requirements. In summary, the following are the key components associated with capacity building:

- Selecting Host IT Development and Sustainment organisation.
- IT Development team skills, and team structure.
- Domain/Business team composition and roles associated with data analysis, data management and information product development. Including promoting national analytical capabilities – PIK[8], Performance Management Team.
- IT Sustainment and Operations team.
- Organisational structure and design.
- IT Processes.

IT TEAM - SKILLS/ TEAM STRUCTURE

Core IT roles for this type of BI/DW development were previously identified in the MyHDW blueprint. These are common to most IT application development projects, though have a stronger data centric focus, and in addition include specialised tool developer associated with BI and Data Integration tools (Extract Transformation and Load - ETL). These include:

- Project Manager (PM)
- Business Analyst (BA)
- Technical Lead or Project Lead
- Solution Architect
- Data Architect/Data Modeler
- Data Base Administrator (DBA)
- Systems/Network Administrator
- Software and Data Testing Lead

8 *Pusat Informatik Kesihatan (PIK)*/ Health Informatics Centre (HIC) is a unit under the Planning Division of the Ministry of Health (MoH) Malaysia.

- Quality Assurance (QA) Lead
- ETL Developer
- BI Developer
- Application/Solutions Developer-web/Java
- Technical Writer
- Information Security Lead
- Project Coordinator

BUSINESS (DOMAIN) TEAM STRUCTURE

Business skills and resource capacity development is also an important part of ensuring the successful implementation of MyHDW, and the effective use of its data and information. The MyHDW Blueprint also identifies core roles, associated with domain or business functionality. Within the Ministry of Health (KKM), Pusat Informatik Kesihatan (HIC) and Malaysians Health Performance Unit, both have roles to play in the development and sustainment of data and analytic activity. PIK is responsible for data management associated with data collection systems used for analysis purposes. In addition, its role encompasses Health Informatics Standard. Malaysians Health Performance Unit is responsible for information product development such as Health Indicators, Benchmarks, Key Performance Indicators et al. Both these organisational areas have a key role in the development, sustainment, data quality, security and the use of information associated with MyHDW. It is important that both groups are staff sufficiently to operate in an effective manner. The implementation of MyHDW will almost definitely increase the need for resources in these areas, and a sustainment plan should be developed to define these increasing requirements. Also, given the overlap in nature of their roles and based on international examples, they will need to closely work together to ensure high quality and usable information and reports are harvested from MyHDW. Major roles within these groups that require ongoing development and sustainment include:

- Senior Domain Specialist
- Health Informatician
- Health Informatics Standard Specialist
- Data Analyst
- Statistician
- Data Scientist
- Medical Record Officer
- Privacy Officer
- Trainer

SKILLS/TEAM STRUCTURE IT/BUSINESS - OBSERVATIONS

For the development and implementation aspects of MyHDW 1.0, a wide range of IT roles were assigned to the project team. Overall, after some initial fine-tuning, there was a good level of resource available to execute project work. Furthermore, based on the skills information gathered during the review process, on average initial team experience level was adequate. A few team members had substantive experience. Two roles were highlighted as being critical for success in initiatives such as this, namely, Project Manager and Business Analyst. Concerning the Project Manager role, it is mostly considered good practice to assign Technical Project Managers to this type of initiative, ideally those with large project BI/DW experience. This is a particularly critical role and should be recruited carefully, and monitored to ensure good fit. Similarly, the BA role specifically Senior BA is another key role, and traditionally would be responsible for leading all aspects of requirements specification. This latter point allows for example, the Project/Technical Lead to focus on solutions design and architecture. That said, it should be also noted that in some organisations the Technical Lead is involved in requirements gathering. There is some debate concerning best practice in this area. Furthermore, to support project work other than monitoring the two roles mentioned, it is important that project documentation and artefacts are of a good quality, organised and available. Future projects should monitor this QA/Project coordination related activity.

The following table outlines approximate peak allocations during Phase 1/2015-2016 activities, established on the initial project plan.

The following are some sample values from the PMP:

System	Name	FTE[9] Count
MyHDW	Total Resource Pool	48
SMRP/PRIS	Total Resource Pool	46
MyHDW	Ave Resource Burn Rate/Allocation Peak Months Nov 2015 to May 2016	37
MyHDW	Ave Resource Burn Rate/Allocation Peak Months Nov 2015 to May 2017	36
SMRP/PRIS	Skills Assessment Received	24
MyHDW	Skills Assessment Received	14
Common	Skills Assessment Received	14
PIK	Skills Assessment Received	14
BPM[10]	Skills Assessment Received	5

Table 1: *PMP and Skill Assessments Sample Value*

9 FTE- Full Time Equivalent. An FTE is the hours or days worked by one employee on a full-time basis. The concept is often used to convert time worked by several part-time employees into the time worked by full-time employees.

10 *Bahagian Pengurusan Maklumat* (Information Management Department) functions as consultant on MOH's behalf for all ICT projects in the ministry.

IT/BUSINESS – ORGANIZATION STRUCTURE AND PROCESSES

To effectively develop, maintain and in future run MyHDW 1.0, several common IT organisational structures and processes may need to be established with whomever is responsible for this work. These are as follows:

- IT- Architecture and Standards: It is common in most IT functions in mid to large-size organisations to have an Architecture and Standards area or department. In terms of healthcare data warehousing, this is particularly important given the high numbers of architecture artefacts and standards and the cross-organisational nature of these types of developments. It is recommended that in the mid to long-term, we review the need for this type of arrangement to be provisioned.
- IT - Operations and ITSM: Once MyHDW 1.0 is built and established, it will need to be operationalised, maintained, and enhanced if we are to ensure good user satisfaction with the product. Consideration should be given in the mid to long-term, how best we provide the necessary services to do this. IT Service Management (ITSM) is one common method used to systematically run a medium to high complexity IT organisation and its implementation should be considered. ITSM is a substantial undertaking and priority needs to be given to which modules are required. It is recommended only a simple ITSM implementation or equivalent is examined at this stage, specifically Service Desk, Change Control and Incident Management applied to applications, operations and infrastructure should be sufficient. Given the long timelines required, careful consideration should be given to planning and change control. MIMOS Berhad has a good deal of this infrastructure already established.
- IT/Business - Security Team: An independent Security Team to manage and monitor was established for MyHDW. This is responsible for the development and maintenance of an Information Security Plan (ISP) and Information System Management System (ISMS) aligned with *RAKKSSA*[11] *1.0, Public-sector Cyber Security Framework*.
- IT/Business - Testing Team: An independent Testing Team was also established for MyHDW once the project was underway. This team is responsible for Testing and Data Verification related activities.
- It is planned that a Transfer of Technology (ToT) will occur between MIMOS Berhad and Bahagian Pengurusan Maklumat (ICT Division of KKM). To facilitate this process, it is important that sufficient resources are allocated and that these have the correct level of skills. This activity should be carefully monitored for feasibility and aligned with customer satisfaction feedback.
- Business/IT - Web team: While there are tasks associated with customizing portal webpages, at the time of writing there is no dedicated web team for MyHDW with a business focus. This type of Content Management function required such as to publish analytical material on websites, maintain portals is a common maturity step once a health analytic environment

11 Rangka Kerja Keselamatan Cyber Sektor Awam (Cybersecurity Framework for Public Sector) is used as a guide by the public sector to enhance cybersecurity protection and manage government ICT assets. It sets a standard on how to identify, detect, respond, and recover from cyber attacks and system failures.

is established. Longer term it is recommended this is considered as funding allows.

- Business - Privacy Team: Many international organisations with a health analytics function often employ specialty privacy roles to ensure good practice, legislative adherence and breech management. This direction is aligned with current Ministry of Health (KKM) policy. It is recommended that additional capacity associated with this role be developed in the mid-term.
- Business/IT - Data and Information Governance: Organisational structures and artefacts associated with this function have been instigated as part of MyHDW 1.0 for this important item. See Data and Information Governance section below.
- Business - Health Analytic Products, Reports and Indicator development: In the mid to long-term this becomes a significant area of investment and value. This function provisions the development and implementation of high quality and relevant analytic products, particularly as relates to Health System Management. These might include specialised reports, Health Indicator development, planning tools etc. It is recommended that consideration be given concerning the levels of investment needed to support these.
- Business: Overarching analytic capacity. National analytic capacity including data analyst training, certification and possibly promotion of the Data Scientist role will need to be considered to maximize the return on investment on MyHDW. International experience shows that commonly there is a lack of qualified resources capable of doing data analytics.
- IT/Data Centre: As mentioned above, MIMOS Berhad has been selected as the initial data centre for the project. Originally, it had been planned that this would be located and provisioned within the federal Data Centre at PDSA. Due to limitations associated with non-governmental development staff access and other factors, in discussion with MAMPU, it was agreed upon that it would be more optimal to initially provision these services within the MIMOS Data Centre. Ongoing evolution and scaling of this data centre should be closely monitored and decisions made concerning the long-term location of this function.

SKILLS TRAINING

It will be necessary in the short to mid-term, as part of capacity development, that BI/DW IT skill, Data Science, Health Analytics, and Statistics training be incorporated into HR and organisational practice. This ranges from careful talent management, recruitment practices through to training programs aligned with project needs an individual career development. Specific programs currently in place at MIMOS Berhad to facilitate this are as follows:

- TRI (Technology Recipient Incubation)
- CODE 8 (Centre of Domain Excellent 8 Labs)
- Transfer of Technology (ToT) activities associated with MyHDW
- Post-Graduate Training (Master and PhD Level)

MALAYSIAN HEALTH DATA WAREHOUSE (MyHDW)

2015-2016 START UP: INITIATION

REQUIREMENT SPECIFICATIONS

Requirement Specifications

It is normal practice in system development to gather business and system requirements during the system development life-cycle. In a waterfall style project, this is normally done in two sequential stages, Business Requirements and System Requirements. In this process, user requirements are elicited from key stakeholders and domain experts, and systematically documented in a way that can be translated into system specifications, and design considerations. It is most common that this work is led by a Senior Business Analyst (SBA) or on small projects a Business Analyst (BA). From this process, two system artefacts are produced; a Business Requirement Document (BRD) and a System Requirement Document (SRD), which normally go through a formal approval once complete as part of the project management life-cycle. In some instance, the suffix 'Book' is used instead of 'Document', and this is the case for the MIMOS Berhad SDLC. Thus for this project, these documents are known as the Business Requirement Book (BRB) and System Requirement Book (SRB) respectively.

In terms of requirements gathering to support the development of MyHDW 1.0, while this began within the formal project structure in August 2015, there were several antecedents that supported and accelerated the project requirements gathering work. In terms of chronology, this began in 2011 with the initial MyHDW Blueprint work. As part of creating a roadmap, and portfolio plan in collaboration with an international consultant, this also laid the groundwork and documented the initial requirements for MyHDW. This was done in a series of workshops run at that time and in subsequent years. The MyHDW Blueprint laid out a plan based on lessons learned in the international arena for a Healthcare Analytics system, based on traditional BI/DW approaches. This was later supplemented to add the inclusion of semantic approaches utilizing semi-structured data, modelled through SNOMED CT. MIMOS Berhad also supported and implemented a full-scale beta release of MyHDW. These factors contributed ultimately to the requirement gathering work carried out under the auspices of this project. Reference to the MyHDW Blueprint was used as a touchstone to frame much of the activity.

As part of the consultant review process, the BRB and SRB documents produced in September and October 2015 were reviewed and feedback shared with the project team. These artefacts were based on MIMOS Berhad SDLC requirements specification templates. Requirements gathering was done in an intensive workshop style format. Approximately, one-month duration was allocated to the completion of each document. This is considered a rapid schedule given the size and complexity of MyHDW 1.0. Despite this and some course corrections, the documents were produced and approved approximately in line with project timelines.

LESSONS LEARNED – BUSINESS AND SYSTEM REQUIREMENTS

The BRB and SRB review process gave an opportunity to provide feedback to; better align these artefacts with international and industry best practice, include BI/DW and health analytic specific material and course corrections necessary for effective project delivery. Overall, given the timeline of the requirements gathering, the newness of this type of initiative to the host organisation and the considerations as described above, the following are the lessons learned associated with the review of the Requirements Specifications for MyHDW 1.0:

- The project team delivered the BRB and SRB's broadly within the project timelines. The workshop style 'intensive' approach used during much of the stages, likely contributed to the ability to complete these complex documents on time as did the strong experience of the Project Leads assign to the project.
- There appeared some misinterpretation as to the way in which BRB and SRB's should be written. Current requirement specification practice for both MIMOS Berhad and KKM, places a different emphasis of the business and system requirements within development artefacts. In particular, the BRB should primarily be composed of business scenarios and non-system specific details. These latter are more the domain of the SRB. It is recommended that reference be made to international industry practice for Business Analyst such as BABOK[12] and that resources assigned to this role are either certified in this, receive training or have equivalent experience.
- The BA role is a critical resource for health applications and BI/DW initiative. Under normal circumstances as mentioned above a Senior BA would be accountable for the completion for BRB and SRB material. It is recommended that this factor be considered for future project phases, and careful recruitment and selection of this role be undertaken. Also, see footnote to previous bullet concerning training and possible certification.
- Care should be taken to precisely and comprehensively define scope within BRB and SRB documentation. If this is not done, it can cause significant downstream problems in project delivery.
- BRB and SRB artefact templates should contain specialised sections associated with BI/DW projects. These were provided by the consultant for MyHDW 1.0, and include for example Data Extract Specifications, Dashboard and Report Template subsections. The current SRB template conforms to these recommendations.
- In regards to BI/DW specific systems, artefacts as mentioned above are also necessary to align these aspects of the project with a recognised and effective BI/DW methodology and architecture. For this project, it was determined that these will be based on the Kimball' Data Warehouse Lifecycle[13] methodology. This is a well-established approach to BI/DW systems, and while it is primarily based in traditional structured data implementations it can also scale well too BDA augmentation.
- In the preparation of BRB material, it is particularly important that specifications fully capture the business context, and potential business benefits, and value associated with the system concerned. This is necessary to ensure that the developed system delivers value, and is implementable in a business context, and does not become technology led.
- Change Request (CR) processes need to be formally adopted within the project life-cycle to ensure that additional work associated with extra requirements and functionality changes are recognised. For MyHDW, this was instigated at the beginning of 2016. Problem Report (PR) is the error or defect found during the development stage. Both CR and PR's

12 Business Analysis Body of Knowledge. See https://en.wikipedia.org/wiki/A_Guide_to_the_Business_Analysis_Body_of_Knowledge

13 Kimball Data Warehouse Lifecycle from: The Data Warehouse Lifecycle Toolkit , Second Edition, 2008 by Ralph Kimball, Margy Ross, Warren Thornthwaite, Joy Mundy and Bob Becker Published by Wiley Publishing, 2008

need to be discussed and ratified by the Change Control Board (CCB). The CCB committee consist of the Project Chairman, Project Director, Project Lead, SME, and team members. During CCB meeting the CR and PR will be presented, and the team will provide impact analysis on the resource, timeline and costs. If there is an impact to the project, the next step is to present in the Project Team meeting, Technical Committee meeting and finally in the Steering Committee meeting.

- With a longer-term view, Agile or Rapid style project management and system development techniques should be investigated as an alternate to a waterfall style approach. That said, implementing Agile style project management should be done in a very careful and controlled manner and may take some years to implement. At this stage the recommendation is simply to investigate this notion.

BI/DW SPECIFIC REQUIREMENTS

As mentioned above, BI/DW projects often utilise a custom version of requirements that have a greater emphasis on information, and data requirements compared to traditional approaches. The following, outlines an example of generalised BI/DW Requirement Specification:

- **Business Scenarios & Business Questions**: In Data Warehousing/BI systems, it is important to understand the business usage situation that might benefit from improved informational knowledge. Furthermore, targeting specific questions that support business processes provides a strong underpinning for design purposes. Categorizing this information in terms of Dimensions and Facts (Metrics), allows the efficient creation of an analytics system in line with data warehousing/BDA good practice.
- **Information Requirements**: Information is commonly defined as data in context of a specific situation or business process. It should be noted that in MyHDW the adoption of the MyHIF is the underpinning for defining and specifying this type of requirement. *See below for further details of MyHIF.*
- **Data Sources (feeds into the warehouse)**: This refers to data sources that are used to load the data warehouse from structured files or tables. The specification of the source data combined with process and triggering event information allows for effective development of ETL processes and the automation in the design stage of data warehousing/BI/BDA initiatives.
- **User Types and Stakeholders**: The classification of users into different categories, sometimes known as personas is necessary to categorize the different types of information systems, reports, analytic environments, dashboards and their associated functionality that these different user groups interact with.
- **Dimensions, Metrics, Hierarchies**: Data warehousing systems, commonly categorized data into either dimensions or metrics. Dimensions should be considered descriptive fields that also can be used to filter data. Metrics or Facts as they are sometimes known, are commonly numeric values that occupy the cells in the body of a report. For example, average length of stay, procedure duration etc.
- **Dimensional Modelling** - Some methodologies recommend data modelling at this stage.

- **DW Architecture**: Describes the desired high level Data Warehouse architecture of importance is the approach used to conform or integrate data within the architecture.

KIMBALL DATA WAREHOUSE LIFECYCLE[14]

As mentioned above during project execution, it was determined that for the structured data aspects of MyHDW that the Kimball data warehouse life-cycle methodology be adopted to supplement and guide future system development life-cycle artefacts. This life-cycle is shown further illustrative purposes below.

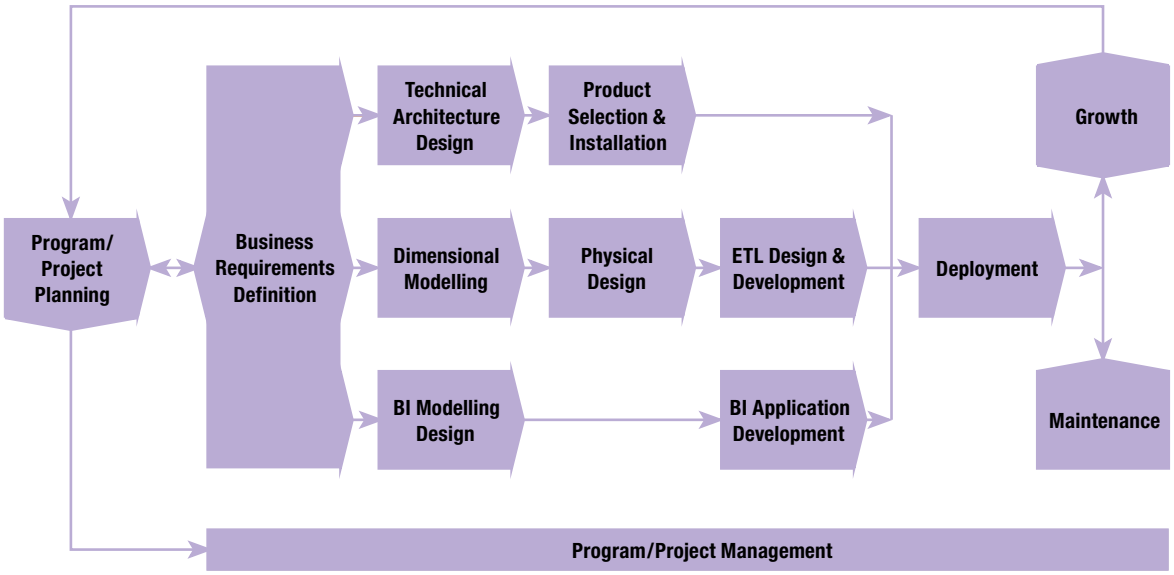


Diagram 3: *Kimball Data Warehouse Lifecycle*

14 Kimball Data Warehouse Lifecycle from: The Data Warehouse Lifecycle Toolkit , Second Edition, 2008 by Ralph Kimball, Margy Ross, Warren Thornthwaite, Joy Mundy and Bob Becker Published by Wiley Publishing, 2008

MyHDW Implementation Methodology

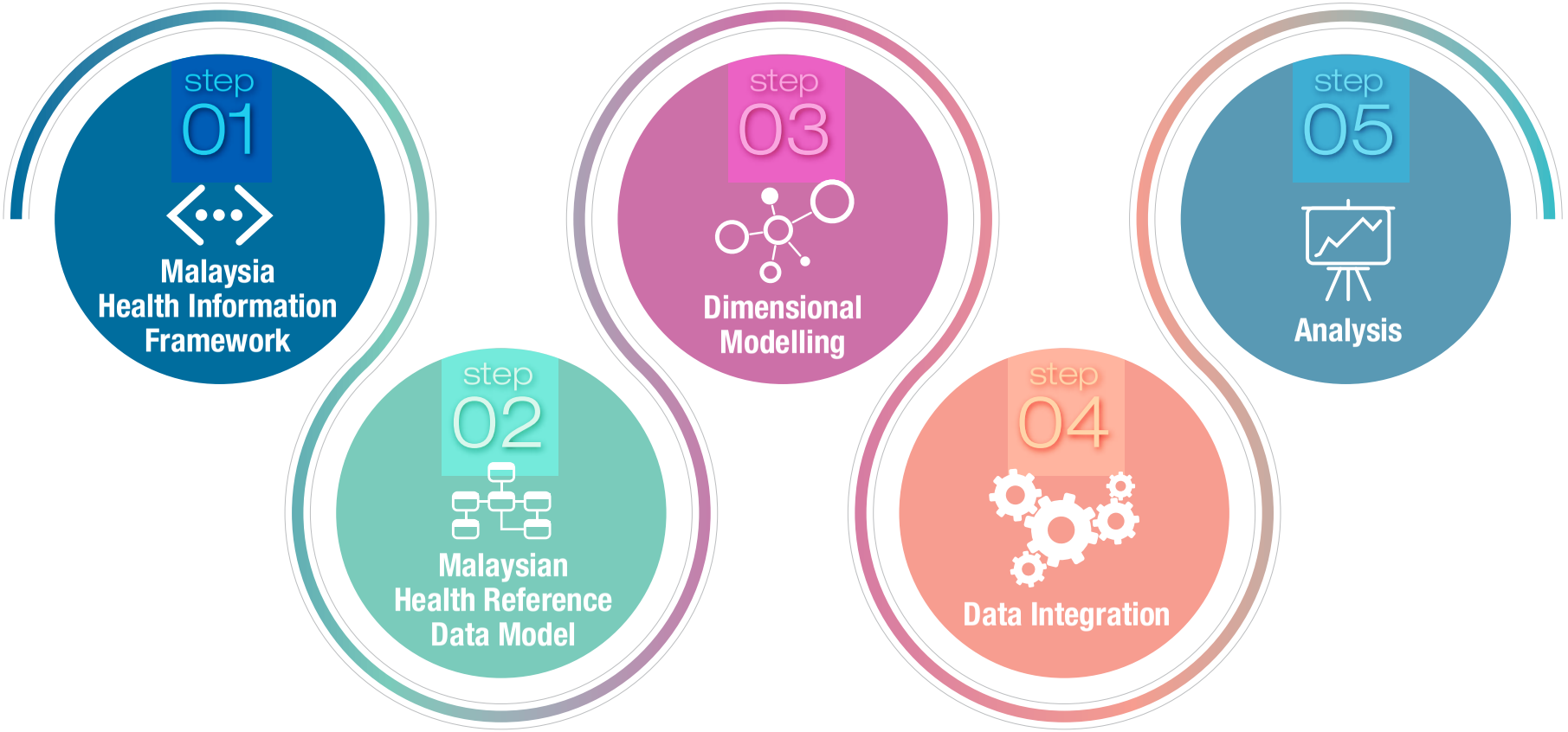


Diagram 4: *MyHDW Implementation Methodology*

MALAYSIAN HEALTH DATA WAREHOUSE (MyHDW)

2015-2016 START UP: INITIATION

HEALTH INFORMATICS STANDARD IN MALAYSIA

Health Informatics Standard in Malaysia

Health Informatics Standard in Malaysia for semantic interoperability cover categories of International Classification of Diseases (ICD), Malaysian Health Data Dictionary (MyHDD), Logical Observation Identifiers Names and Codes (LOINC), SNOMED International including Reference Set development, and Malaysian Health Reference Data Model (MyHRDM). Pusat Informatik Kesihatan (PIK) develops, maintains and monitors utilisation of these standards.

INTERNATIONAL CLASSIFICATION OF DISEASES (ICD)

Malaysia has a policy to use International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10) for diagnosis and International Classification of Diseases, 9th Revision – Clinical Modification (ICD-9CM) for procedures. ICD-10 was introduced in 1998 and fully implemented the following year. There are legacy data codified in ICD-10 but in various editions. It was only in 2014 that funds were available to purchase ICD-10 2010 Edition and ICD-9CM 2013 Edition and allow the standardisation of diagnostic and procedure data across hospitals within and outside of KKM.

For quality of coding, there are 3 main activities conducted throughout the year:

- Training of Medical Officers in documentation of diagnosis and procedures
- Training and Certification of coders
- Coding Error Rate Study

Medical officers are given training on proper documentation of diagnosis and procedures in the Admission and Discharge Form, PER-PD301. This form is mandatory for all inpatient hospital discharges. For diagnosis, it can be further separated into mortality and morbidity documentation. PIK has printed a pocket book as a reference to all doctors, which is also available as an Android mobile application. Proper documentation of the diagnosis will facilitate the correct coding in ICD-10. Initially, PIK conducted the training with State Departments for House Officers. As the need grew, PIK proposed to include the training in a pre-service course for new doctors. Currently, the training is coordinated by Training Management Division who handles the pre-service course. Trained officers from the State Departments as well as from PIK are provided with an hour slot.

Training and certification of coders allow identification of Medical Record Officers who are eligible to code or code under supervision. Initially, eight expert coders were identified as evaluating officers based on the best marks in ICD-10 coding examination conducted by the International Training Centre for Case MIX and Clinical Coding, 2008. Their function includes formulating questions, conducting the exam and marking the exam papers. Candidates identified consist of Medical Record Officer (MRO), Assistant MRO (AMRO), and Record/Operational Clerks. Participants now include coders from private and non-KKM hospitals.

Coders undergoes ICD-10 training, and certified at 3 levels. Level 1 measures competency in ICD-10 coding for single conditions. There are 40 questions to be answered within 3 hours and only those with 90% and above mark can proceed to level 2. Level 2 measures competency in ICD-10 coding for multiple and complex conditions/procedures. There are 40 questions to be answered within 3 hours. Again, to proceed to Level 3 they'll have to pass 90%. The final level features real-life documentation of diagnosis that they see in every hospital discharges. There are 30 questions to be covered in 3 hours but the passing mark is at 80%. Certified coders are those who pass at least level 2 by 50% and are those who can code independently. Those who pass all 3 levels will be certified as expert coders and they will be the reference for future ICD activities.

To monitor the quality of coding, PIK conducted coding error rate studies. The initial study in 2011 evaluates the baseline situation using 2008 SMRP inpatient data which showed 37.5% error rate for 4-digit coding and 21.3% error rate for 3-digit coding. This study became a part of a National Indicator Approach (NIA) for the Quality Assurance Program (QAP). Subsequent studies in 2013 (using 2012 SMRP inpatient data) and in 2014 (using 2013 SMRP inpatient data) showed only slight improvements, as in Table 2. This was thought to be due to a high turnover of new Medical Record Officers, and House Officers who needs to be trained. PIK then strategised to introduce centralised coding to address this issue.

Centralised coding is led by a team of certified coders in PIK which was established in 2014. They are responsible to recodify the discharge diagnosis in PER-PD301 form. In 2016 (using 2014 SMRP inpatient data), error rate for centralised coding has improved and almost meets the target compare to hospital-based coding.

Year of study		2013	2014	2016	
Year of data sample	2011	2012	2013	2014	2015
Target	<30%	<25%	<20%	<15%	<10%
Hospital-based coding error rate		22.9%	22.4%	21.9%	
Centralised coding error rate				15.9%	

Table 2: *Centralised Coding Error Rate*

SNOMED CT INCLUDING REFERENCE SET (RefSet) DEVELOPMENT

Development started in 2013 after Malaysia membership with IHTSDO in late 2012. The first workshop was held in February 2013, which was attended by a senior representative of IHTSDO. The Cardiology and Oncology head of service was also present at the workshop, along with respective team members, who jointly made the decision to focus on terminology used in the respective clinical patient registry. MIMOS Berhad was also invited due to its interest in ontology and ICT expertise. Based on further discussions, the output would be to develop a reference

set based on clinical terms and a tool to harmonize and codify at the backend database with SNOMED CT.

RefSet development focused on the Cardiology domain with emphasise on buy-in process among the stakeholders. The methodology of RefSet development and stakeholder engagement was presented at the SNOMED CT conference in 2014.

RefSet development expanded from Cardiology Registry RefSet to include all Cardiology related terms and pharmaceutical products used by the discipline.

MIMOS Berhad developed MyHarmony as the tool to harmonize and codify clinical terms in unstructured data (narrative text) with an analytic component to demonstrate higher value in identifying cardiac cases, as compared to equivalent queries utilising structured data and classifications (ICD-10) approaches. MyHarmony is currently being further refined to handle post-coordination and contextualisation in alignment with SNOMED CT guidelines. In terms of BDA and next generation analytics, MyHarmony is capable to facilitate and has a direction to increase utilisation of unstructured data and semantic data integration.

Logical Observation Indentifiers Names and Codes (LOINC)

LOINC is another international terminology standard used in Malaysia. All KKM laboratory pathologists have agreed to use LOINC to standardise their Laboratory Information Systems (LIS). Three workshops were conducted to map the laboratory orders to LOINC from 2014 to 2016. In the first workshop, two representatives from Regenstrief Institute who developed and maintained LOINC was present to guide the participants. RELMA and LOINC online browsers were used. Participants from different health facilities with LIS mapped their own list of laboratory orders to LOINC. Each orderable were mapped based on matching parts in LOINC which are component, property, timing, system, scale, and method.

One of the issues found during these exercises were the mapping of one-to-many. For example, the test for Erythrocyte Sedimentation Rate (ESR), has 4 LOINC codes with specific method and one LOINC code without specified method. The pathologists were advised to confirm on the method used in their facility for specific mapping to be done. Or else, the ESR will be mapped to LOINC code with an unspecified method.

Other than using LOINC to standardise laboratory terms, it is also proposed to be used for analytics. Based on the collaboration between LOINC and IHTSDO, Malaysia will codify laboratory questions in LOINC while the answers will be coded in SNOMED International. A demonstration of the use of LOINC in MyHarmony, was presented in the SNOMED CT Expo 2016. This feature will be further enhanced in the future for analytics.

National Health Data Dictionary (NHDD)

National Health Data Dictionary (NHDD) is the health component of Data Dictionary Sektor Awam (DDSA – Public Sector Data Dictionary). Any health component definition that is developed by

the Ministry of Health (KKM) for MyHDD will be adopted into DDSA. It is managed by Malaysian Administrative Modernisation and Management Planning Unit (MAMPU).

The first NHDD was developed in 2003 by KKM for the Health Informatics industry. The objectives of NHDD are to provide semantic understanding of data elements used across health care industries; conform to nationally accepted protocol standards; and to be consistent with international standards wherever applicable. This first NHDD document was published in 2007. It consisted of a minimum required dataset for demographic data collection, which was known as Person Record Dataset.

A series of workshops involving various stakeholders were then conducted in 2012 for further development and revision of this NHDD. It was also renamed as Malaysian Health Data Dictionary (MyHDD) which incorporates data elements from Patient Registries, and the Malaysian Health Reference Data Model (MyHRDM). The first workshop in 2012 was attended by representative of Healthcare Information Technology Standards Panel (HITSP) and Joint Learning Network (JLN) for Universal Health Coverage to confirm on the methodology to develop MyHDD using a consensus-driven approach. JLN for Universal Health Coverage noted the consensus-driven methodology and has invited PIK, Planning Division to present MyHDD in their international workshops and published MyHDD's content on their website (Openhdd.org).

A total of 11 datasets were made available online on KKM's website in 2012 namely: Revised Person Record Dataset, Discharge Summary Dataset, Emergency Dataset, Cancer Dataset, Radiotherapy & Oncology Dataset, Nuclear Medicine Dataset, Obstetric & New born Dataset Oral Health Dataset, Pharmacy Dataset, Forensic Dataset, and, Mental Health/ Psychiatry Dataset.

MALAYSIAN HEALTH INFORMATION FRAMEWORK (MyHIF)

Information Modelling is a necessary and sometimes overlooked technique which was utilised in MyHDW 1.0. In terms of a national standards perspective. The original MyHDW Blueprint recognised the need for this type of work and outlined an artefact, Malaysian Health Information Framework (MyHIF) to accommodate this. MyHIF is essentially a requirement gathering tool for Health Data Warehouse design to allow a global understanding of the information need in a holistic and system wide manner. Specifically, MyHIF is composed of a grid with rows and columns and details the following information:

- Question - performance based question that addresses business improvement need. For example, how can we improve access to emergency department for patients in need? What is the workload of Day Care Services in the hospital to plan for more of such facilities?
- Function - Function affected (e.g. Access, Quality, Efficiency).
- Information Users - Position or users that might use the information product, once developed, to inform action that can improve patient care. For example, Minister of Health, Senior Managers in KKM, Hospital Directors and State Health Directors.

- Candidate Information Product - Likely type of information products for examples, Health Facts, Health Indicators, Statistical Report, KPI, and, Dashboards.
- Available Data/Infrastructure - Does a data source exist? If so what is the data source? Are there standards in place for these data sources?
- Frequency - Ideal frequency that these data and information would be available through MyHDW. For example, daily, monthly, quarterly, or yearly.
- Use of Information Products - How the information/product could affect healthcare performance. For instance, identify and share best practices to reduce waiting time, or monitor health tourism returns from participating Health Tourism Hospitals.
- Priority - The relative priority of the question(s) and associated information products. The evolution and development of a data warehouse system such as MyHDW likely occurs over many years. Practically, there is a finite amount of resource available for this work. Therefore, it is necessary to prioritise which information is needed from MyHDW to allow for its effective evolution.
- Order of Magnitude - Cost to develop and maintained information products. Indicate a nominal value (e.g. High, Medium, Low) to answer the question(s) associated with an information product or report.

Early versions of this product were produced prior to the start of the project within a national context and then further enhanced with specific MyHDW 1.0 examples as part of the requirements specification activities.

For further details of MyHIF, please see the *MyHDW Blueprint 2011 to 2013.*

MALAYSIAN HEALTH REFERENCE DATA MODEL (MyHRDM)

To support the development work of MyHDW 1.0, SMRP 2.0 and PRIS 1.0 and the associated Conceptual Data Warehouse Architecture, it was necessary to establish an initial approved version of MyHRDM and make this available to the development team. The initial development of MyHRDM concerns building consensus on key Concepts. A Concept sometimes called an Entity, is related to a precise definition that is universally agreed on by the stakeholders involved, for example, Encounter and Visit; these concepts if agreed upon in various systems like SMRP, PRIS and TPC/OHCIS, will allow for the creation of a model that is repeatable and consistent in all the mentioned systems. By defining these two Concepts in a clear and agreed-upon manner, we were able to build reporting and information products efficiently and with the minimum degree of misunderstanding. In addition, query questions that involve multiple Service Type such as the movement of patients from Inpatient to Rehabilitation services for example, will be easier to perform in contrast to not having the model.

It should be noted that an often-overlooked assumption associated with these types of data models is that in addition to providing standardised business definitions, these are ultimately intended for deployment in Relational Database Management Systems (RDBMS) such as Oracle

or SQL Server. This fact combined with the mathematical precision of the models explains the importance of these types of structures when building data centric systems. Until recently, for the last 30 to 40 years, nearly all serious system development involving large amounts or complex data have utilised RDBMS's and thus the need for modelling to ensure these are defined in an efficient, internally consistent and high performing manner. With the advent of newer technology such as Hadoop, Graph Databases, Document Databases etc. over the last while, real alternatives to RDBMS's are now viable, that said, at the time of writing they are still used for the majority of deployments and hence the need for Entity/Relation(E/R) data models. While there is also a movement within the industry towards increasing use of semantic relationships (semantic web, ontology models, Resource Description Framework (RDF) etc.), it is likely that traditional structured data as expressed through data models and databases will sustain and work as part of a broader environment including these semantically integrated data sources. As such, ongoing maintenance and development of a minimum set of core structure data concepts will still be required and the ongoing maintenance and development of MyHRDM is planned.

In 2013, an initial exercise was undertaken to build the HRDM and define several concepts to support this as part of a MyHDW workshop. This work was successful within the confines of an exploratory workshop, but required further work to complete to a level consummate with use in the development of deployable solutions such as those which are the topic of this report.

MyHRDM WORKSHOP NOV 4 - NOV 6 2015 – SMRP/PRIS

During November 2015, a MyHRDM Workshop was arranged with the objective of completing an initial approved version of MyHRDM for use by the project team. This would include an approved set of Concepts and Definitions and an initial MyHRDM Data Model including those concepts. While recognising that this model would involve a holistic approach, only key concepts or entities associated with MyHDW 1.0, SMRP 2.0 and PRIS 1.0 were included.

The participants present were shown basic techniques utilised in building data models for healthcare systems development. These techniques were elaborated on to demonstrate the importance of enterprise level data models in the use of Health Informatics Standard, and as an underpinning for building high quality and cost effective healthcare systems. International examples were provided, including process and governance considerations. The positioning of a Malaysian Health Reference Data Model (MyHRDM) as a necessary supplement to the Malaysian Health Data Dictionary (MyHDD) was also discussed, and recognized by those present as being of importance. There was also recognition that the addition of concept definitions would greatly benefit data elements in the MyHDD, by better setting context for these items. The use of the term Reference data model was explained to the group and compared with the more traditional terms used in data modelling, such as Conceptual, Logical and Physical data model. It was noted that when developing systems within large organizations, it is best practice to develop an Enterprise Data Model; in healthcare, this is similarly required. The complexities and non-standardisation of processes in healthcare delivery, complex stakeholder relationships and governance warrant the use of a Reference data model approach, as this greatly increases the chance of successfully completing an initiative of this nature. This Reference Data Model is like a

Conceptual Data Model[15,16], but limited to only the most necessary data concepts. Internationally there have been several parallel initiatives which recognise the need to limit modelling activities and their associated early binding to business rules to the minimum. Associated with this is that implementation structure such as a schema is deferred until required. The reference model associated with MyHDW utilises this approach. This will allow for a rapid development and utilisation of a National Data Model for healthcare or similar without excessive timelines and work effort. A similar approach has been used in other international organization such as the Canadian Institute for Health Information and this was used as a reference[17].

Following this primer, an exercise was done with the whole group to ratify and approve the Concepts (Entities) to be included in this model as well as Concept Definitions. In addition, any Core Attributes from MyHDD in common usage, associated with the concept were identified. Also, a data model was produced along with its associated relationships that utilise these concepts. This was discussed with the group present and this and the Relationships ratified.

MyHRDM WORKSHOPS DEC 2,3,7, 2015 - TPC-OHCIS/ PhIS

Subsequently, additional workshops sessions with Primary Care and Pharmacy domain experts from TPC-OHCIS and PhIS took place December 2,3 and 7, 2015. Given the successes of the previous workshops, it was felt that the opportunity should be taken to attempt to add these subject areas (Primary Care and Pharmacy) to the MyHRDM model. This would further accelerate its development, and provide the possibility of utilisation of the reference model by the related systems. A similar process to the above was undertaken, and a total of about 20 new concepts were identified. Other than concepts specific to the domain mentioned, several generic concepts associated with Episodes were also defined and added. These latter concepts are somewhat more challenging to model, and are particularly useful to connect Clients across the continuum of care. Furthermore, while the preliminary work to establish the relationship associated with these new concepts was not completed with the domain experts, an initial design was undertaken by the modelling team.

The following table identifies the updated Concepts and their definitions, and the diagram following this is the Reference Data Model. These are based on all the sessions to date, including all workshops mentioned above.

15 ISO/TR 2221:2006: Health Informatics - Good Principles and Practices for a Clinical Data Warehouse. International Organisation for Standardisation, Geneva, Switzerland.

16 ISO/TS 29585:2010: Health Informatics - Deployment of a Clinical Data Warehouse. International Organisation for Standardisation, Geneva, Switzerland.

17 See Canadian Institute for Health Information CRDM Toolkit: https://www.cihi.ca/en/submit-data-and-view-standards/methodologies-and-decision-support-tools/data-architecture

MyHRDM CONCEPTS AND DEFINITIONS

Concept/ Entity	Definition
Client	A PERSON who receives health service.
Discipline	A SPECIALTY delivered through an ORGANISATION.
Donor	A CLIENT who has consented to donate or has donated cells, tissues or organs for the purpose of transplantation.
Encounter	An interaction between a CLIENT and one or more PROVIDERS.
Health service Event	A past, current, planned or requested healthcare act.
Intervention	An activity(s) that is intended to observe, monitor, assess and/or change the state of the health of a CLIENT.
New-born	A CLIENT aged less than or equal to 28 days of life.
Next of Kin	A PERSON who is a relative or a friend of a CLIENT.
Organisation	A social unit of PERSONs that is structured and managed to meet a need or pursue a collective goal.
Observation	Information derived from performance of a health related activity to a CLIENT.
Person	An individual.
Place	A physical point or area in space.
Programme	A collection of activities delivered through a SERVICE in an ORGANISATION designed to meet a healthcare need.
Provider	A PERSON who has delivered, is delivering, or has the potential to deliver health care related services or goods.
Service	A grouping or collection of healthcare activities delivered directly or indirectly through an ORGANISATION to a CLIENT.
Specialty	A SPECIALTY delivered through an ORGANISATION.
Standard Assessment	An authorized tool used to collect a set of constituent OBSERVATIONS and INTERVENTIONS .
Visit	An interaction between a Client and one or more SERVICES.
Emergency Contact	A PERSON who is related to a CLIENT who can be contacted in situation of emergency.
Referral	A request from one PROVIDER to another PROVIDER or ORGANISATION to deliver one or more health services to a CLIENT.
Diagnostic and Imaging	A series of diagnostic and imaging activities by the PROVIDER for the CLIENT which cover the capturing and interpretation of images.
Immunization	Administration of a vaccine by the PROVIDER to CLIENT for the purpose of preventing the spread of infectious disease.
Lab Test	A series of lab testing activities by the PROVIDER for the CLIENT which cover the capturing and interpretation of biological samples.
Medication Prescribed	The medications dispensed by PROVIDER to a CLIENT.
Medication Dispensed	The medications dispensed by PROVIDER to a CLIENT.
Maternal Health	A care provided to a female CLIENT from conception till 6 weeks post-delivery associated with antenatal, intra-natal and postnatal.
Provider Screening	A screening PROGRAMME for PROVIDERS associated with CLIENT and PROVIDER safety.

Concept/ Entity	Definition
Nutritional Status	The summary of CLIENT's health condition related to nutrition.
Health Status	The summary of CLIENT's health condition as a result of OBSERVATION and STANDARD ASSESSMENT performed by PROVIDER.
Child Health Status	The HEALTH STATUS of a CLIENT who is defined as a child.
Pre-school Health Status	The HEALTH STATUS of a CLIENT who is defined as being of pre-school age.
School-going Health Status	The HEALTH STATUS of a CLIENT who is defined as being of school-going age.
Adolescent Health Status	The HEALTH STATUS of a CLIENT who is defined as being of adolescent age.
Adult Health Status	The HEALTH STATUS of a CLIENT who is defined as an adult.
Elderly Health Status	The HEALTH STATUS of a CLIENT who is defined as an elderly person.
Episode	A collection of HEALTH SERVICE EVENTs related to health conditions, type of care, VISITs, ENCOUNTERs and other factors.
Episode of Health Condition	Collection of VISITs and/or ENCOUNTERs related to one particular instance of a health condition or collection of health conditions.
Episode of Care	Collection of VISITs and ENCOUNTERs related to one particular instance of a SERVICE or collection of SERVICEs.

Table 3: *MyHRDM Concept and Definitions*

MyHRDM MODEL V1.1

Diagram 5: MyHRDM V1.1

The following diagram is an initial draft of the MyHRDM version 1.1 model, based on the above workshop discussions. This is subject to further refinement and approval.

MyHRDM V1.1, USAGE, APPROVAL PROCESS AND GOVERNANCE

Resources

Critical to the advancement of this work, and prior to the workshops commencement was the fact that experienced resources were allocated in a meaningful way to this modelling activity, provided by MIMOS Berhad. In addition, supporting services were further provisioned by MAMPU to ensure broader alignment with government standards and systems. PIK of Planning Division also agreed to provide secretariat services as well as additional SME resources. It will be important to sustain this level of support and structure, if the model is to be maintained and further evolved.

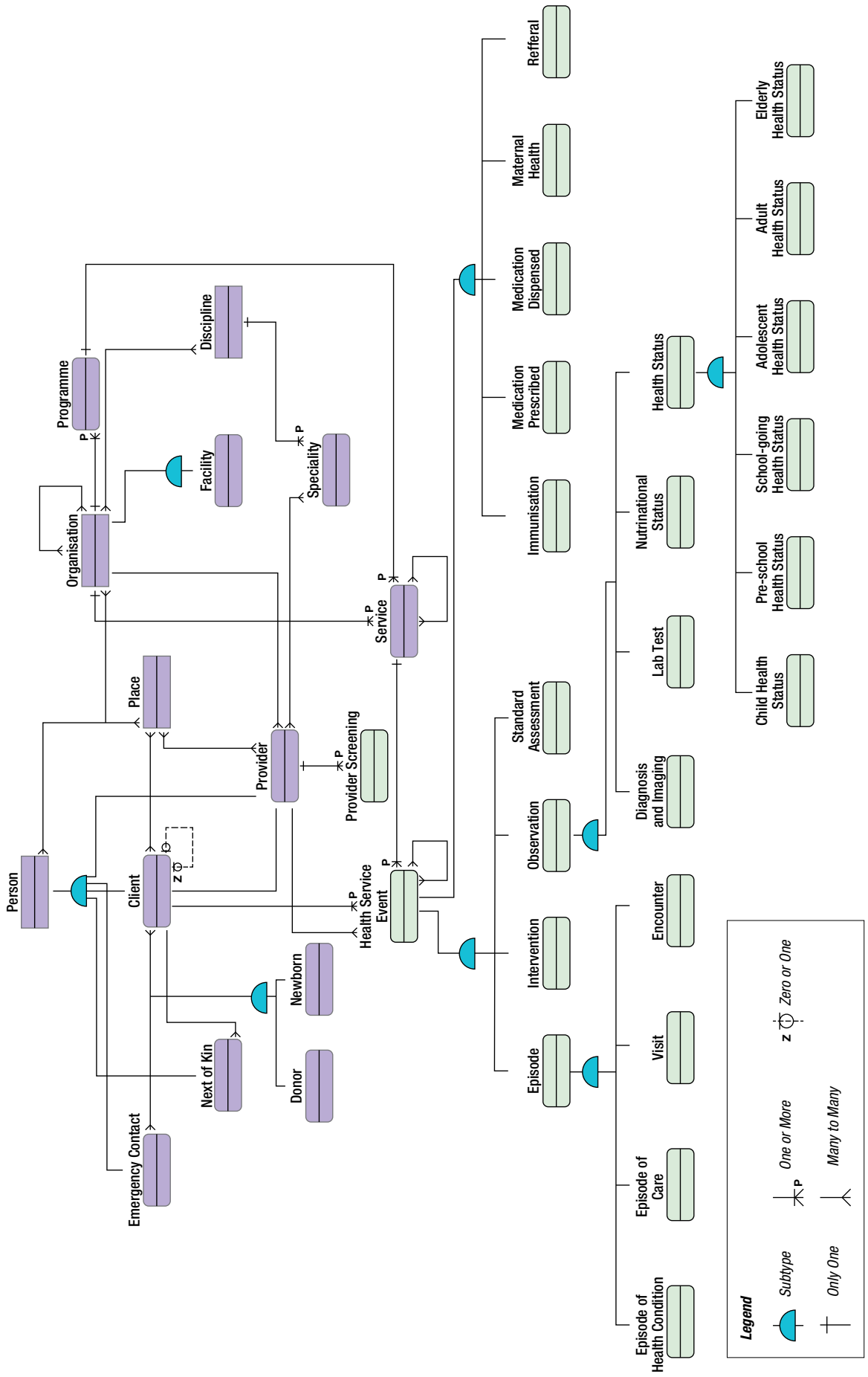


***Diagram 5:** MyHRDM v1.1*

While a ratification process is required for formal approval, as an interim measure the draft model is being fully utilised as a working model by the MyHDW 1.0 project, and used as a reference for all data modelling activity. This is an important achievement and will provide many benefits to the program associated with good design, improved data integration, and ultimately improve return on investment. A process will need to be developed to ratify the model with the MyHDW Technical Committee, and ultimately through a process involving the National Health Informatics Committee (NHIC).

MALAYSIAN HEALTH DATA WAREHOUSE (MyHDW)

2015-2016 START UP: INITIATION

ARCHITECTURE

Architecture

To build high quality and cost effective information systems, it is necessary that these are designed holistically including all key aspects of technical and business implementation. Traditionally, this type of design activity is known as architecture. A common analogy, used to describe the necessity of architecture is that of a blueprint used in the design and construction of a house. One common scheme[18] for breaking down architecture into major components, recognizes four types of architectural artefacts. These are Data Architecture, Technical Architecture, Application Architecture, and Business Architecture. These components are collectively known as Enterprise Enterprise Architecture(EA)[19].

To develop and implement a digital information system requires that software is specified, designed, and implemented in a technical environment such as Java or a Relational Database Management System (RDBMS), for example Oracle or SQL Server. As part of this design and implementation process, it is common that time is spent on detailing Data Architecture, and Application Architecture as part of the System Development Lifecycle (SDLC) to support this. A common type of Data Architecture artefact used during this process is an Entity/Relational Data Model. There are also models developed for application architecture such as Class Models, and Sequence Diagrams et al.

In Healthcare IT System Development and Analytics, in addition to the architecture facets mentioned above, there are often parallel efforts associated with what is known is Health Informatics or Health Informatics Standard. Sometimes, these are known as Content Standards. The objective in doing these activities is to establish a consistent means to identify clinical activity and observations such as associated with diagnosis, intervention, prescription medication and others. The term often used to ensure the consistent and uniformed usage of related information is Semantic Interoperability. The objective of this approach is to ensure consistent meaning is applied at a low level, independent of the implementation method used, or across clinical domains so that predictable results can occur. Example classification and terminology standards within this realm are ICD, SNOMED CT, LOINC and so forth. Additionally, Informatics Standards are also commonly developed for sharing of messages between digital systems e.g. HL7. It is important to note that there is an interrelationship between Enterprise Architecture and Health Informatics Standard, and these must work together to ensure that high quality, reliable and efficient systems can be built. It should be noted that this fact is also consistent with frameworks such as TOGAF, which recognise industry standards that fit within architectural building blocks.

DATA ARCHITECTURE

For MyHDW 1.0, there is a strong focus on Data Architecture that is due to the nature of the systems being developed, specifically BI/DW. In addition to data models such as the MyHRDM, and downstream Logical and Physical Data Models used in application development associated

18 TOGAF - http://www.opengroup.org/subjectareas/enterprise/togaf Also see IASA http://iasaglobal.org/itabok/capability-descriptions/

19 See MAMPU EA initiative for example; http://www.mampu.gov.my/web/en/enterprise-architecture

with SMRP 2.0 and PRIS 1.0, a category of Data Architecture is required and to be adapted to Data Warehousing. Data Warehouse Architecture is a specialised topic and is required in this project as the underpinning for the core design of the Health Data Warehouse. There is a good deal of literature and industry experience associated with this type of architecture, and this is leveraged in the design of MyHDW, and in the earlier Blueprint work. In the current project, this is known as the Conceptual Data Warehouse Architecture.

BDA environments by contrast, use in some instances is a completely different approach to creating analytic environments, in which very little or no data modelling is performed, and extract transformation and load processes are kept to a minimum using the Hadoop ecosystem and other types of NoSQL databases. It should be recognised that a good deal of historic experience associated with data warehouse architecture is as described below, and is based on the implementation of structured data implemented in an RDBMS. It is certainly true that in MyHDW this is a fundamental part of the early design. That said, this design has also been supplemented as will be described to incorporate these evolving BDA techniques specifically through the deployment of Hadoop, Natural Language Processing, Document and Graph Databases.

DATA WAREHOUSE ARCHITECTURE SELECTION

As mentioned, a key aspect of MyHDW is the Architectural Framework which underpins its design and implementation at a conceptual level. this is recognised as a part of the Kimball Life-cycle as shown above and is often done at the end of the Requirements or beginning of the Design phase of this type of initiative. During October 2015 a detailed briefing on Data Warehouse Architecture and Data Architecture methodologies and techniques such as Data Modelling was given by the project consultant. Following this, a discussion occurred with the objective of picking a suitable Data Warehouse architectural methodology for MyHDW 1.0. Two of the more common approaches[20] are outlined below:

Hub and Spoke

- Based on an enterprise wide analysis of requirements and data.
- Developed in an iterative manner subject area by subject area. A highly scalable and maintainable infrastructure with a clear focus on an enterprise view of data.
- Data in the hub for this style of data warehouse is stored in a normalised (3rd normal) form commonly in a traditional entity/relationship schema.
- Data in the hub is stored in an atomic or granular form.
- Dependent data marts are created utilizing the centralised hub data.
- Most users query the dependent data marts.
- Dependent data marts may be implemented using a range of different schema forms including commonly Star Schema.

20 Data Warehouse Architectures: Factors in the Selection Decision and the Success of the Architectures, Hugh J. Watson, Thilini Ramachandra, 2005

Data Warehouse Bus Architecture (Kimball)

- In this approach, the first data mart is built for a single subject area approach using dimensions and measures utilizing conformed dimensions and facts that will be used with other marts.
- Additional marts are developed using these conformed dimensions, and these allow for data integration (or linkage) across different subject areas.
- Schemas are implemented in star schemas which represent the data in a dimensional manner.
- It is common for a business requirement for a process subject area to form the underpinning of the first Data Mart.

During the discussion with the Project Leads in October 2015, it was agreed upon that a modified form of the Kimball Data Warehouse Bus Architecture be utilised in preference to a Hub and Spoke approach. This decision was reached based on the following reasoning. The Data Mart Bus Architecture (Kimball) is:

- A simpler design and better suited to the maturity level of the project and team.
- Likely, a quicker and cheaper solution to implement over the short to mid-term.
- Would reduce risk by not introducing excess complexity into the design, while being able to service the needs of a Healthcare Data Warehouse.

Following this initial discussion and decision point, an additional session was carried out in which the necessary aspects of the architecture were adapted for potential BDA capabilities, to be incorporated into the architecture, and to thus increase its voracity given the current state of these technologies, and to better align with MIMOS platform development and directions. Combining these considerations along with several other best practice approaches, the following Conceptual Data Warehouse Architecture was agreed upon to use in MyHDW 1.0.

CONCEPTUAL DATA WAREHOUSE ARCHITECTURE:

The following diagram describes the Conceptual Data Warehouse Architecture for MyHDW 1.0. This both incorporates the Data Mart Bus Architecture (Kimball), Data Acquisition components, Big Data Analytic technologies, Source of Truth Data, Harmonisation and Codification Services, Presentation Services including BI and Statistical tools, as well as a variety of supporting services such as Master Data Management (MDM), Security and Privacy etc.

It should be re-emphasized that while the current implementation for MyHDW 1.0 focuses on structured data implemented through a traditional BI/DW design as this has a proven track record with this type of data, that future considerations including the MyHDW Phase 2 implementation will include various facets of Big Data Analytics (BDA). The primary vehicles for back-end services for this may cover a spectrum of NoSQL type technologies including Natural Language Processing (NLP), Hadoop ecosystem/HDFS, Document Databases (e.g. Apache Drill), Graph Databases. In terms of data integration, industry trends currently suggest a move towards semantic

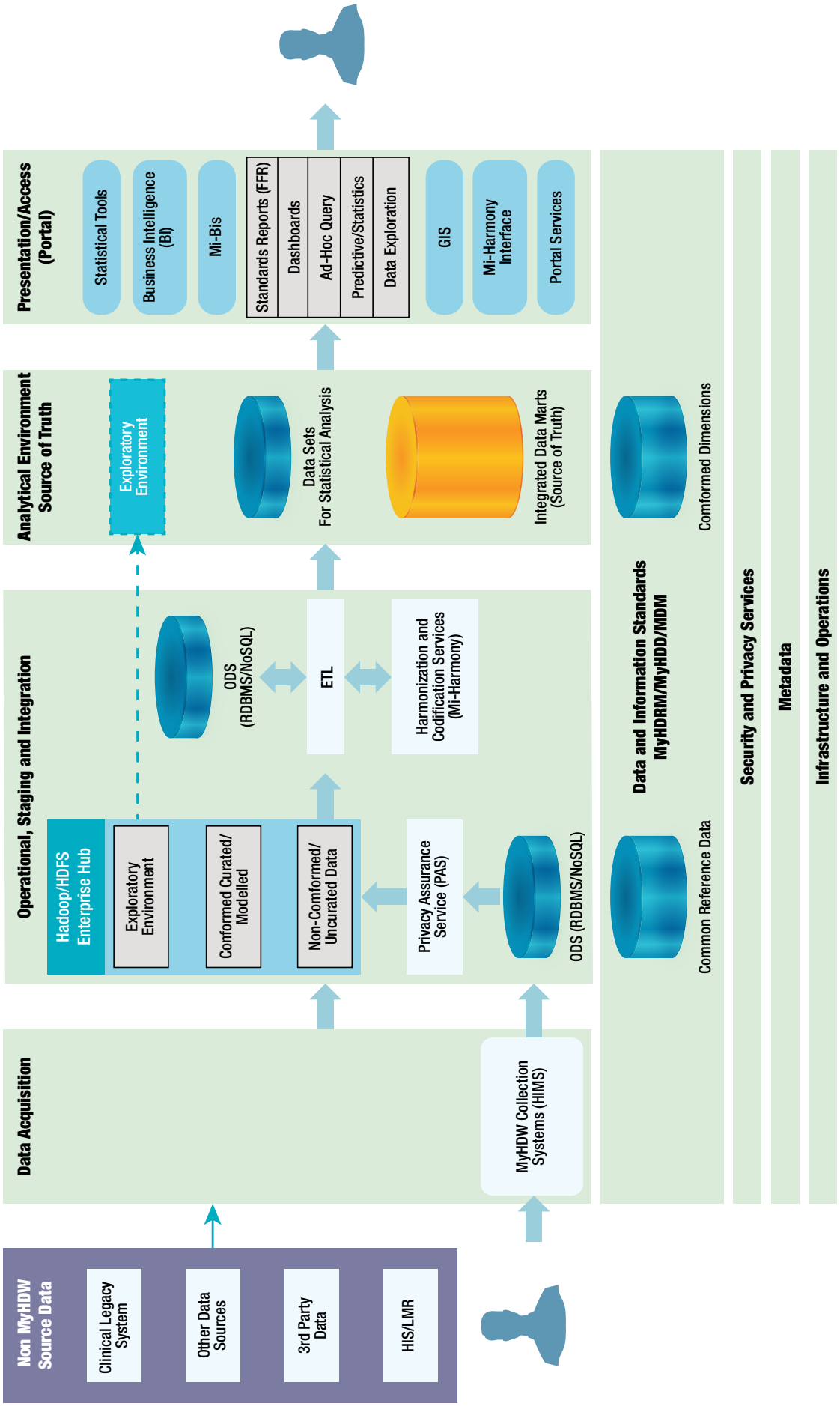


Diagram 6: *MyHDW Conceptual Data Warehouse Architecture*

integration. Likely, again a hybrid approach to data integration is to be expected. MyHarmony is a step towards this. A balanced approach that takes into consideration an understanding of the productivity, and maturity of any of these new technologies was kept in mind in developing this architecture and associated downstream implementation.

Non-MyHDW Source Data

This area, outside of the main MyHDW operating environment is a location of source data from a variety of systems and organisations. This would include any legacy system data, data from third-party sources such as census or geographic data, HIS data, and possible future data sources etc.

Data Acquisition

These are data acquisition collection systems such as SMRP 2.0 developed, based on the core data architecture associated with MyHDW. Specifically, MyHRDM and MyHDD as implemented through Conformed Dimensions and Common Reference Data.

Operational, Staging and Integration

This is a rich environment that contains all the core structures to both provision operational data for the Data Acquisition systems such as SMRP 2.0, provide for a HDFS/Hadoop System for a Landing Area (BDA Platform) for all data in an encrypted form, Harmonisation and Codification services as well as a RDBMS based Staging Area to support ETL activities.

Analytical Environment: Source of Truth

This is curated data that has been cleaned, enriched and standardised, and would be considered the **Source of Truth** for Reporting and Analytic products that require a high degree of accuracy and trust. Data in this area will be stored in the form of integrated Data Marts i.e. Kimball Data Mart Bus Architecture. Data in future releases will be drawn from these Source of Truth Data Marts in the form of Datasets to support use of this data with Statistical Tools. A virtual BDA exploratory area may also be exposed through this layer. While this will not necessarily have the same level of data quality as the rest of the Source of Truth, this will expose parts of the HDFS Landing zone in a privacy sensitive manner to allow sandbox style analytics.

Presentation/Access

A presentation or access layer, managed through a Portal environment and supporting services such as Security Services. This would include BI Tools such as Mi-BIS, GIS and in future releases of Statistical Tools.

Master Data Management (MDM)

This is common data based on MyHRDM and MyHDD standards to support both operational systems such as SMRP 2.0, as well as a query based MyHDW. The standards are implemented through Common Reference Data and Conformed Dimensions.

MyHRDM/MyHDD

Malaysian Health Reference Data Model (MyHDRM) and Malaysians Health Data Dictionary (MyHDD). These standards (see section below), when implemented will form the basis of the MDM services above.

Security and Privacy Services

Approved security and privacy services to support the appropriate use of data in the MyHDW environment

Privacy Assurance Services

As a supplement to the overarching Privacy and Security Services (PAS), a dedicated Privacy Assurance Service will be implemented to ensure that all health data moved into the Landing Area is processed in such a way to ensure individual privacy is fully protected. This may include techniques such as pseudonymisation, scrambling, and removing identifying information et al. In principle, all clinical data will pass through the service.

Metadata

Data and information concerning these facets within the MyHDW environment. The objective of this service is to provide analytic and associated methodology information to inform query, and analysis development and downstream decision-making.

Infrastructure and Operations

Basic infrastructure and operations services such as hardware provisioning and ITSM or similar services to allow for the effective appropriate development, and operations of the MyHDW environment.

LOGICAL DATA WAREHOUSE ARCHITECTURE – PHASE 1

The framework diagram for Phase 1 can also be represented logically as follows:

Diagram 7 Logical Architecture MyHDW (Phase 1)

LOGICAL DATA WAREHOUSE ARCHITECTURE – PHASE 2

The framework diagram for Phase 2 can also be represented logically as follows and includes greater detail concerning new defined Phase 2 specific functionality.

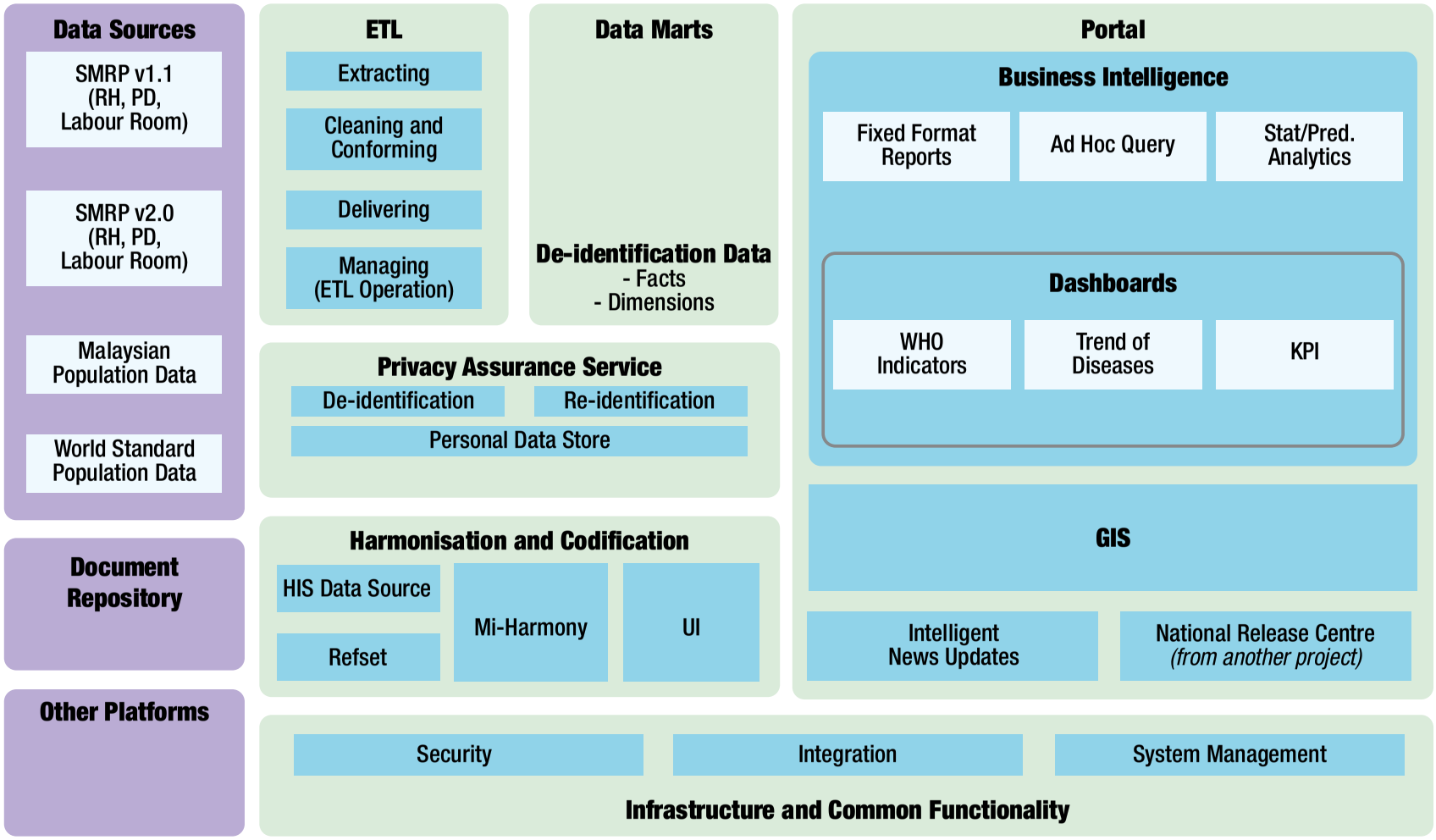


Diagram 7: *Logical Architecture MyHDW Phase 1*

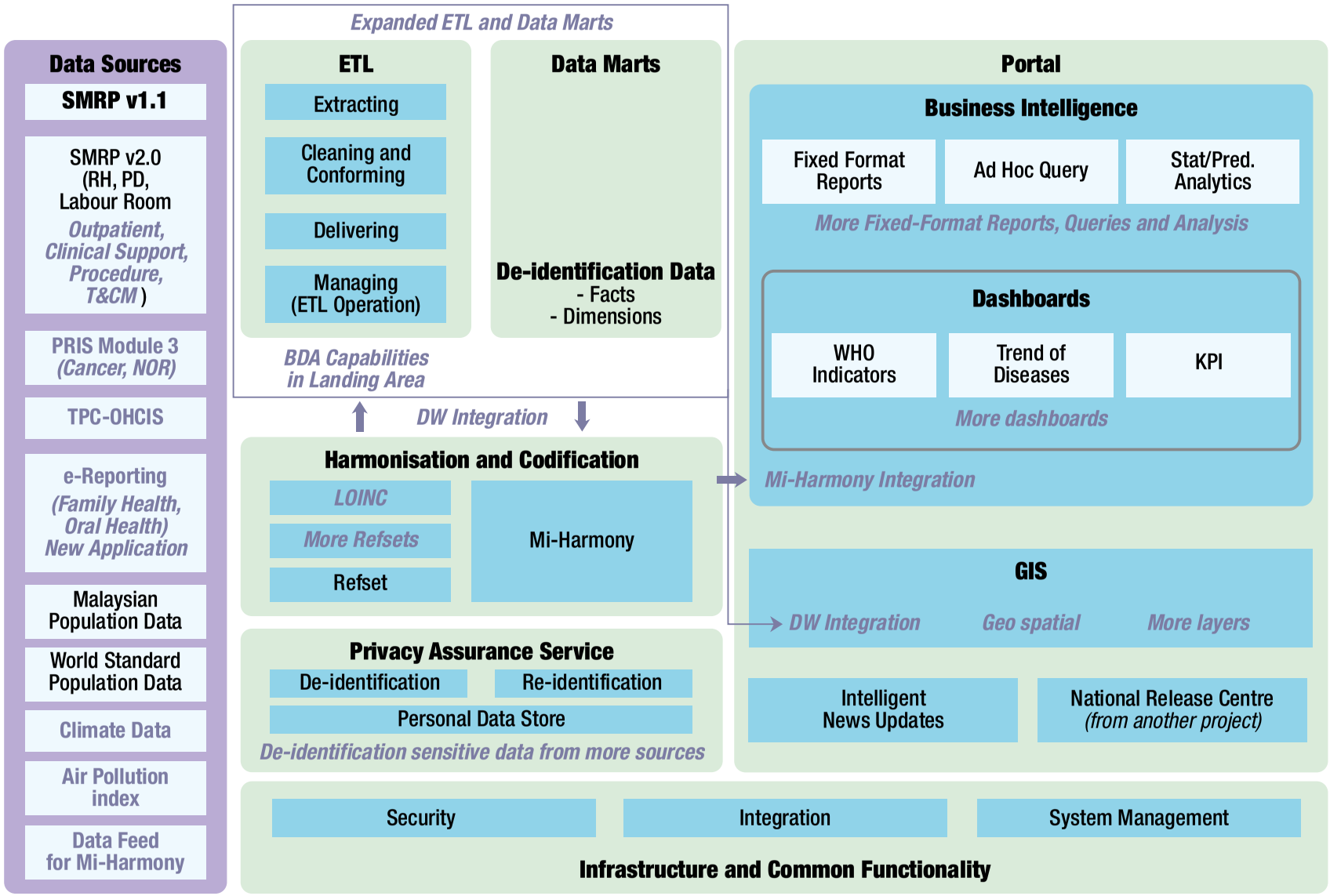


Diagram 8: *Logical Architecture MyHDW Phase 2*

DIMENSIONAL MODELLING

Given the selection of the Kimball Data Warehouse Bus Architecture as described above, Dimensional Modelling was an important part of MyHDW as associated with the BI/DW environment. These Dimensional Models form the basis of the backend structured data utilised in all data query activity for MyHDW 1.0, and utilised by front-end BI query tools such as Mi-BIS. Dimensional Models were designed and built on MyHDW 1.0 requirements documentation. The data models and architects were also advised by the consultant to utilise easily available dimensional modelling guidelines[21] and templates to accelerate this work. Models were created that comprehensively cover at a granular level all available Inpatient and Daycare information source from SMRP. During the project, these were initially designed for SMRP 1.1, and then extended to accept data from SMRP 2.0. In addition, Population Census data models were also built as were other aggregate or specialised dimensional models based on source data.

Modelling resources from MIMOS Berhad were assigned to the project and worked on these models guided by the international consultant, project lead and the guideline reference in the footnote below. These models cover a wide range of query options from standardise reports all the way through to Ad-hoc and statistical queries. Furthermore, these are fully aligned with MyHRDM and can allow integration across different subject areas, effectively allowing data to be linked. For example, it is possible to link Inpatient data to Census data allowing the calculation of Crude and Standardise Rates. As additional data sources are added, it will be further possible to link across the Client continuum of care and Service Types such as tracking patients from Emergency treatment into Inpatient for example. While the detailed design of these models and the associated methodology to create them is outside the scope of this report, it should be mentioned that the work to create these models was successful and done in a timely manner.

LESSONS LEARNED – DIMENSIONAL MODELLING

- Utilise industry standards methodologies, and ensure you have experienced data architects with healthcare or similar BI/DW experience on the team to guide the process.
- Ensure data modelling and data architect resources have experience in BI/DW systems.
- Utilise a professional Data Modelling or Data Architecture tool.
- The use of a reference data model such as MyHRDM increases the ability to integrate data, and to ensure concept consistency with the data collection systems.

DATA MODELS – LOGICAL AND PHYSICAL SMRP 2.0, PRIS 1.0

While the focus of the quality review process was primarily on the BI/DW analytic environment, reviews were also performed on SMRP 2.0 and PRIS 1.0 system artefacts including the Entity/Relation Logical and Physical models used to support the front-end application of the systems.

21 The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modelling, 3rd Edition, Ralph Kimball and Margy Ross, 2013

These models also followed the MyHRDM standard and as such aligned well with the MyHDW 1.0 BI/DW environment. Overall, the development was also done successfully, and in a timely manner.

EXTRACT TRANSFORMATION AND LOAD (ETL)

In BI/DW systems, perhaps the most significant amount of work (60 to 70% of the development effort) occurs in developing ETL processes to load the Data Warehouse. The Conceptual Data Warehouse Architecture above as well as the logical architecture shows the points were ETL activity occurs. This is primarily extracting data from source systems such as SMRP, transforming this and loading this into structured Data Marts. Furthermore, given the utilization of the Kimball, data Warehouse Life-cycle Toolkit methodology approach, certain key specification and design artefacts well utilised as they are incorporated into this. Most notably, Source to Target documentation, and design templates known as Kimball's 34 ETL subsystems.

In addition, the MIMOS Data integration/ETL tool Mi-Morphe was utilised by the project team for ETL development. Resources from MIMOS Berhad also made up the ETL developers on the project team. Overall, the ETL design and development has created a solid foundation for future work, and delivered what was specified on time for MyHDW 1.0

LESSONS LEARNED – ETL

- Utilizing an industry-standard methodology (Kimball life-cycle toolkit).
- Ensuring development resources have solid BI/DW ETL experience.
- The utilization of a professional ETL or data integration tool.
- The utilisation of an automated data warehouse testing tool (see testing section below).

In future phases and as appropriate to the type of data in query activity concerned, it may be possible to utilise non-RDBMS data sources for query purposes such as Hadoop, Document databases etc. This has the potential to reduce the development effort associated with ETL, and thus potentially reduce costs and development time. That said, it will still be necessary for a portion of the system to utilise structured data and in these cases ETL processes such as described above will be necessary.

TECHNICAL ARCHITECTURE

The physical design of the system is associated with a comprehensive server infrastructure installed in the MIMOS data centre. Running on this are several data management systems such as RDBMS, and Hadoop/HDFS configurations and services. A significant amount of Physical Design documentation associated with this has been compiled by the team, and was available for review. This comprehensively covers the following areas:

- High-Level Data Flows

- Hadoop Architecture, Hortonworks Data Platform 2.4
- Database Architecture - PostgreSQL Plus Advanced Server (PPAS) EnterpriseDB (EDB) V9.5.1
- Table Space and Storage Layout
- Physical Data Models
- Index Design
- Partition Design
- Summary Management
- Performance configuration
- Performance benchmarking
- High Availability configuration
- Initial disaster recovery (DR) discussions

As mentioned, there is a good deal of work and documentation associated with the technical architecture that was produced by the project team. This was reviewed as part of the consultant review processes and found to be an appropriate quality. Detailed description of lessons learned and technical details are outside the scope of this document.

SECURITY ARCHITECTURE

In terms of security architecture, the design of MyHDW 1.0 includes the notion of 'security and privacy by design'. The conceptual, logical architecture of the system includes a Privacy Assurance Service (PAS) which encapsulates core functionality to protect patient/client privacy through a variety of techniques such as anonymized, de-identification, data pre-processing, two-factor authentication etc. Its design and usage is aligned with the projects Information Security Management System and Plan.

APPLICATION ARCHITECTURE

No significant review was undertaken associated with the Application Architecture utilise by the Java development for SMRP 2.0 and PRIS 1.0 as this was outside the scope of the review process.

MALAYSIAN HEALTH DATA WAREHOUSE (MyHDW)

2015-2016 START UP: INITIATION

MIMOS BERHAD INFRASTRUCTURE, PLATFORM AND MyHDW SOLUTION

MIMOS Berhad Infrastructure, Platform and MyHDW Solution

INFRASTRUCTURE

As mentioned in the capacity building section above, establishing a secure, scalable and high performing infrastructure was a key part of the start-up activities of MyHDW. After consultation and discussions with a number of government agencies, it was determined that MIMOS Berhad would be a suitable location for this. Originally, the governmental data centre at PDSA had been identified to provide for this capability but due to some limitations and in discussion with MAMPU ICT Technical Committee (JTICT MAMPU)[22], it was agreed upon to site the data centre at MIMOS Berhad. This provisional arrangement has been set up with the notion of utilizing MIMOS Berhad as an 'incubator' facility aligned with its ITC innovation role. Once the MyHDW project has been established and stabilised, further discussions will occur regarding the long-term infrastructure arrangements. Currently, this is for a period of five years with the infrastructure purchased as part of this agreement remaining assets of the Ministry of Health.

Consistent with the above directions, and as part of the MyHDW 1.0 project a budget was approved for hardware procurement for the data centre. After consultation, it was agreed upon for practical reasons that MIMOS Berhad would manage the hardware procurement. Aligned with MIMOS Berhad directions, a comprehensive deployment of commodity servers was specified and procured and implemented in the current MIMOS data centre. In addition, consistent with MAMPU guidelines, connectivity was provisioned through 1Gov*Net[23] to allow network traffic in a secure manner to all MyHDW sites in KKM and non-KKM facilities such as University, Army and Private Hospitals..

An additional benefit and point of alignment in utilizing MIMOS Berhad for the infrastructure was its significant investment and development in technology 'platforms' related to Security and Big Data Analytics. Given the overall governmental direction to invest into local technology and infrastructure as possible, this seems synergistic.

In terms of project implementation during MyHDW 1.0 Milestone steps 3, 4, 5 and 6, infrastructure hardware and software platforms were specified, procured, installed and configured. This activity occurred the last quarter of 2015, consistent with the project plan, and to enable the availability of servers for the purposes of development and early testing. Utilization monitoring of the servers was undertaken and reports associated with shared with the project governance to understand level of usage. In addition, a substantive amount of testing and verification of the Production, Testing and Development environments was undertaken to ensure this work was successful and appropriately done. Associated with this work was verification performed by MIMOS Berhad, BPM as well as review by the project Consultant. BPM provided final independent review of User Acceptance and Final Acceptance Tests (UAT/FAT).

22 Malaysian Administrative Modernisation and Management Planning Unit (MAMPU). See : http://www.mampu.gov.my/web/en/jtict-eng

23 http://www.1govnet.gov.my/

Security considerations were also paramount during this implementation. To specify and enforce these, an Information Security Plan (ISP) was developed aligned with the formation of a Security Steering Committee and ongoing monitoring. This was further aligned with the recently published RAKKSSA 1.0, Public-sector Cyber Security Framework.

PLATFORM TOOLS-MIMOS BERHAD

A number of factors were considered in selecting the correct technology associated with analytic work in a Health Data Warehouse and BDA environment. As part of the original MyHDW Blueprint work, discussions occurred associated with selecting the correct technology once the project was initiated. In 2013 workshops were held in Kuala Lumpur to discuss criteria for product selection and to examine feasible options for MyHDW. These discussions included KKM and MIMOS Berhad staff and were facilitated by an international consultant. It was determined that time that in terms of pure functionality some of the best of breed commercial tools have the strongest capability. The types of tools examined were; Business Intelligence (BI), Extract, Transformation and Load (ETL), Data Integration and Relational Database Management (RDBMS) products. As part of this process platform products from MIMOS Berhad were included, as were a shortlisted selection of commercial products.

Following this activity Proof of Concepts (POC) and then subsequently a Beta release of SMRP were developed by MIMOS Berhad utilising their Platform technology and infrastructure.

In 2015 with the initiation of MyHDW 1.0 these factors were re-examined, when combined with; licensing costs, infrastructure alignment, security, subsequent ongoing refinement of MIMOS Berhad platforms and government directions for investment into local technology. It was determined that MIMOS Berhad platforms would be employed where possible in MyHDW 1.0. It was also recognised that any gaps between these platforms and best of breed technology from commercial vendors would be analysed and options considered to close potential gaps. In addition, if requirements could not be met completely these locally developed platforms would be supplemented as necessary with commercial products.

While not part of the initial evaluation of platform products, a Graphical Information System (GIS) was also included as an independent implementation in MyHDW 1.0. Furthermore, the MIMOS platform Mi-Harmony which is a product being developed to auto-code clinical data utilising Natural Language Processing (NLP) and the clinical terminology SNOMED CT is also being implemented as a separate product line in MyHDW 1.0.

In addition, at an administrative level within the MyHDW 1.0 project, the Mi-Doc, document management tool, was also utilised which allowed the storing and tracking of MyHDW documentation from 2011.

PLATFORM DEPLOYMENT IN MyHDW 1.0

The following diagram outlines the MyHDW 1.0 Logical Architecture overlaid with MIMOS

Platform technology and its utilisation in the project. This includes a wide range of capabilities including BI, ETL, Portal, Security, Content Management and Infrastructure tools. These are described below along with generic functionality and also any evaluation work done as part of MyHDW 1.0.

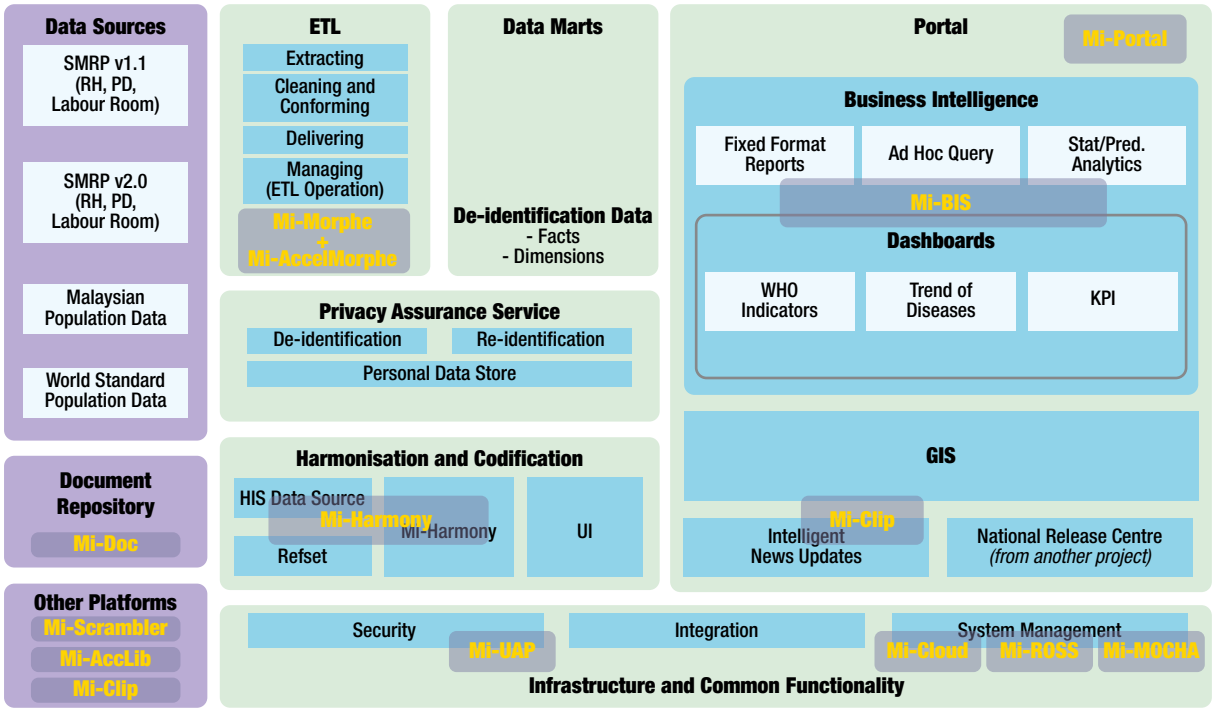


Diagram 9: *MIMOS-Platform Tools Associated with MyHDW 1.0 Based on Logical Architecture*

ENTERPRISE PORTAL: Mi-PORTAL

As a part of a number of Presentation Layer components as described by the Conceptual Architecture above, an Enterprise Portal is required as an overarching container for all presentation services, query tools and access control. MyHDW 1.0 utilises the MIMOS platform product Mi-Portal. This also is integrated with a number of other platform products such as Single Sign-On (SSO), User Access Control etc. As part the project, a user experience team (UI-UX Team) provided input to further refine the look and feel and user friendliness of this environment. Images from the current MyHDW 1.0 Portal pages are shown in diagram10.

BUSINESS INTELLIGENCE (BI)/ ANALYTIC TOOL: Mi-BIS

In analytic and data warehousing environments it is common in recent history to query data stored in a relational database management system often specifically designed for analytic activity, utilizing a query or reporting environment known as a BI Tool. Example commercial tools of this nature might include Microstrategy, Tableau, Cognos, Business Objects etc. These are characterised by the ability in a graphical point-and-click environment to query and interact with data in a high performing and feature rich environment. Normally these are implemented

in the front-end or presentation layer of a BI/DW environment. Interestingly the term Business Intelligence was more related to the notion of garnering 'intelligence' from a business's data to allow it to better inform its decisions. Although it is now more commonly associated with the reporting tools themselves.

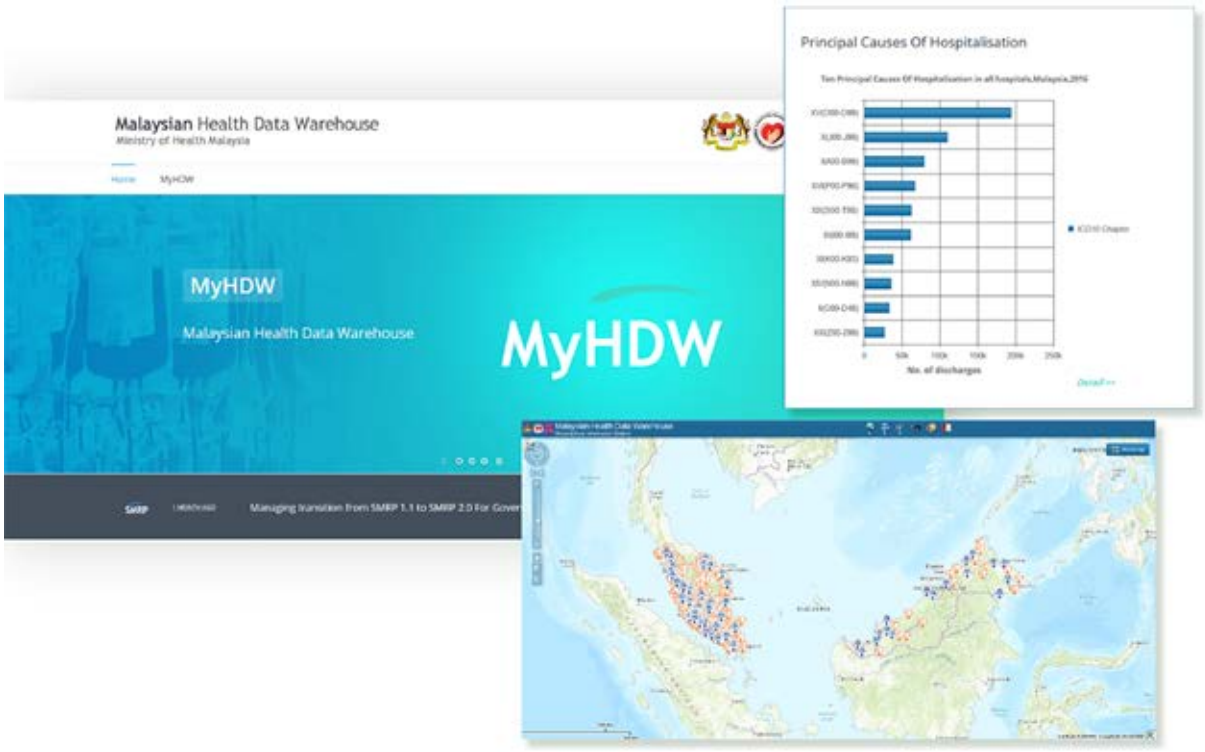


Diagram 10: *MyHDW 1.0 Portal Screenshots*

For MyHDW 1.0 the MIMOS platform Mi-BIS is utilised. With the evolving landscape associated with next-generation analytic products aligned with the paradigm shift into Big Data Analytics (BDA), new tools are emerging associated with querying non-structured and semi structured data. In addition, the differences between query tools associated traditionally with statistical analysis and BI tools is becoming blurred. The MyHDW Blueprint identified the need for both BI tools and statistical tools. That said, at the time the Blueprint material was written these were clearly different. As part of discussions that occurred during the MyHDW 1.0 implementation, the need to understand the differences between BI tools and Statistical tools and how they are used in healthcare analytics became apparent. To assist in understanding, material was produced, and shared with team members and the governance.

DIFFERENCES IN USAGE BETWEEN BI AND STATISTICAL TOOLS

As mentioned historically speaking, there are two core types of query tools; BI and Statistical. Each have a distinct set of features and orientations. Experience in long-term deployments in Healthcare Data Warehousing and associated analytic environments highlight differences in the ways certain categories of 'power users' utilise query tools. Two critical features observed from these real-world examples are that:

- It is difficult to deploy Statistical Tools to remote sites as often these have had historically, limited administrative and security features. In addition, the tools are complex to use and as such require a good deal of training and support.
- Those users from a research or statistical background may resist the usage of BI Tools and this has often led to change management issues and limited uptake in these types of environments.

A summary comparison between the two types of tools can be found below:

BI Tools	Statistical Tools
Good ease of use	More complex to use
Queries are primarily built by a point and click method	Queries are primarily built using coding techniques. Though some point-and-click interfaces are available.
Data is mostly accessed via RDBMS tables commonly in Dimensional Models or Star Schema	Data is commonly accessed through proprietary datasets though some RDBMS access supported.
Supports Production Reporting	These tools have very good flexibility due to the use of coding but for the most part are less productive than BI tools
Supports Dashboards	
Often supports Data Exploration	A full range of statistical functions including those required to support predictive analysis are available
Supports Ad-hoc/Self-Service Query	
Modest support for statistical functions	Support for Ad hoc/Self-Service query
Supports strong administration, security and audit	Support for tool administration, security and audit is often limited
Is easy to deploy at remote sites	Deployment to remote sites not always feasible due to the point above

Table 4: *Comparison of BI Tools vs Statistical Tools*

As mentioned a new class of tools is emerging associated with BDA and paradigms that extend this. These will be further examined in *Moving-forward: MyHDW Phase 2 and Beyond* section below.

HISTORY OF Mi-BIS

An evaluation of Mi-BIS 1.2 was originally undertaken in June 2013 in comparison with a number of other commercial and Open Source BI Tools. This was done under the auspices of a MyHDW workshop carried out at that time. One item that was noted regarding Mi-BIS 1.2 was the fact that the tool was implemented using a Client/Server model and lacked a web interface. The availability of a web interface greatly simplifies the deployment and support of the tool at remote sites. Subsequent to this a web implementation version was developed and this was utilised during the MyHDW Beta Release. This release also allowed for continual evaluation of the Mi-BIS platform. In addition, as part of this process, an ongoing dialogue with a number of BI tool vendor's continued to ensure a good understanding of the key features necessary in these types of products for a successful deployment of MyHDW.

Mi-BIS USAGE IN MyHDW 1.0

Mi-BIS V1.6 for MyHDW7 is proposed version that will be used in the development and the first live deployment of MyHDW 1.0. The tool supports basic requirements for Fixed Format Reports (Standard Reports), Dashboards, and Ad-hoc Query capability. The underlying technology for much of the Mi-BIS platform is based on the Open Source BI tool, SpagoBI Competency Centre. The most recent version of this tool is V5. In addition, a custom built Java application has been written to better support Ad-hoc query, and also a graduated implementation of the open source statistical platform R has been integrated into the Platform.

As part of the review process in MyHDW 1.0, a Business and Technical Options Analysis was carried out to look at opportunities to supplement Mi-BIS with the use of a commercial tool or alternatively accelerating product development for Mi-BIS. This was to allow additional functionality such as Data Exploration, interactive statistical analysis and advanced use of functionality to be incorporated into the presentation layer to supplement Mi-BIS capability. It was determined that for Phase 1 any additional commercial tool were not be procured as the basic requirements as specified were met by Mi-BIS. That said this may be considered for the next phase.

STATISTICAL AND PREDICTIVE ANALYTICS

The MyHDW Blueprint and subsequently MyHDW 1.0 recognizes the need for statistical tools including those that can do predictive analytics. The scope of MyHDW 1.0 while not fully implementing these at this stage included some functionality associated with statistics in its scope. This functionality will be delivered by Mi-BIS through its integration with the Open Source statistical product R. As mentioned above, the distinction between BI tools and statistical tools is becoming less apparent as demonstrated by both types of functionality being available in Mi-BIS.

ETL AND DATA INTEGRATION TOOL - Mi-MORPHE

In traditional BI/DW environments data is commonly extracted from source systems, transformed for analytical purposes and loaded into a relational database (RDBMS). This process known as Extract Transformation and Load (ETL) is a standard part of systems development and in mature environments involves utilisation of ETL or data integration tools. Example commercial versions of these tools are Informatica, Ab Initio etc. For the purposes of MyHDW 1.0, MIMOS Berhad platform technology Mi-Morphe is utilised.

Mi-Morphe is primarily based on the open source BI/ETL tool Pentaho. The current version used by the product is Pentaho Data Integration (Kettle) 5.4. Mi-Morphe extends the basic Pentaho Data Integration features by for example adding the following:

- Job Scheduling
- Administration

- Custom plug-ins. A number of these utilise in-house developed accelerator technology utilizing GPUs. For example, de-duplication, scrambling functions etc.

ETL development is commonly the most expensive and time-consuming part of building and maintaining Data Warehouses. With the paradigms shift into BDA the ability to reduce project costs associated with ETL and in addition make data available in a timelier fashion is becoming a reality. Primarily through the utilisation of the Hadoop ecosystem and techniques associated with using data in its native and transform state such as in the MyHDW Landing Zone which is similar to the notion of a Data Lake[24].

RELATIONAL DATABASE MANAGEMENT SYSTEM (RDBMS) AND HADOOP

The database management system use of structured data in MyHDW 1.0 is PostgreSQL Enterprise Edition. This commercial version of an open source RDBMS provides a comprehensive set of features including High Availability. To support future expansion into BDA capability such as in Phase 2 and beyond MyHDW 1.0 architecture also includes the use of a Hadoop/HDFS environment. This utilises the Open Source platform HortonWorks.

GEOGRAPHIC INFORMATION SYSTEM (GIS)

Currently the GIS capability the MyHDW is implemented as a standalone system loaded with Oral Health data. It is not incorporated fully into the MyHDW architecture at this stage other than being part of the Presentation Layer. The technical underpinning for the GIS system is based on ESRI ArcGIS Server. Future phases of MyHDW will more fully incorporate GIS capabilities into its architecture.

NATURAL LANGUAGE PROCESSING (NLP), HARMONISATION AND CODIFICATION SERVICES - MI-HARMONY

One of the challenges in Healthcare Analytics is acquiring high quality health data as part of the clinical workflow. The advantages of doing this is both in terms of reducing the burden of data collection and also to ensure high-quality data based on the original clinical information recorded by the healthcare provider. Traditional models associated with data collection often involve the manual classification of diagnostic and intervention information into standards-based classifications such as for example ICD-10. With the advent of improvements in natural language processing and the advancement of BDA techniques including semantic data integration (Semantic Web) the ability to obtain precise meaning from clinical notes in semi or unstructured

24 A **data lake** is a collection of storage instances of various data assets additional to the originating data sources. These assets are stored in a near-exact, or even exact, copy of the source format. The purpose of a **data lake** is to present an unrefined view of data to only the most highly skilled analysts, to help them explore their data refinement and analysis techniques independent of any of the system-of-record compromises that may exist in a traditional analytic data store (such as a data mart or data warehouse). See : http://www.gartner.com/it-glossary/data-lake/

form becomes possible. This has the potential of not only realizing efficiencies as mentioned but also allowing greater precision associated with the underlying data. This latter point relates to the additional amount of meaning in raw data as opposed to that that has been classified. The analytic component of Mi-Harmony was able to demonstrate higher precision in identifying Cardiac cases as compare to equivalent queries utilizing structured data and classification approaches associated with ICD-10.

Mi-Harmony development started in 2013 after Malaysia's membership into IHTSDO[25]. It utilises Natural Language Processing (NLP) guided by SNOMED CT RefSets to automatically produce SNOMED codes and other means to Harmonize and Codify data. In MyHDW 1.0, this will be implemented as a standalone product in three facilities that have Cardiology Services. The initial version of Mi-Harmony has a comprehensive set of Cardiology RefSets data. It is planned that future phases of MyHDW will extend RefSet development to include all Cardiology and Cardiothoracic terminology. Furthermore, it is being refined to handle more complete post coordination and contextualization in alignment with SNOMED CT guidelines.

In terms of providing metadata to Mi-Harmony for its processing capability ontology's were created associated with the mention RefSets. Utilization of Mi-Harmony in the future state MyHDW architecture will also contribute towards semantic data integration capabilities. The Mi-Harmony Platform was presented to SNOMED CT conferences in 2014 and 2015.

The schematic of the Mi-Harmony platform architecture shown below:

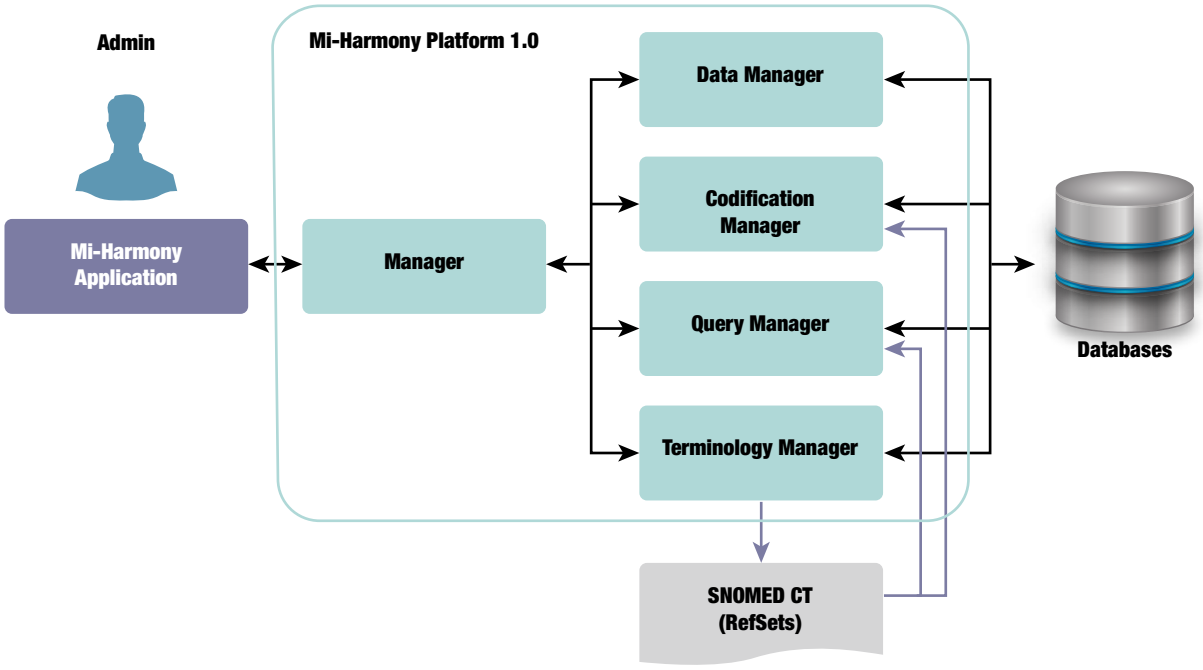


Diagram 11: *Mi-Harmony Logical Architecture*

25 SNOMED CT (Systematized Nomenclature of Medicine -- Clinical Terms) which is a standardized, multilingual vocabulary of clinical terminology used by physicians and other health care providers for the purpose of electronic exchange of clinical health information. See : http://www.snomed.org/

MyHDW SOLUTION

Building and maintaining an enterprise wide health analytic capability or Healthcare Data Warehouse is a complex endeavor. MIMOS Berhad in addition to developing a number of Platform technologies as described above also has substantial ICT processes and resource capabilities which have been used as the underpinning to provide a broader MyHDW Solution or Practice Model. These capabilities include for example System Development Life-Cycle, Project Management Lifecycle, Development Standards and Quality Assurance processes. To some degree these are being refined and further developed through the operation of the MyHDW 1.0 project. This is allowing both the capacity development, lessons learned from international and industry experience and program and process development to allow a full BI/DW solution to be provisioned. Furthermore, the MyHDW Solution has the potential to be generalised and implemented outside of the healthcare environment.

MALAYSIAN HEALTH DATA WAREHOUSE (MyHDW)

2015-2016 START UP: INITIATION

TESTING AND QUALITY ASSURANCE

Testing and Quality Assurance

As mentioned previously in this report the development team associated with this project, while very experienced in software development including testing, generally had less experience in testing activities associated with BI/DW initiatives and particularly those associated with the healthcare domain. To address this risk material was collated associated with successful strategies and approaches in testing BI/DW initiatives with a particular focus on healthcare. This was based on industry best practice and international experience. Briefing material was prepared by the project consultant in consultation with the QA and testing team at MIMOS Berhad. Information sharing sessions were instigated and this material was presented and discussed. A summarise version of this is presented below for reference purposes.

TESTING APPROACHES AND STRATEGIES IN BI/DW INITIATIVES

Testing and Quality Assurance are key activities within BI/DW projects and programmes. It is often surprising, to those who have not undertaken the development of these types of projects before just how much resource needs to be allocated to these tasks and the type of resource needed.

Broadly speaking these two terms can be defined as follows:

- **Testing** is the process of verifying and validating that a software program works as expected.
- **Quality Assurance (QA)** is a set of activities designed to ensure that the development and/or maintenance process is adequate to ensure a system will meet its objectives.

A key aspect of testing activity in BI/DW initiatives is related to data verification. While traditionally testing activities are associated with 'software testing' for the purposes of these types of projects, this activity also includes data verification as part of the testing process.

QA OBJECTIVE FOR MyHDW:

The working definition for MyHDW defined originally in the 2011 Blueprint document is: *'A trusted source of truth of comprehensive healthcare data structured for query and analysis purposes'*. In terms of QA, the notion of 'trusted source of truth' is particularly germane in this context. If data and information is to be used to inform decisions that affect the health system, it is critical that it is reliable, accurate and considered a credible source by those who use it. Hence, it is necessary to have the checks and processes in place to ensure this is the case and that evidence of these have visibility if need be to the end user community.

With this in mind, the following QA objective for MyHDW is proposed:

To foster the highest levels of trust in MyHDW as a Source of Truth of Healthcare Data and Information.

To achieve this objective, it will be necessary that processes be in place for each stage of the development life-cycle and also in future maintenance stages. Specifically in development, this would encompass:

- Data Acquisition/Collection.
- ETL/Data Integration.
- Database/Data Storage.
- Data Access and Presentation.

DATA WAREHOUSE/BI TESTING – DIFFERENTIATORS

As intimated DW/BI QA and testing have a number of unique facets which differentiate them from standard software development approaches. These have been observed both in industry practice and also based on international experience in building these types of systems in the healthcare arena. The following outlines these key differentiators:

- Testing is heavily data centric and complex. This complexity is present in all DW/BI projects but is emphasised in the healthcare subject area due to its intrinsic complexity.
- QA/Testing is time consuming and requires significant resource allocation. For example, approximately 60% of the resources assigned to ETL development/design resources would be commonly allocated to QA/Testing.
- Specialised skills/resources are required to execute data validation and verification process. This is especially the case with Healthcare which is mentioned is particularly complex.
- Data loss during data integration process or in the movement from source systems.
- 'Dimensionalising' Data i.e. the movement of data into a Star Schema can introduce many hard to detect errors.
- Often testing needs to be done on production volumes to uncover certain errors.
- BI semantic layer/metadata can add complexity to the testing approach. BI semantic layer is the meta data that is often defined in BI Tools and is used to define how the tool will interact with a backend data source.
- QA Automation is less well supported than in traditional applications.

PRINCIPLES OF EFFECTIVE DW/BI QA

The key principles for effective QA to a greater degree, is standardized across the IT industry. Starting with these and applying the unique aspects of DW/BI the following are identified as the key principles for setting up an effective QA and testing approach:

- The quality of an effective test depends on the documentation completeness and accuracy in terms of the Business and System Requirements and Project Planning material.
- Testing of DW/BI systems is as mentioned, data centric. To be successful, testing will need to utilise both real data and mock data to reproduce the most common error situations that are encountered in ETL.
- The testing team should include a variety of technical, SME and domain expert roles. This virtual team should be identified during the project planning phase.
- QA as part of the data warehouse life-cycle, should be aligned with system design. The test phase should be planned at the beginning of the project. This will allow the specification of testing goals and quality levels and the settings of associated expectations.
- End-to-end traceability should ideally be incorporated into test plans and strategy.
- Key Performance Indicators (KPIs) and benchmarks along with the associated criteria for QA certification should be agreed upon in an explicit way with the executive sponsors of the project.
- Evaluate and consider acquiring a DW/BI Automated testing tool.

MyHDW CONCEPTUAL ARCHITECTURE IN RELATIONSHIP TO TESTING

Each aspect of the MyHDW Conceptual Architecture will need to be tested to cover an end-to-end scenario. The greatest volume of testing will occur in the ETL processes between Operational/ Staging and Analytic Environment. Appropriate resource management in terms of availability and type of resource will be particularly necessary during this stage as well as the use of automation. The following are illustrative tests and approaches for each part of the architecture:

Data Acquisition

- Data profiling: Data profiling activity helps in understanding the nature of source data. A significant degree of risk in ETL processes is unexpected data values and conditions in source data. Data profiling early on in the project help ameliorate this risk.

Data Architecture and Integration

- Validating the Source to Target Mapping: Ensuring traceability throughout.
- Validating the Data Model: This involves validating data structure with the requirements specifications.
- Check for alignment with MyHRDM.

Operational/Staging: ETL testing is probably the most complex and critical testing phase, because it directly affects the quality of data

- ETL testing.

Database/Storage

- Validating data loads.
- Performance and Scalability: Database and file service stress and performance testing.

Presentation Layer

- Report and dashboard layout validation as per mock-ups and data validation as per requirement specifications.
- Ad-hoc query validation as per specification.
- GIS reporting as per specification.
- End to End Testing.

TESTING OVERVIEW FOR MyHDW 1.0

The following figure shows the overarching testing aspects implemented as part of the MyHDW 1.0 project including test phases and types, test process, artefacts, testing tools and team composition. Of note is the utilization of an independent testing team.

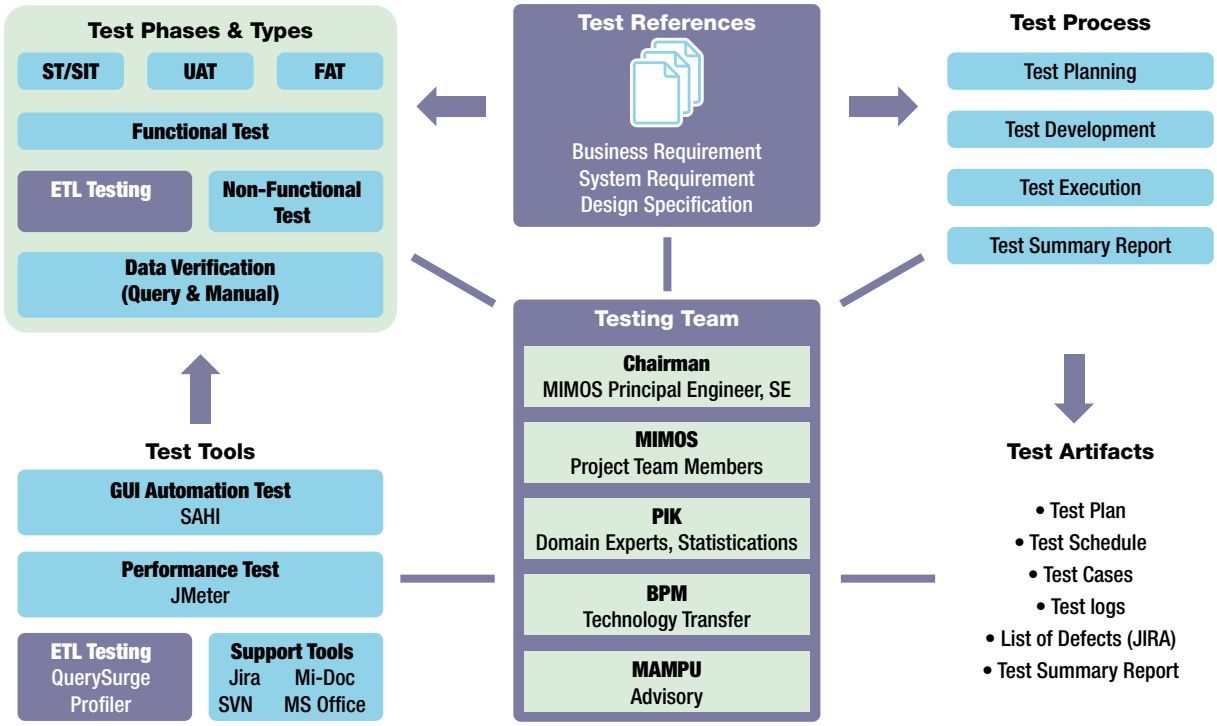
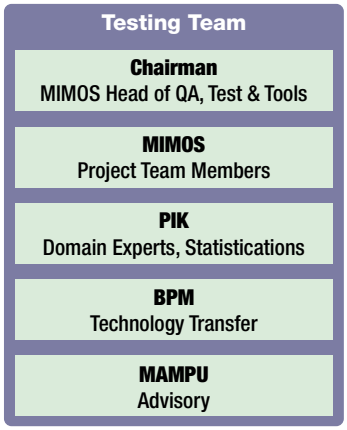


Diagram 12: *MyHDW Testing Overview*

INDEPENDENT TESTING TEAM

An independent testing team was instigated during Q1 of 2016. This was in recognition of the critical role associated with the credibility and quality of the data within MyHDW and processes and accountabilities associated with this. The team composition and keep principles shown in the figure below:



Key Points

- Independent Committee & Team
- Consists of IT Technical & Domain expert members
- MyHDW as a Health Analytics Environment needs to credible and considered a Source of Truth for Malaysian Healthcare data
- Testing is a critical project activity and resource intensive
- Testing is related to both Functional and Data Verification aspects

Diagram 13: *Independent Testing Team*

TEST PROCESS OVERVIEW

In line with normal software testing processes, a series of steps were developed associated with the overall testing pathways for the project. In addition to the normal types of activities found in application development initiatives associated with functional testing these also include data verification steps.

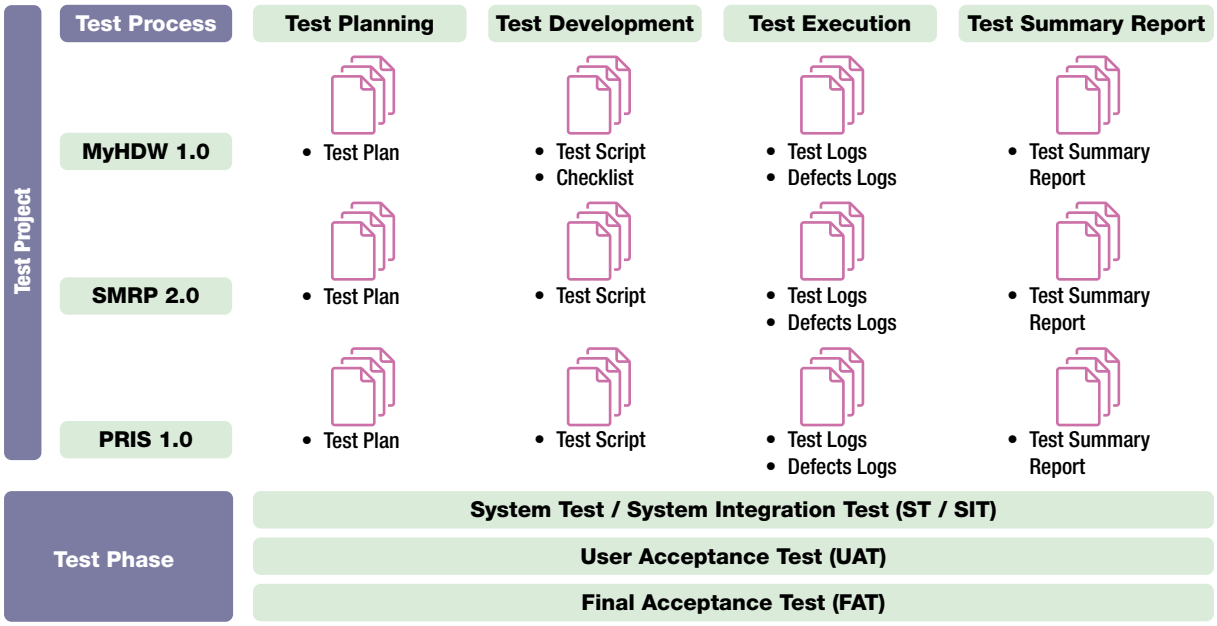


Diagram 14: *Test Process Overview*

DATA VERIFICATION METHODOLOGY

As mentioned data verification was an important part of the testing process of MyHDW 1.0. Based on international experience in testing and verifying healthcare data in similar environments the following scheme was developed by the Testing Team.

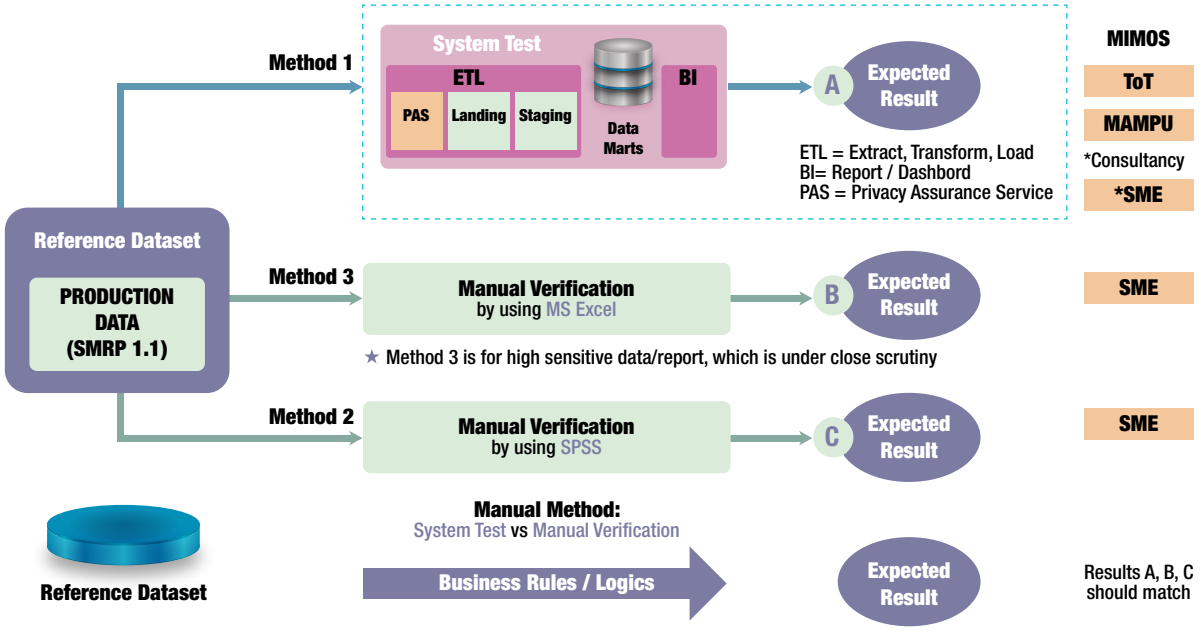


Diagram 15: *Data Verification Methodology*

AUTOMATED BI/DW TESTING TOOL:

As part of the initiation activities for MyHDW 1.0 it was recommended that investigation and evaluation be done regarding the acquisition of automated testing tools for data warehousing situations. Based on this evaluation a testing tool was selected and implemented for use in the project. This tool allows the automation of testing ELT code and data verification in BI/DW and BDA environments.

QUALITY ASSURANCE

In parallel to the testing activities described above, a number activities were initiated to support overall quality of MyHDW and to ensure that it ultimately meets its objectives. These quality assurance processes include:

- Quality review and guidance by project consultant
- Establishment of an initial Master Data Management (MDM) program.

- Formation of Data and Information Governance Steering Committee and program of policies and guideline development and approvals.
- MIMOS Berhad internal IT processes and standards.
- Document Management Standards-utilisation of Mi-DOC document management tool to store and manage all project documentation.
- Specialised product review such as the BI Tool evaluation and options analysis carried out during MyHDW 1.0.
- Development of a comprehensive Project Glossary to ensure consistency of terminology across team members.
- Ongoing monitoring of testing processes and traceability.

LESSON LEARNED – TESTING AND QA

Testing and Quality Assurance for this project was a highly complex and important stage. The following is a summarisation of the key lessons learned that may benefit future phases of this initiative or similar projects:

- As identified in the initial consultant planning discussions with the testing team, testing for this type of initiative required specialised resources and sufficient allocation associated with these to complete all the work. The resource allocations used in this project should be used as a benchmark for future initiatives.
- The Data Verification Methodology developed during the project is an effective means to ensure high-quality data.
- Data Verification testing as identified in the early planning stages would best be incorporated into the normal project testing steps such as UAT, SIT etc. as opposed to a separate activity.
- ETL testing should also be included into standard project testing steps such as UAT, SIT etc.
- Utilization of automated testing tools for ETL testing has been an effective process and should continue for future projects.
- A separate additional testing step will be required to test MyHDW with production data after SMRP 2.0 data is live and stable. While the use of mock-up data is useful and necessary, it is also required that the full breadth and depth of production data be utilised in a final testing step as described.
- Separate testing steps will also need to be applied after each load of historical data. These tests can be considered one-offs.
- The utilization of an Independent Testing Team has been a useful and necessary approach.

MALAYSIAN HEALTH DATA WAREHOUSE (MyHDW)

2015-2016 START UP: INITIATION

DATA SOURCES AND ACQUISITION

Data Sources and Acquisition

As was outlined in the original Blueprint report, a number of key principles need to be followed to ensure that data is collected and acquired in a manner that is both cost efficient, accurate and with the right degree of timeliness. Furthermore, this process should align with the clinical workflow as much as possible so as to ensure minimal impact to healthcare providers while capturing as much high quality data as needed. For example, it is desirable that atomic or granular data is collected in preference to aggregate data. The latter is limiting in terms of its ability to support a broad range of analytics. Moreover, duplication of data entry should be avoided in alignment with the principle stated above associated with efficiency. Overall it is important that both primary usage and secondary usage is aligned in Digital Health systems design to facilitate this.

Consistent with the original Blueprint portfolio plan MyHDW 1.0 will develop an improved version of an acute service data collection system (SMRP 2.0), initially beginning with Inpatient and Daycare, in addition to laying the foundation for a generalised registry data collection system (PRIS 1.0) beginning with Cancer data. Population Census data is also utilised and integrated into the BI/DW environment. In addition to these systems developed within the MyHDW Conceptual Architecture, a wide range of other source data will be potentially used by the system as determined by the scope of future phases. Overall Hospital facilities visit data reflects the following categories of services:

- Inpatient
- Daycare
- Outpatient (Specialist Clinic, General Clinic, Acute and Emergency)
- Traditional and Complementary Medcine
- Procedure
- Population and Disease Registry and Clinical Support (Physiotherapy, Occupational Therapy, Speech Therapy, Audiology, Diet, Social Welfare, Radiology, Pathology Laboratory, Nuclear Medicine and Forensic).

Furthermore, in the Public Health domain, the following are categories of services provided;

- Oral Health
- Family Health
- Family Planning
- Water Quality
- Rural Water Supply
- International Port Entry
- Food Quality
- Food Safety

- Pharmaceutical
- Human Resource
- Health Expenditure
- Genomics
- Facilities and Services
- CASEMIX
- MyHIX
- HIS/TPC-OHCIS

In terms of the frequency by which data is captured to some degree, this is determined by how it will be used in analytic products and downstream business processes. Despite the ever increasing trend to capture data in real-time, sometimes this is not needed and maybe even not optimal. That said, there is certainly benefit in building data capture that services both primary and secondary usage into normal workflows so that all points of data capture are setup both for functional clinical usage and also analytic. In terms of domains which will regularly need Near or Near Real-Time data these would include Disease Control and Disaster Management. In MyHDW, the Health Information Framework (MyHIF) is one means by which the timing associated with reports and/or alerting is specified. In addition, BDA techniques allow for timely data capture primarily due to the utilisation of original source data from the Hadoop Landing area without the need of significant ETL.

In terms of enriching the MyHDW datasets with supporting data, partnerships with interagency groups such as Department of Statistics, Department of Environment and Department of Meteorology will be instigated to ensure these data sources are available in a timely manner to be integrated into MyHDW and thus potentially increase the return on investment and value of the source of truth data and associated analytics.

MALAYSIAN HEALTH DATA WAREHOUSE (MyHDW)

2015-2016 START UP: INITIATION

DATA LOADING, MIGRATION AND INTEGRATION

Data Loading, Migration and Integration

It is common practice in BI/DW environments associated with structured data to load the newly developed environment in a stepwise fashion normally beginning with historical data associated with the subject area of concern. This is particularly needed to support trend analysis. Following this, additional data sources to enrich analytics are added for example Master Data such as Diagnostic codes, Population Census statistics etc. Finally, live transactional data begins to be added to the environment.

In addition to the above, there is a good deal of complexity in MyHDW 1.0 associated with the interdependency between the new SMRP 2.0, PRIS 1.0 systems and the MyHDW analytic environment. It is ideal to have stable source data before building a structured BI/DW environment. In MyHDW 1.0 this wasn't possible due to project considerations so a number of strategies have been adopted to allow the building of this, while these new collection systems are being developed. Details concerning the historical loading of data, how data will be migrated into the new system and also information around integrating data from data capture systems and MyHDW 1.0, are summarized below:

- **Historical Load of 2012-2016 SMRP 1.0/1.1 Data**: Historical data for Inpatient and Daycare will be loaded from SMRP 1 from 2012 to 2016 Q2. Estimated Q3 2016.
- **Historical/Live Load of 2016 SMRP 1.1 Data**: SMRP 1.1 2016. Estimated Q1 2017
- **Historical data from 1999 to 2011** will be available offline for analytical purposes.
- **Live transactional data from SMRP 2.0**: Estimated start Q1 2017. Complete year dataset ('Closed Year'). Estimated Q1 2018.
- **MyHDW 1.0 features 2 Iterations** to accommodate integration and migration issues as mentioned above.
- **Iteration 1**: is associated with loading the BI/DW environment with SMRP 1.1 data and creating a set of reporting products, Fix Format reports and Dashboards associated with this.
- **Iteration 2** is associated with loading the BI/DW environment with SMRP 2.0 data. The above set of reports are updated to reflect the new data, new Fix Format and Dashboard reports are added, also functionality for Ad-hoc Query and Statistical capability including Predictive analysis.
- **SMRP 2.0** adds additional variables and security features on top of SMRP 1.1.
- **Cancer** visit data from SMRP 2.0 comprises what is known as PRIS 1.0 Module 2B functionality. Further development of PRIS 1.0 will extend this.
- **Integration**: This project stage includes all interdependencies between SMRP version is and in addition that necessary to include non-Ministry of health facilities such as University, Army and Private Hospitals.

- **Externally facing data sharing and integration**: this includes, National Registration Department on Notification of Birth and Death. Also Public Health Programmes on iKelahiran.

The above steps are currently managed as part of a comprehensive roll-out plan.

MALAYSIAN HEALTH DATA WAREHOUSE (MyHDW)

2015-2016 START UP: INITIATION

DATA AND INFORMATION GOVERNANCE

Data and Information Governance

GENERAL CONSIDERATIONS

A key aspect of BI/DW systems and particularly that associated with healthcare is Data and Information Governance. This is a broad topic and its applicability is highly context sensitive. In the case of healthcare there are sensitivities and legislative constraints around access to data, privacy, appropriate use of data, data quality etc. In other domains this list may be different. Experience has shown data governance is a vital aspect for the success and long-term viability of a Healthcare Data Warehouse. The MyHDW Blueprint report certainly recognised and made visible the need for this to be a well-established function within a program that would run and maintain MyHDW.

Outside of healthcare there is a body of knowledge associated with Data and Information Governance. In terms of setting broader context, it is recommended that we draw on this best practice and adapt to a healthcare setting rather than reinventing the wheel. While healthcare is perhaps one of the more complex types of systems, Data and Information Governance issues are relatively similar across industry best practice. One Data and Information Governance definition is as follows:

Data and Information Governance is the exercise of authority and control over the management of data as an asset[26].

Data and information governance is implemented in projects and programmes via a variety of means. These may use of a governance structure and a clear set of accountabilities. Commonly this would establish or define practices and policies to ensure that data and information is:

- **Secure** - That only people who are authorized to can access data and information.
- **Accessible** - that data and information is accessible by those who need it, are authorized and that the correct format and tools are available to them.
- **Credible, Data Quality** - that data is reliable, valid, trusted, and produces the results that are expected.
- **Monitored and auditable** -that there is a clear understanding how data is being used, by whom as well as an understanding of the lineage of the data.
- **Consistent and efficiently managed** - This includes how analytic and reporting products and methodologies are standardised such as Health Indicators, Benchmarks, Key Performance Indicators (KPI) etc. Also how data is standardised in terms of consistent code usage, classifications and terminologies.

26 DMBOK Data Management Body of Knowledge. See: http://www.dama-dmbok.org/

HEALTHCARE SPECIFIC CONCERNS

Data and information governance is of particular importance in a healthcare setting. This has been seen by similar initiatives across the international landscape and a number of lessons can be drawn from these examples. Common focus areas in this domain are; appropriate access to data, privacy and security concerns and controls, maintaining good data quality, consent and system and standards development alignment. It has been found also useful to make a distinction between how data is used and specifically whether it is for direct clinical care (**primary usage**) or for **secondary usage**. This latter term refers to any usage in which data collected for direct clinical purposes or perhaps as part of the specific collection activity, is used for non-clinical analytics and reporting. This distinction is made as there are different sets of legislative controls and privacy considerations between these two types of data. A list of these types of secondary usage categories are shown in the table below.

Secondary Use Category	Description	How Used
Health System Management	Improve the efficiency and effectiveness of the health system	Cost management Strategy, planning and policy development Resource allocation, priority-setting and funding Resource utilization System performance reporting
Research	Insights in regards to improved medical treatments and programs of care, and better understand the performance of the system and health of the population	Clinical research studies Comparative effectiveness and evaluation Retrospective analysis of policies, initiatives and interventions Population research Modeling and simulation
Surveillance	Provide data on certain disease/events which leads to preventive and control activities Evaluation of public health programs/ practices Alert on potential outbreaks/risks	Disease Surveillance Public health education Public health reporting
Program Management	Improve delivery of clinical care and evidence-based best-practice	Quality control and improvement Disease management Patient safety initiatives Monitoring access to care

Table 5: *Secondary Data Usage*

MyHDW DATA AND INFORMATION GOVERNANCE

In recognition of the importance of this topic, MyHDW 1.0 instigated a number of activities to advance this type of governance aligned with project objectives. Approved by the project Technical Committee, a Data and Information Workshop was held in Kuala Lumpur March 2 to March 4, 2016 involving a broad range of senior healthcare stakeholders. Its objective was to get input from this group and better understand the specifics of healthcare related data and information governance in Malaysia. Participants came from the following organisations and groups:

- Data consumers: public, KKM (MHP, NIH, programmes), power users, super users
- Source data system (SMRP, PRIS) users
- Data owners of PRIS
- Agencies (JPN, DOSM, Met Dept., JAS)
- International Bodies (WHO, UNICEF, UNHCR)
- Data manager (PIK/HIC)
- Data processor (MIMOS Berhad)

The agenda for the workshop broadly covered the topics below:

- An Introduction to Data Governance in Healthcare
- Access
- Privacy
- Security
- Use of Data and Information
- Data Quality
- Consent

This was a productive and thoughtful session and gave a good opportunity for healthcare stakeholders to share their views and concerns related to data and information governance associated with MyHDW and Secondary Use Data. The workshop employed a breakout group format and presentations were done following extensive discussion on the topics above.

DATA AND INFORMATION GOVERNANCE STEERING COMMITTEE

As a follow on from the above the MyHDW Data and Information Governance Committee was formed and its inaugural meeting took place at the end of June 2016. This meeting saw the approval of the steering committee terms of reference in addition to general principles for its

operation. The committee will be responsible to the Pusat Informatik Kesihatan of KKM and is chaired by TKPK[27] (P&ST KKM). In summary the approved Terms of Reference for this committee Identifies the Following Responsibilities:

The establishment of policy and guidelines for:

- Data Access Management
- Master Data and Reference Data
- Use of Data and Information
- Consent and Data Ownership

The establishment of policy and framework for:

- Data and Information Quality
- Data and Information Privacy[28]
- Policy on Data and Information Management
- Data retention and data archival policies

The membership of the MyHDW Data and Information Committee is as follows:

Chair: TKPK (P&ST KKM)

Members:

- KKM CIO
- Director of Planning
- Telehealth Unit (UACP owner)
- Representative from Registry owners (PRIS data provider)
- Representative from Medical Programme (SMRP data provider)
- Representative from Disease Surveillance
- Malaysian Healthcare Performance
- National Institute of Health (NIH)
- Malaysian Administrative Modernisation and Management Planning Unit (MAMPU)
- Technical Lead for Rangka Kerja Keselamatan Siber Sektor Awam (RAKKSSA)

27 Timbalan Ketua Pengarah Kesihatan (Deputy Director General of Health)

28 Data Security governed a separate Security Committee guided by RAKKSSA Version 1.0, April 2016.

- Department of Statistic Malaysia DOSM)
- Representative from Finance
- KKM Legal Advisor

Secretariat: PIK

The priority drivers for the committee's work include legislation and sensitivities associated with healthcare data and particularly personally identifying information (PII), work associated with the RAKKSSA and its implementation through the associated Security Committee. In addition to specific requirements of certain health care registries. Longer term there will be a broad requirement for good data and information governance and this committee is well-positioned to guide, monitor and have accountability for this.

In establishing Data and Information Governance in the start-up phase, MyHDW has achieved a major milestone in terms of system and operating maturity.

MALAYSIAN HEALTH DATA WAREHOUSE (MyHDW)

2015-2016 START UP: INITIATION

PRIVACY AND SECURITY

Privacy and Security

Health data in MyHDW refers to Secondary Use Data, which is used for making policy and Health System Management. Health data is highly sensitive. Sensitive data is when personally identifiable information is linked to clinical information. Some requests specifically refer to data and others refer to information generated from such data, and data that are linked to another set of data to provide more value-added information.

Malaysian legislation stipulates key control around privacy and confidentiality of personal data which includes Personal Data Protection Act 2010 (Act 709) (PDPA); Private Healthcare Facilities and Services Act 1998 (Act 586); Medical Act 1971; Code of Professional Conduct from Malaysian Medical Council (MMC). In addition, there are existing policies and guidelines for reference issued by Ministry of Health.

The establishment of MyHDW provides for such avenue and requires proper governance in place to manage the ownership and accountability and access when such data or information are provided. Data and Information Governance Committee was established in the project headed by Deputy Director General of Health (Research and Technical Support) and reports to Health Informatics Council chaired by the Director General. The Term of Reference includes establishing policy, guideline and framework for data access management, data ownership, data and information quality, privacy and information management. Immediate focus is on user access. Stakeholder meeting was undertaken to address the issues and present the output to the committee and later to be endorsed by PIK before implementation. The remaining areas are privacy, use of data, consent and ownership will be addressed accordingly.

RAKKSSA provides coherent and comprehensive approach to security management. A Security Committee was established for the project, chaired by the Technical Lead of RAKKSSA. This committee is independent of the Project Implementation Team.

Full security controls include an information security plan which have been implemented in MyHDW. The plan governs the integrity, privacy, security, and confidentiality of MyHDW's information, especially highly sensitive information, and the responsibilities of departments and individuals for such information. Inappropriate use exposes MyHDW to risks including malware attacks, compromise of network systems and services, and legal issues.

During the development of the project there are requirements for all individuals to undergo security clearance at both the Ministry of Health and Chief Government Security Office (CGSO).

Role-based access control to data and information in MyHDW is enforced. Another key security management includes the process of de-identification and re-identification of patient's record. In addition to complex technical controls, Non-disclosure Agreements are used in to increase accountability of data and information users where technical approaches are not feasible.

MALAYSIAN HEALTH DATA WAREHOUSE (MyHDW)

2015-2016 START UP: INITIATION

SYSTEM ROLLOUT

System Rollout

In relation to System Rollout, the Project Glossary has defined the following terms that are widely used in the project:

Formal release	Software released from development after successfully meeting the objective and criteria of UAT, FAT and Release Readiness Review (RRR) gate checkpoint. This release is ready for deployment for production.
Go Live	A milestone to indicate that the system has been successfully deployed to the intended environment and become operational to selected users.
Rollout	An activity to deploy the system to the intended environment over a period of time and become operational to all users.

Table 6: *System Rollout Project Glossary*

Within MyHDW there are components that were developed in silo during Phase 1 i.e. Mi-Harmony and Geographic Information System (GIS). For the purpose of this rollout, they will still rollout but separately, and will be integrated with MyHDW in Phase 2.

The first to rollout was Mi-Harmony on 4 August, followed by MyHDW on 5 September, GIS on 6 October, SMRP Phase 1 comprising of Inpatient and Daycare modules on 1 November and PRIS National Obstetric Registry (NOR) on 2 December 2016.

KEY CONSIDERATIONS FOR THE ROLLOUT PLAN

For planning of rollout, the following were the key considerations:

1. **The integration and dependencies** between MyHDW, SMRP, PRIS systems and other system.
2. The system deployment and migration dependencies on stakeholder's system environment
3. The readiness prior to rollout to stakeholders
    a. Data and Information Governance
    b. Access to and Support for the Government Network (1GovNet)
    c. Customer Support Readiness
4. **Communication Plan** – stakeholder's awareness of what's coming, through website; portal and social media
5. **Training Plan** – readying the users to use the system. Train-the-trainer approach for mass training in future

THE INTEGRATION AND DEPENDENCIES BETWEEN MyHDW, SMRP, PRIS AND OTHER SYSTEMS

MyHDW, SMRP, PRIS, MyHarmony and GIS Go-Live following individual Go-Live milestone dates and subsequently rollout. However only upon the systems being integrated that the data and information processes are seamless and be fully represented. The following describes the integration dependencies:

a. For project Phase 1, MyHarmony Go-Live in August 2016, integrated with MyHDW 1.0 at single sign-on. MyHarmony will be integrated with MyHDW in Phase 2.

b. MyHDW Go-Live in September 2016 prioritizing with historical data from SMRP 1.1:

    i. 1st historical data load until June 2016. Therefore, the dashboards, graphs and reports will show data as at June 2016.

    ii. 2nd batch of historical data load until end 2016. By March 2017, MyHDW will be able to present the latest year i.e. 2016 annual result on the dashboards, graphs and reports.

c. Upon SMRP 2.0 full rollout in January 2017, the data from the enhanced SMRP 2.0 will be seen at MyHDW.

d. Upon SMRP 2.0 rollout only then the notification to PRIS 1.0, for the Non Obstetric Registry (NOR) will be populated, followed by Cancer Registry. Subsequently for Phase 2, data from PRIS will be populated to MyHDW.

SYSTEM DEPLOYMENT AND DATA MIGRATION DEPENDENCIES AND CONSTRAINTS ON STAKEHOLDER'S SYSTEM ENVIRONMENT.

MyHDW being a newly developed and implemented solution has key dependency on the readiness of the source systems and ability to upload the required data. However, the following are the constraints and dependencies for the other systems rollout.

1. SMRP System Integration – readiness of the private hospitals, other non-KKM hospitals, and Army Hospitals, and integration with systems in National Registration Department and Department of Statistics including iKelahiran.

    b. For the SMRP rollout is for the data coming in from other than KKM hospitals systems. Some of the hospitals rely fully on their vendors to maintain their system and to provide the integration requirements.

    c. Longer time required for engagement with other than KKM hospitals for them to agree and comply with KKM requirement for data to be transferred.

2. GIS in Phase 1 has integration through the single sign-on at MyHDW portal. The existing GIS will maintain only entry channel through MyHDW. There are administrative considerations to be addressed for the maintenance of existing production licenses and data migration. Thus the visualization layers and data integration to MyHDW will be performed in Phase 2.

3. MyHarmony in Phase 1 also has integration through the single sign-on at MyHDW portal. However, Mi-Harmony rollout has its own implementation track on Cardiology specialty, where three (3) facilities were proposed by the Head of Cardiologist for the rollout; Hospital Serdang, Institute Jantung Negara (IJN) and Hospital Sultanah Bahiyah. The implementation constraint is mainly in getting the acceptable format of data set from the hospitals for purpose of harmonization and codification.

THE READINESS PRIOR TO ROLLOUT TO STAKEHOLDERS

Data and Information Governance

Related to the use, access, quality, privacy and security of data and information. The establishment of the governance will give level of assurance to stakeholders to use the systems.

Data and Information Governance Committee was established, with first meeting convened on 27 June 2016, with the information gathering done with relevant stakeholders in March 2016 and further updated in workshop in August 2016.

The followings are in place for the rollout:

1. Policy and Guideline for Data Access Management
    a. Data Access Policy – General Statement
    b. Data Access Matrix
    c. User Registration and Account Activation Process
2. Policy and Guideline for Master Data Management

With the Data Access Matrix available, subsequently the establishment of the application administrators for SMRP and PRIS are assigned for the system to start operational.

Refer Table of Content on Data and Information Governance page 88.

Access to and Support for the Government Network (1GovNet)

The system rollout for the KKM Hospitals runs within the Government Network and for the Private and Non-KKM Hospitals their access to the system will eventually go through 1GovNet.

For KKM's systems operations running through 1GovNet, the access control is being managed by KKM's IT Department, Bahagian Pengurusan Maklumat (BPM). The communication and managing of the network access from KKM's offices and hospitals are controlled at KKM level.

Customer Support Readiness – readiness to support systems upon Go-Live and rollout

MIMOS Customer Helpdesk Centre is already setup to cater for Customer Support for MyHDW, SMRP and PRIS upon Go-Live and rollout. The following are in place for the Customer Support:

1. 1 300-88-8030 Hotline number dedicated for Helpdesk support
2. The use of Mi-Latte; MIMOS Helpdesk Ticket System for online ticket escalation
3. 24 x 7 Helpdesk Support
4. 3 Support Levels
    a. 1st level – Helpdesk
    b. 2nd level – from MIMOS (7 persons) and PIK (2 persons)
    c. 3rd level – MIMOS and PIK.

The 2nd level support from PIK will be channelled for issues/tickets on data, KKM policy and processes in using the systems in the daily operations.

Training Plan

The training is conducted as end-to-end training from SMRP to MyHDW as the target user groups are the same; mainly the Medical Record Officers (MROs) and the Assistant MROs. The MROs and AMROs are expected to enter into SMRP the patients' records and subsequently generate the Fixed Format Reports through MyHDW. The MROs are also expected to assist the Directors at Hospitals, District and State levels in churning out the reports and statistics which can be produced from MyHDW.

Train-the-Trainer

In addition to the MROs, there will be cases whereby the other administrative personnel and nurses assigned to enter patients' records and generate reports. To cater for the mass group of users later, the training plan follows Train-the-Trainer approach.

Thirty-two (32) key champions were appointed from the MRO and AMRO group to attend 3 full days end-to-end training from 10 to 12 October 2016. Refer table below session 1.

User Training

After the Train-the-Trainer session, 13 training sessions was held to cover end-to-end training. The trainers from the Key Champions above also facilitated these training sessions. In total 401 participants already trained for KKM Hospitals and 300 participants trained for Private Hospitals. The following is the Training Schedule.

No	Dates	Region	Attendees	Location
1.	10 – 12 Oct	User Champions from MRO Group	32 Key Champions	Pearl Int'l Hotel
2.	9 – 10 Nov	Zone Selatan (Johor, Melaka and N. Sembilan)	62	MIMOS
3.	13 – 15 Nov	Zone Sarawak	60	Kuching
4.	16 – 17 Nov	Zone Timur (Kelantan, T'ganu, Pahang)	84	Hospital Tumpat
5.	22 – 24 Nov	Zone Sabah	48	Grand Borneo Hotel
6.	28 – 29 Nov	Zone Utara (Penang, Perak, Perlis, Kedah)	51	MIMOS
7.	29 – 30 Nov	Zone Labuan	29	JKN Labuan
8.	1 – 2 Dec	Zone Tengah (Selangor)	67	MIMOS
9.	18 – 19 Oct	Private Hospitals – session 1	45	MIMOS
10.	24 – 25 Oct	Private Hospitals – session 2	52	MIMOS
11.	1 – 2 Nov	Private Hospitals – session 3	51	MIMOS
12.	7 – 8 Nov	Private Hospitals – session 4	64	MIMOS
13	14 – 15 Nov	Private Hospitals – session 5	56	MIMOS
14	13 – 14 Dec	Private Hospitals – session 6	31	MIMOS

Table 7: *User Training Details*

Trainers and Facilitators

For the training sessions, at project level, MIMOS and PIK have identified potential trainers and facilitators. This group has worked together to prepare the Training Modules for each system and gone through rehearsal prior conducting the first few training sessions. The following is the list of trainers and facilitators:

No	Areas	Trainer and Facilitator
1.	Overall and Lead	Dr Md Khadzir Sheikh Hj Ahmad Pn Kasinah Bidin En Wan Zawawi Md Zin Pn Norikma Ilias
2	Coordinator	Pn Zarith Sofia Zainal Abidin Pn Nurul Shahida Isnani Cik Sarah Masmerah Hanafiah
3	User Registration	Pn Nur Shuhada Husin En Sulaiman Ibrahim En Wan Mohd Nasrun Wan Sulaiman
4.	User Authentication through Mi-UAP	Cik Aini Mohamad Daud Cik Nurfarahani Jailani Cik Sarah Masmerah Hanafiah En Badrul Hisham Harun

No	Areas	Trainer and Facilitator
5.	MyHDW	Dr Teo Wil Ly Dr 'Ismat Sulaiman Dr Mohd Syazrin Mohd Sakri Cik Azlina Mohadzir Cik Aini Mohamad Daud Pn Fauziah Hanim Jahidin Pn Rahayu Abdullah Sani Cik Nurfarahani Jailani Cik Nur Syafiqah Munir
6.	SMRP	Pn Syamimi Mokhtar En David Liew En Badrul Hisham Harun Cik Nurul Syafira Fadzil Pn Hanani Azmi Cik Sarah Masmerah Hanafiah
7.	SMRP System Integration	En Damanhuri Ahmad Dusuki Cik Nor Shuhada Sulaiman
8	Helpdesk	En Ahmad Zuhairi En Jegathisan Gonasegaran En Calvin Magen R Magendramani Cik Nursyuadah Abdullah

Table 8: *Trainers and Facilitators List*

The training modules cover the following:

1. User Registration
2. User Authentication through Mi-UAP
3. MyHDW
4. SMRP
5. SMRP Integration.
6. Helpdesk
7. The participants are provided with the link to access the Training Modules above.

Echo Training

After Train-the-Trainer, the 32 key champions from respective State and hospitals continued with Echo Training where they then practice using the systems and conduct further training to the other target users at their own premise. The echo-training has commenced in December 2016 and expected to be ongoing until March 2017.

MIMOS has already setup Training Environment within KKM's production environment, and after strict Security Posture Assessment (SPA) has opened access for purpose of the training.

Communication Plan

The communication for the awareness and buy-in for the project to the various levels of stakeholders are on-going done by PIK and through various channels. The following are the identified channels:

1. Information and Governance Committee Meetings and workshops
2. Engagement Meetings with Private Hospitals and Non-KKM Hospitals for system integration with SMRP
3. Engagement Meetings prior rollout for Mi-Harmony
4. Engagement Meetings with Department Heads, Registry Owners and key representatives for requirement gathering sessions
5. Throughout 13 Training sessions for SMRP and MyHDW

All the above engagement meetings are accompanied by official correspondences to notify the coming rollout milestones and activities.

Operational Readiness

The focus and highlight for system operations are the use of SMRP and MyHDW where the full rollout is target for 1 January 2017. PIK has given the direction that SMRP data entry for every census year is closed by the end of Feb the following year. As KKM users start using SMRP 2.0 on 1 January 2017, it is anticipated that the heavy usage SMRP 1.1 to close 2016 census year.

The following summarises the readiness for SMRP 2.0 rollout and PIK Operations:

TEAM	ROLE
MIMOS Berhad	1. Solution readiness
	2. Infrastructure readiness
	3. Helpdesk readiness (to receive and follow-through tickets status)
	4. 2nd and 3rd level support group
	5. Frequently Asked Question (FAQs)
PIK	1. Cut-off date for user registration
	2. Portal User registration & user roles assigned – MyHDW Portal Administrator and SMRP Application Administrator
	3. User maintenance process (changes and updates of users)
	4. 2nd PIK Support Level
	5. MyHDW Portal content maintenance

TEAM	ROLE
1. JKN 2. KKM 3. Non-KKM 4. Private Hospitals Note: • BPM and/or • appointed IT Personnel • MRO	1. Check https://myhdw.moh.gov.my is accessible from facility 2. Check https://myhdw-test.moh.gov.my is accessible from facility 3. Monitor and manage network bandwidth for JKN and hospitals 4. Infrastructure – laptop/PC minimum specs 5. Key Champions – as liaison or reference for issues raised 6. Users registered and trained

Table 9: *SMRP 2.0 Roll-out and PIK Operations Readiness*

In addition to above, MIMOS has facilitated the establishment of the following processes:

1. Operating Process for Changes to Production System and Notification of Planned Downtime.
2. Operating Process for Issues Escalation and Notification of Unplanned Downtime
3. Backup and handover procedure

MALAYSIAN HEALTH DATA WAREHOUSE (MyHDW)

2015-2016 START UP: INITIATION

MOVING-FORWARD: MyHDW PHASE 2 AND BEYOND

Moving-forward: MyHDW Phase 2 and Beyond

PHASE 2 MyHDW

It is proposed that future phases of MyHDW will expand data sources and also reporting and analytic capability of the environment. The latter will see the inclusion of Big Data Analytics integration. This will allow a greater number of healthcare related questions to be answered, improve timeliness of information and utilisation of semi and unstructured data where the business use case is applicable.

It is proposed that MyHDW Phase 2 begin immediately after the current phase ends and run for a period of two years from 2016 to 2018. Phase 2 will include the following:

- Additional data sources including; Outpatient, Clinical Support Services, Family Health and Oral Health.
- Standard reporting such as Fix Format Reports, Dashboards and also Ad-hoc Query utilising data from; PRIS Modules 3/4, Outpatient, Clinical Support Services, Procedures, Traditional and Complementary Medicine, Family Health and Oral Health.
- Full integration of GIS capability into the MyHDW architecture. In addition to broader security capability this will include functionality associated with facility planning to include, contour formation, drainage, site location and traffic. This latter development of GIS layers will be guided by various KKM divisions and contribute to improved site and disaster management planning.
- Expansion of Mi-Harmony capabilities. This will include integration into the core MyHDW architecture also, ongoing development of additional SNOMED CT RefSet data to incorporate additional healthcare domains.
- The expansion of analytic tools particularly associated with statistics and possibly data exploration.

Please see figure below (replicated from architecture section) which outlines the above items.

BIG DATA ANALYTICS

Big Data or Big Data Analytics (BDA) is a relatively new approach to how analytic data is considered and extends or augments traditional data warehousing approaches. It is commonly defined as what is known as the three or four "Vs". These are associated with increased data; Volume, Variety and Velocity. Sometimes a fourth or even fifth "V" is added to denote Value or Voracity. Notwithstanding definitions, BDA represents a paradigms shift both in terms of technology and how analytics can be performed and how this might be consumed in an increasingly prevalent way to inform decision-making within organisations. This is an ever evolving and increasingly complex area perhaps even more so in the healthcare domain. It would be fair to say that care should be taken to understand the maturity and likely productivity of technologies, approaches

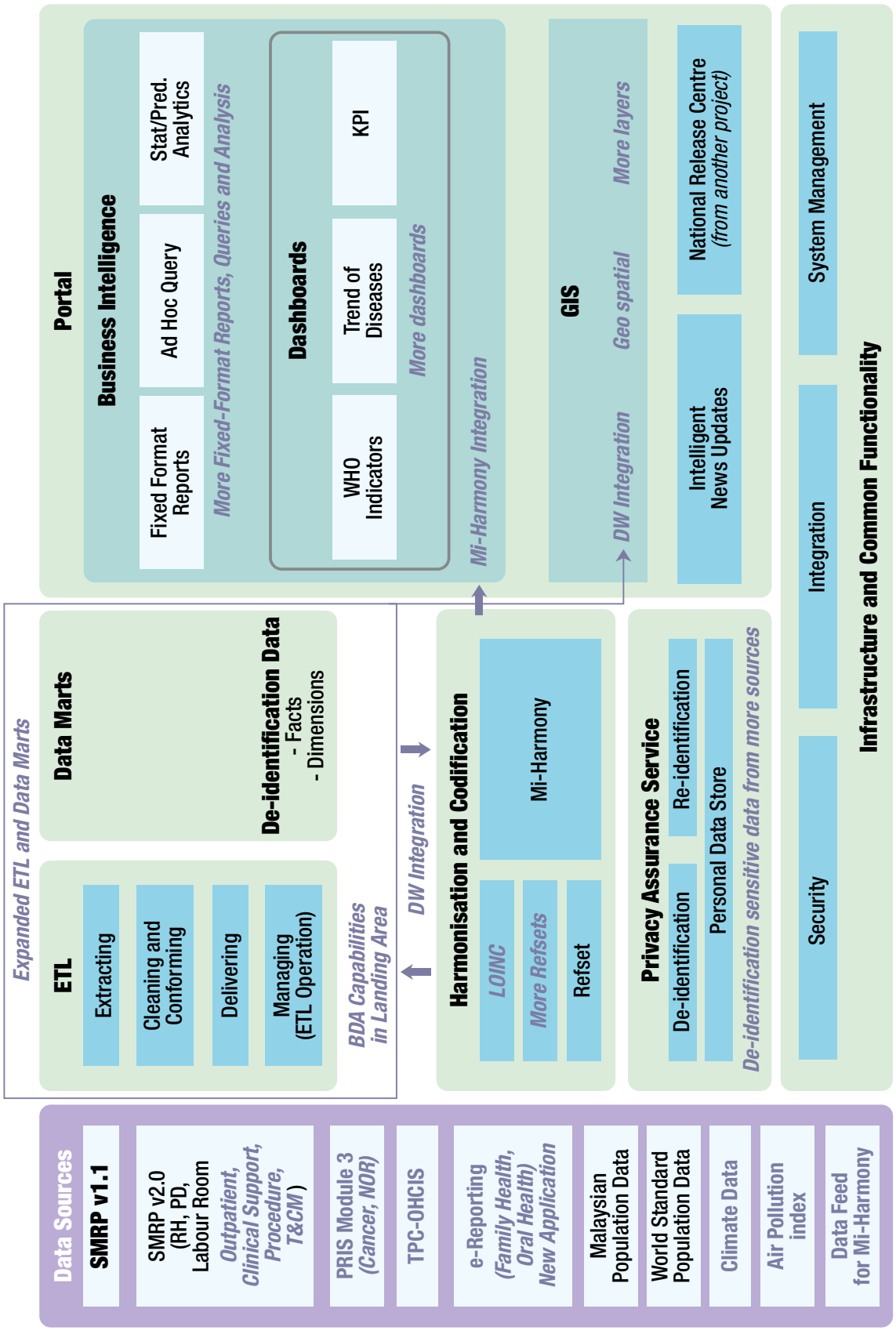


Diagram 16: *MyHDW Phase 2 Architecture*

and the availability of appropriately skilled resources. Furthermore, which of these is best suited to the business use case under consideration.

MyHDW 1.0 primarily focused on establishing the fundamental underpinnings for a national health analytic system, both technically and in terms of supporting program services. The data architecture of this phase, as has been previously mentioned focused on an implementation of a traditional structured data warehouse. It also recognised that future phases would likely include semi or unstructured data. Furthermore, this data might be stored in something other than an RDBMS such as PostgreSQL in either Hadoop/HDFS or another NoSQL data management system such as a Document or Graph Database. MyHDW 1.0 prepared for this through the use of a Hadoop/HDFS Landing Zone. This is broadly equivalent to what is sometimes known as a Data Lake, in which large amounts of raw data are stored in a native format normally in Hadoop. This allows data to be utilised in downstream traditional data warehouses or directly into predictive or data discovery analytic tools. In MyHDW all data is stored in this Landing Zone in an encrypted state and further protected by aaccess control and Privacy Assurance Service (PAS).

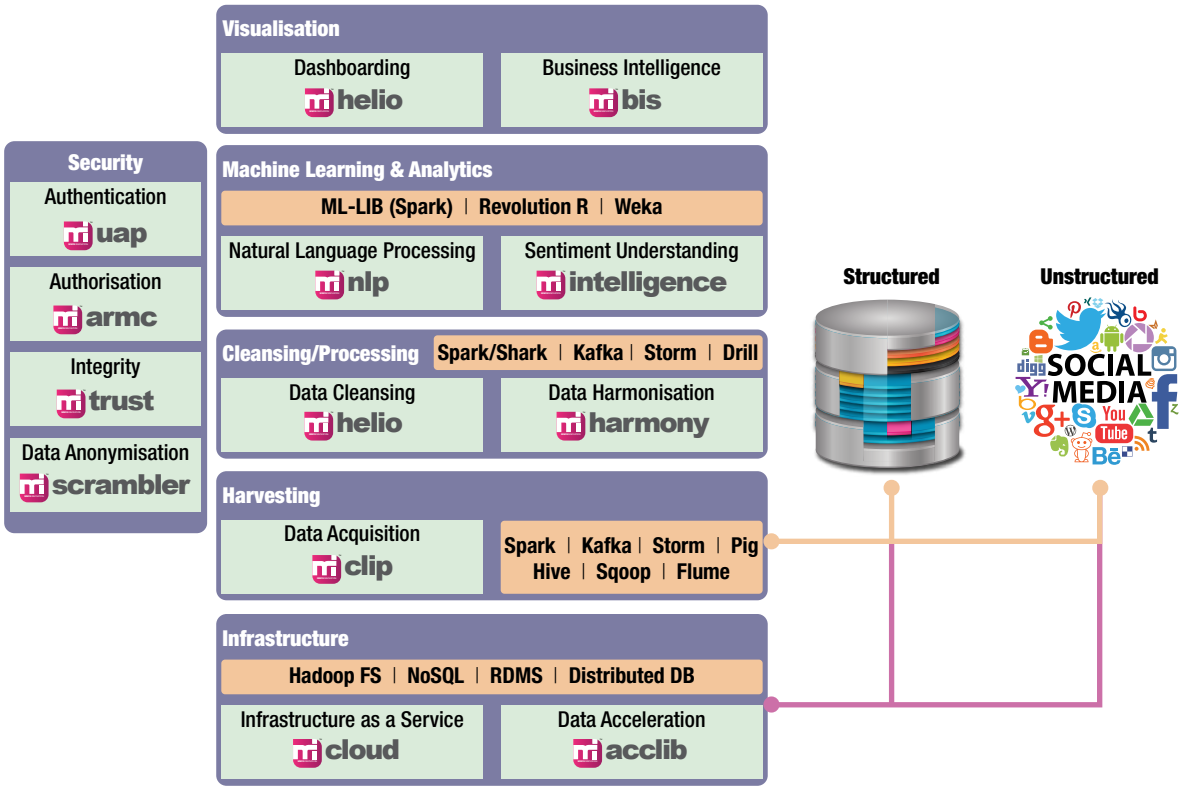


Diagram 17: *Big Data Software Stack*

It should be recognised that while BDA offers new and exciting types of analytic capabilities, that a traditional structured data warehouse is also needed for certain types of high precision analytics. To a degree a structured data warehouse augmented by BDA techniques achieves the best of both worlds and services the broadest range of analytic use cases. For example, structured data in MyHDW is particularly suited for Health Indicator and Key Performance Indicators utilised in Healthcare Planning and Management. In fact, any type of traditional reporting, dashboards and Ad-hoc query are likely best done utilising a backend of structured data. By contrast more exploratory data investigation utilising analytic or statistical tools in which the questions to

be answered are not previously known may best be done against data stored in HDFS or a Document/Graph Database. In this approach, data is not pre-structured through a data model until it is actually used, this is sometimes known as "schema on read" or "deferred binding".

Another area in which BDA offers significant advantages is in the reduction of *data latency*. In this data may be available far more quickly than in traditional models including real or near time data. This also has the added benefit of significantly reducing costs associated with ETL. That said is important that the technology, approach and architectural facet is suited to the business use case. Recently these combinations of traditional data warehousing involving structured data and BDA approaches are known as Logical Data Warehouse or Data Management Systems Analytics.

MIMOS Berhad is the last two years been refining and investigating approaches to BDA and developed a number of platform technologies and an overarching *Big Data Software Stack* and architecture consistent with these. This is shown in the figure below.

EMERGING TRENDS IN HEALTHCARE ANALYTICS

Increasing complexity in healthcare systems and data is driving change in terms of systems that are best suited to meet the challenges of providing high quality analytics to support these. One challenge is how to keep up with this constantly evolving landscape. It is certainly recognised that we live in a world where both structured and unstructured data need to be aligned and that there are ever-increasing pressures to improve access to data and information. In addition, ensuring an appropriate level of data quality is also an ongoing consideration. By contrast privacy and security concerns need careful and significant attention to ensure these are in line with legislation and continuing public trust.

New data sources are also prevalent particularly those associated with the Internet of Things of which about 30% is composed of healthcare data including machine-readable data from sensors and medical devices etc.

In terms of Health Analytics aligned with EMR/EHR programs, internationally there have been some good progress particularly under "meaningful use" and "accountable care" initiatives in the US. Along with this increased interest in Health Analytics and Health Data Science. Similarly, there are increased investments into predictive analytics as opposed to analytics on historical data. Other specific analytic techniques that continue to be utilised in the more prevalent way include Text Analytics, Data Mining and Optimisation. In terms of the healthcare verticals which are emerging Risk Management and Population Management would appear to be of increasing interest. As are Client Experience and Pathways of Care.

SNOMED CT is also recently being considered for direct analytic purposes though it is still early to determine how effective this will be. Similarly, SNOMED CT RefSets do provide a means to drive precise meaning in specialty areas such as Primary Care and certain other disciplines like Cardiology.

There are new interoperability standards such as HL7's FHIR[29] which is a more granular way to exchange data without the rigidity of traditional HL7. The latter is becoming increasingly paired with other standards such as SMART[30] which allow a variety of mobile devices to interact in a practical manner with EHR's.

In terms of self-service models Open Data is becoming an increasingly popular and effective means increase access to trusted data and information.

Future phases of MyHDW was certainly recognised the augmentation of traditional data warehousing approaches to add BDA technology and approaches thus creating a logical data warehouse architecture. In addition, the potential to capture data from clinical sources is certainly considered as is a wide variety of tools to query and perform analytics on data in MyHDW's analytical data management system.

29 Fast Healthcare Interoperability Resources (FHIR). See : https://www.hl7.org/fhir/

30 Substitutable Medical Apps, Reusable Technology (SMART). See : http://smarthealthit.org/

MALAYSIAN HEALTH DATA WAREHOUSE (MyHDW)

2015-2016 START UP: INITIATION

SUSTAINABILITY

Sustainability

A successful systems initiative ultimately involves not just the initial development of the product but also how this product is run to ensure good user satisfaction and providing actual and perceived high value. This requires a long-term or program view to be in place to ensure these factors are achieved and maintained. Practically this sets the time horizon of three years plus as a minimum for considering systems of this nature. Particularly in terms of funding, budgeting, resource allocation and planning.

Associated with MyHDW 1.0 there are a number of sustainment activities which to some degree are outside the main development tasks. These include:

Sustainment Activity	Current Owner
Data Quality	PIK
Support and Training especially eLearning	PIK for Health Informatics Standard. MIMOS Berhad for support and training associated with the system
Architecture: Application, Data & Technology	MIMOS Berhad
Standards; Integration with other Health Informatics Standard	PIK
Release and Deployment Planning	PIK/MIMOS Berhad
Operations and Maintenance – IT Service Management (ITSM[31]/ITIL)	PIK
Future Planning and Phases	PIK/MIMOS Berhad
Capacity Management	MIMOS Berhad
Business Intelligence Competency	PIK/MIMOS Berhad
Master Data Management Program	PIK/MIMOS Berhad
Portfolio/Program Management[32]	PIK/MIMOS Berhad
Product Management[33]	PIK/MIMOS Berhad
Human Resource Management	KKM/MIMOS Berhad

Table 10: *Sustainment Activities*

Increasing maturity and experience with Healthcare Data Warehousing shows a common trend of moving from the project view to a longer term Program or Portfolio one, once initial deployments are rolled-out and stabilised. For example, the coordination of Architecture and Standards, Master Data Management (MDM) etc. Overall a MyHDW Solution or Program might include the following components:

31 IT service management (ITSM)is an integrated, process based, set of best practices to manage IT services to meet unique customer requirements and priorities.See : http://www.itsm.info/ITSM.htm

32 Program Management: "A group of related projects managed in a coordinated way to obtain benefits and control not available from managing them individually. Programs may include elements of related work outside scope of the discrete projects in the program." – PMBOK

33 Product Management – Part of above with a holistic focus on the relative cost and value of products.

- Initiation/Blueprint/Business Case/Funding
- Capacity Building
- Architecture and Standards
- Software Development
- Data Curation: Data Management and Governance
- Systems Integration
- Deployment, Training, Support
- Operations/Hosting
- Enhancement and Maintenance
- Ongoing Intake Management
- Product Management including Value (Value Stream), Customer Satisfaction, Product Roadmaps, Product Cost etc.

Some of the above sustainment activities were discussed during the project implementation during an information sharing workshop on this topic.

From the perspective of an Operating Model it is also important to consider the role of each partner organisation to include considerations associated with Product Development, Systems Integration (SI) and Hosting/Application Service Provider (ASP) services.

In addition to process related items, Human Resource (HR) management is a critical item. It is necessary we have the right type and quantity of resources both in terms of the technical team and also the business and analytical teams. Experience has shown that a good deal of additional resource is required on an ongoing basis both to build, operate and use a Health Analytic environment such as MyHDW. This inevitably is at some cost though in a broader context the value obtained from the system should outweigh this. In MyHDW 1.0 a good deal of HR activity was associated with recruiting and/or allocating project resources associated with both the technical and business teams. A number of these roles are new and may be associated with privacy, security, data quality or data analysis. Longer term strategies associated with talent management, retention and employee satisfaction etc., will need to be established and key performance indicators associated with these tracked. One critical item that will need close monitoring is staff retention as we transition between project phases as will occur during August 2015.

Short to mid-term strategies associated with the transfer of technology (ToT) duties to internal KKM ICT staff should also be monitored to ensure feasibility. One strategy associated with this has been to ask technical personnel to self-evaluate their technical skills in areas such as RDBMS, ETL and BI.

Sustainability and Program activities are highly detailed topics and are outside the scope of this report at this stage other than the summary outlined above. These will require further examination in a future report or project stage.

MALAYSIAN HEALTH DATA WAREHOUSE (MyHDW)
2015-2016 START UP: INITIATION

CONCLUSIONS

Conclusions

This initiative set out to build the foundation of a National Health Analytics environment to provide timely self-service access to information to manage and optimise the Health System, implement improved Surveillance and facilitate Clinical Research, in addition to other types of health information requirements. This is the first time that a system such as this has been developed in Malaysia. It was recognised early on that one critical factor for a successful implementation would be an understanding of the starting point of local capacity capable of delivering such a complex system, and this was investigated as part of the initial Blueprint work. In addition, a number of deliberate strategies were adopted to reduce risks and improve the conditions for effective and high quality deliverables. Certainly a willingness to learn from others, both locally and internationally has been used to good effect. In addition, collaborative and innovative partnerships across government agencies have also been instrumental in furthering this work. In summary the following strategies were used to facilitate the planning, inception and implementation of MyHDW:

- Careful planning and project inception involving comprehensive stakeholder engagements and consultations. This was documented during the period 2011 to 2013 in the MyHDW Blueprint material and this and the associated portfolio plan, was used in an updated form to guide the implementation work associated with this report. Intrinsic in this is also long-term continuity and commitment from all key stakeholders.

- An ICT capacity development strategy involving partnership between the Ministry of Health (KKM) and MIMOS Berhad. This recognised the need for an 'incubator' approach for the systems development work associated with the project as well as the need for skilled and committed resources. Furthermore, the requirement for high security and comprehensive data centre services was also enabled by this arrangement. In addition, the use of local technology and infrastructure, a direction that have been previously set was synergistic and allowed access and further development to the MIMOS Berhad internally developed platform technologies which were employed on this initiative.

- The utilisation of a business domain led initiative. Historically these types of initiatives have often been led from the technology function within an organisation. It is generally accepted at this point that this approach is not optimal and that having a senior domain expert or subject matter expert (SME) to lead initiatives such as this is the best way to ensure a system that is clearly aligned to business drivers and recognises value and return on investment.

- Leveraging existing best practice in the planning and design of large-scale Health Analytic systems and implementations associated with Business Intelligence, Data Warehousing and Big Data Analytics. It has been recognised from the earliest planning stages of this project that learning from others is an effective strategy to reduce risk. Engagement of an experienced international consultant and in addition to ongoing information exchange through international and national analytics and Informatics groups and forums underpin the basis of this approach.

- The use of techniques which allowed the accelerated development of Health Informatics Standard to support the ITC portion of the project in addition to improved semantic interoperability across healthcare domains. Specifically, this relates to the development

of the Malaysian Health Reference Data Model (MyHRDM). An initial draft version of this high-level reference model was developed during the latter part of 2015 and was used as the underpinning of all structured data modelling in the SMRP and PRIS collection systems as well as the BI/DW environment. This model was able to build upon a previous version created in 2013, international examples and a rapid development process to incorporate aspects of SMRP, PRIS (Cancer) as well as parts of TPC-OHCIS. In addition, cardiology RefSets derived from SNOMED CT were developed to facilitate utilisation of the MIMOS Berhad platform tool Mi-Harmony which will be deployed in a subset of three hospital sites during Phase 1. This will allow analytics and auto coding/terminology, associated with semi structured medical data in those sites.

While the above strategies were in most instances successful, inevitably a number of challenges also arose during the implementation, these are noted below as lessons learned and include:

- It is important that that data sources utilised within the analytic environment, especially those associated with structured data are stable and mature prior to loading. In this project phase, SMRP 2.0 was developed at the same time as extract transformation and load processes were being developed for MyHDW and this caused a number of complexities throughout the project stages.
- Ongoing training and coaching may be needed to increase skill levels associated with BI/DW/BDA large-scale development. This is particularly critical in the areas of project management and requirements gathering.
- That some degree of buffer be incorporated into the project timelines to accommodate unexpected requirement changes and situations.
- The use of local technology for a number of the critical components of the system had many significant benefits and is in line with directions associated with the project. It should also be recognised that in many instances these were not of the same maturity as best of breed commercial products. For future phases a number of strategies might be considered to ameliorate this situation including; increased investment into local platform technology to close any gaps that exist, setting expectations with the user community and where necessary supplementation with commercial products.
- To both continue to develop future phases of the system while sustaining MyHDW 1.0 will require careful monitoring to ensure sufficient resources are available to ensure good user satisfaction and also ongoing high quality development efforts.

Overall, MyHDW 1.0 (Phase 1) delivered on time, scope and budget. The partnership between the Ministry of Health and MIMOS Berhad in terms of technology service provisioning, has worked well and in some instances deliverables have even exceeded original scope specifications. Activities following the initial deployment will involve operations, maintenance and enhancements. In addition, MyHDW will continue to evolve to include additional data sources, a broader user base and also BDA capability. It will be important during this time to ensure that sustainment activities such as ongoing funding, program management and capacity management etc., are carefully considered. Furthermore, security and privacy processes should continue to mature and be monitored to ensure the highest confidence in MyHDW by users and the public.

With the foundation of MyHDW established it will also present an opportunity to further enhance information products such as measurement of Health System performance, Health Indicator development, Clinical Benchmarks, Open Data initiatives and so forth. These latter activities will in particular ensure that MyHDW is of ongoing high value and good return on investment.

MALAYSIAN HEALTH DATA WAREHOUSE (MyHDW)
2015-2016 START UP: INITIATION

APPENDICES

Appendix 1 -Blueprint Portfolio Plan

REF	INITIATIVE / TASK	NOTES	2011	2012	2013	2014	2015
11-1	Create Business Case and present to ICT SC		●				
11-2	Establish National Health Informatics Council		●				
11-3a	Develop Start-up Plan			●			
11-3b	Participate in National eHealth initiative				●		
11-3c	Develop 3 MyHDW Portfolio Plan	MyHDW Portfolio Plan				●	
11-3d	Develop MyHDW Requirement and Architecture Report(s)	Establish and document a global requirement for MyHDW. Develop and recommend an architecture to support requirements.				●	
11-4	Develop Infrastructure Plan	Technology, Tools, Infrastructure Plan		●			
11-5	Gap Analysis between HIS and SMRP/Discharge summary data sets.	Gap analysis to determine if HIS can be used for SMRP/ Discharge collections	●				
11-6	Recruit and train core teams - Analysts and IT resources	15 x IT 15 x Analytical/Stats/ Research See Appendix 3 for IT resources		●	●		
11-7	Reference Data Model and Data Dictionary	Establish national data standards (structure) – Data Model and Data Dictionary		●	●	●	●
11-8	Develop Health Information Model/ Framework	Overarching map of key questions that need to be answered to support priority KPI's, 1Care and HIMS and delineate the products and data required to answer them.		●	●		
11-9	Communication Plan	Communication material, roadshow etc. to communicate the MyHDW concept to key stakeholders		●	●		
12-1	Procure and setup technical infrastructure			●	●		

REF	INITIATIVE / TASK	NOTES	2011	2012	2013	2014	2015
12-2	Establish enhanced analytical and technical capacity/ capability in HIC and IT			●	●		
12-3	Build 1 x HIMS databases 'Data Marts' within MyHDW	Establish 1 x Data Marts and Reporting Products – Proposed candidate: SMRP		●	●		
12-4	Renal & Cancer Registries	If possible house 2 x registries ideally Renal and Cancer within the MyHDW infrastructure. These would be separate from MyHDW but could leverage the same technology and services		●	●		
12-5	Establish Master Data Program – Develop Facility and Professionals Master data	Establish program/function in HIC to develop and maintain master or reference data. Initial master data candidates are Facility and Professionals		●	●	●	●
12-6	Strategy, planning and policy development, System performance reporting	Report and Information products development for these areas. TBC		●	●		
12-7	Systems Running - Establish Operations and Support Services	Establish group to maintain and support what has been built			●	●	●
13-1	Add 1 Data Mart to MyHDW from a mature data source- Explore 2 additional Data Marts if resources permit	Establish a Data Mart and Reporting Products from a mature and strategic data source– Likely candidates would be: SMRP, Communicable Diseases, Primary Care, Pharmacy -To be determined based on further analysis. Explore 2 additional Data Marts if resources permit			●		
13-3	Add Public health surveillance data/ functionality	Addition of Public Health Surveillance data/function TBC			●		
13-4	Patient safety initiatives, Monitoring access to care	Develop new reports for Patient safety initiatives, Monitoring access to care			●		
14-1	Add 2-3 Data Marts to MyHDW	New Data Marts To be determine after further analysis				●	●

REF	INITIATIVE / TASK	NOTES	2011	2012	2013	2014	2015
14-2	Other Registries	To be determined after further analysis				●	●
14-3	LHR	Begin to utilise LHR data feeds as possible. Further analysis needed				●	●
14-4	Clinical research studies, Disease and wellness program	Develop new reporting products for these areas				●	●

Color Legend:

Completed Task
In Progress
Future Plan

Acknowledgements Contributors

MIMOS Berhad

Wan Ahmad Jafri Wan Abdul Aziz

Wan Zawawi Md Zin

Norikma Ilias

Dr Teo Wil Ly

Dr Khalil Bouzekri

Fazli Mat Nor

PIK

Dr Ismat Mohd Sulaiman

MyHDW Project

Malaysian Health Data Warehouse team members

Special mention

Azlina Mohadzir and Mohd Faizal Amin for coordinating this publication.

MyHDW Resources in Phase 1

No	Roles	Name	Project Role
1	Project Management	Wan Zawawi Md Zin	Project Director
		Norikma Ilias	Project Manager
		Nur Shuhada Husin	Super User
	BUSINESS		
2	Sr. Business Analyst	Azlina Mohadzir	Business Analyst cum Technical Writer
3	Business Analyst	Aini Mohd Daud	Business Analyst
		Nur Syafiqah Munir	Business Analyst/ Apprentice
4	Medical Record Officers	Norsuhada Sulaiman	SME Lead - System Integration
		Nurulhuda Ramlan	SME Lead - SMRP
		Sulaiman Ibrahim	SME - PMO
		Ayu Fissabieummisukasih Parman	SME - SMRP
		Norati Abdullah	SME - SMRP
5	Statisticians	Zainuddin Ahmad	SME - Statistician
		Wan Roslida Othman	SME - Statistician & Tester
		Sabrina Sabri	Lead SME - Statistician & Test Lead
		Suhaya Komari	SME - Statistician & Tester
		Mohd Rosli Mahmood	SME - Statistician & Tester
		Mohd Mahadzir Tumin	SME - Statistician & Tester
		Azlinda Azrai	SME - Statistician & Tester
6	Change Management Team/ Trainers	Kasinah Bidin	Lead SME - Training
		Dr. Md. Khadzir Sheikh Ahmad	Project Chairman
		Dr. Ismat Mohd Sulaiman	Lead SME - MyHarmony
		Nor Hanani Azmi	SME - Training
		Dr. Teo Wil Ly	Project/Technical Lead (MyHDW)
		Azlina Mohadzir	Business Analyst cum Technical Writer
		Rahayu Abdullah Sani	Developer Front End (BI/ Report Development)
		Fauziah Hanim Jahidin	Developer Front End (BI/ Report Development)
		Nurfarahani Jailani	Developer Front End (BI/ Report Development)
		Syamimi Mokhtar	Business Analyst
		Sarah Masmerah Hanafiah	Apprentice
		Badrul Hisyam Harun	Apprentice

No	Roles	Name	Project Role
7	Audit (Internal)	Yong Seak Sun	Quality Lead
8	Privacy Officer	Dr. Md. Khadzir Sheikh Ahmad	Project Chairman
9	Sr. Domain Specialist	Dr. Md. Khadzir Sheikh Ahmad	Project Chairman
		Dr. Ismat Mohd Sulaiman	Lead SME - MyHarmony
		Nur Shuhada Husin	Lead SME - IT and PMO
		Wan Mohd Nasrun Wan Sulaiman	Lead SME - GIS
		Dr. Nur Shaema Darus	Lead SME - ICD-10
		Dr. Syirahaniza Mohd Salleh	SME - MyHarmony
10	Transfer Of Technology	Adam Malik Sahar	ToT Lead and SME IT - ETL
		Hamidah Mat	SME IT - SMRP
		Norzaidah Mohamad Nor	SME IT - PRIS (TOT)
		Mohd Firdaus Mohamed Khairi	SME IT - BI (TOT)
		Noor Shabirah Shamsuri	SME IT - DBA (TOT)
		Faridah Hanim Haron	SME IT - Testing (TOT)
11	Health Informatician	Dr. Ismat Mohd Sulaiman	Lead SME - SnomedCT
12	Health Informatics Standard Specialist	Dr. Md. Khadzir Sheikh Ahmad	Project Chairman
13	Data Analyst	Dr. Md. Khadzir Sheikh Ahmad	Project Chairman
		Wan Roslida Othman	SME - Statistician & Tester
		Sabrina Sabri	Lead SME - Statistician & Test Lead
		Suhaya Komari	SME - Statistician & Tester
		Mohd Rosli Mahmood	SME - Statistician & Tester
14	Data Scientist		
15	Health Informatician	Dr. Ismat Mohd Sulaiman	Lead SME - SnomedCT
16	Health Informatics Standard Specialist	Dr. Md. Khadzir Sheikh Ahmad	Project Chairman
17	Data Analyst	Dr. Md. Khadzir Sheikh Ahmad	Project Chairman
		Wan Roslida Othman	SME - Statistician & Tester
		Sabrina Sabri	Lead SME - Statistician & Test Lead
		Suhaya Komari	SME - Statistician & Tester
		Mohd Rosli Mahmood	SME - Statistician & Tester
18	Data Scientist		
	TECHNICAL		
19	Technical Team Manager	Dr Teo Wil Ly	Project/Technical Lead (MyHDW)

No	Roles	Name	Project Role
20	Project Coordinator	Zarith Sofia Zainal Abidin	Senior Project Admin
		Nurul Shahida Isnani	Project Admin
		Sarah Masmerah Hanafiah	Apprentice
21	DBA	Nazeri Hambari	DBA (Infra, Integration and Migration)
22	Data Architect / Data Modeler	Dr. Seah Boon Keong	DW Data Modeller
		Vinay Kumar Ramidi	ETL Engineer (Data Modeling, Integration and ETL)
23	Data Architect Consultant	Mr Mark Fuller	Project Consultant
24	System Engineer/ System Admin/ Network Administrator	Mahathir Sulaiman	System/Network Lead
		Yau Tee Kuan	Network Engineer
		Anwaruddin Hamdan	Infrastructure Engineer (Technical Support)
		Virga Ramada Buana	Technical Backend (Infra, Integration and Migration)
		Nazarudin Mashudi	Configuration Engineer
25	ETL Engineer / ETL Developer	Vinay Kumar Ramidi	ETL Engineer (Data Modeling, Integration and ETL)
		Lim Hooi Hwa	"ETL Engineer (Data Modeling, Integration and ETL) *(joined on April 2016)*
		Pramod Reddy Menne	DW ETL Developer
		Mallesham Balagoni	DW ETL Developer
26	Solution Architect (Technical / Solution Lead)	Dr Teo Wil Ly	Project/Technical Lead (MyHDW)
27	Technical Data Warehouse Consultant	Mark John Fuller	Project Consultant
28	Solution Developer / Application/ Solutions Developer-web/Java	Javed Ahmed	Developer Front End (Big Data and Portal)
		Yap Chin Wei	Developer Front End (Big Data and Portal)
		Yusup Andri	Developer Front End (Mi-BIS)/ Apprentice

No	Roles	Name	Project Role
		Nur 'izzati Ahmad Zaki	Developer BackEnd (Mi-BIS)
		Norlisa Ahmad Termizi	Developer BackEnd (GIS)
		Luqman Razak	DW ETL Developer Application Development (PAS)
		Fazli Mat Nor	Mi-BIS
		Wael Hamdan	Developer BackEnd (Mi-Harmony)
		Lew Chia Chyi	Project/Technical Lead (SMRP PRIS)
		Damanhuri Ahmad Dusuki	Developer BackEnd (Integration)
29	Testing and QA / Software and Data Testing Lead	Rajashekara Swamy	Test Chairman
		Catherine Yumar	Test Lead
		Om Seshu Karthik	Tester
		Nor Adlin Mat Noh	Tester
		Yogini A/P Sunthera Sagaran	Tester
		Ahmad Zuhaili Ikhwan Ahmad Khairi	Tester
		Nor Sahreena Mohd Salehan	Tester
		Izzah Khairiyah Abidin	Tester
		Nor Izzati Osman	Tester
		Kalairasu Santhirasekaran	Tester (Performance)
		Sandhya Kiran	Tester (Performance)
		Sabrina Sabri	Lead SME - Statistician & Test Lead
30	BI Developer	Rahayu Abdullah Sani	Developer Front End (BI/ Report Development)
		Fauziah Hanim Jahidin	Developer Front End (BI/ Report Development)
		Nurfarahani Jailani	Developer Front End (BI/ Report Development)
31	Senior BI Developer Consultant	Rahayu Abdullah Sani	Developer Front End (BI/ Report Development)
32	Data Quality Engineer	Dr. Khalil Bouzekri	Project/Technical Lead (Mi-Harmony) (Unstructured)
		Lim Hooi Hwa	ETL Engineer (Data Modeling, Integration and ETL)
33	Information Security Consultant / Information Security Lead	Ng Kang Siong	Security Chairman

No	Roles	Name	Project Role
		Alwyn Goh	Security
		Wong Hon Loon	Technical Backend (Technical Support)
		Sea Chong Seak	Mi UAP Integration
		Tn Haji Jaafar Jamaan	ICTSO, KKM
34	Technical Writer	Azlina Mohadzir	Business Analyst cum Technical Writer
		Nur Syafiqah Munir	Technical Writer/ Apprentice
35	Web/Graphic Design	Hazmi Md Zaki	UI UX Design Lead
		Mohd Akhmal Abd Manaf	UI UX Designer
		Idyawati Hussein@Hussen	UI UX Designer
		Safarihairol Mustafa	UI UX Designer
36	Quality Assurance (QA) Lead	Yong Seak Sun	Quality Lead

** The above list refered to those who are involved in MyHDW Phase 1 (at least 6 month)*

ISBN 978-967-0769-76-9

MALAYSIAN HEALTH DATA WAREHOUSE (MyHDW) 2015-2016 START UP: INITIATION

HEALTH INFORMATICS CENTRE, PLANNING DIVISION
MINISTRY OF HEALTH MALAYSIA
Tel : +603 8883 3128/9 Fax : +603 8883 3169 Email: myhdw@moh.gov.my