Data Framework

An Outline of Data Practices and Procedures for the University of Waikato / July 2017

Author’s name:
Mr. Gary Buczma
UoW Contract Architect

Written on behalf of
Mr. Eion Hall,
Director
Information Technology Services
# Table of Contents

1. **Overview**  
   1.1 Context  
   1.2 Current State Analysis  
   1.3 Looking forward - Where do we want to be?  

2. **The Data Framework**  
   2.1 High-level UoW Data from 30,000ft  
   2.2 Data Goals  
   2.3 Data Architecture Principles  
   2.4 Strategic Initiatives - Medium term initiatives to realise data goals  
   2.5 Detailed Approach for Strategic Initiatives  
      2.5.1 Strategic Initiative 1: Consolidate technologies and complex techniques for Data Transportation/Integration  
      2.5.2 Strategic Initiative 2: Introduce Master Data Management (MDM)  
      2.5.3 Strategic Initiative 3: Develop a Data Services Capability  
      Data Services team - Developing a “One stop shop” for data  
      2.5.4 Strategic Initiative 4: Business Information and Analytics Team  
      2.5.5 Strategic Initiative 5: Develop an Enterprise-wide Operational Data Store (WODS)  
      2.5.6 Strategic Initiative 6: Future Initiatives - An Innovation Data Environment for Academia and Research (IDEAR)  
      2.5.7 Strategic Initiative 7: Future Initiatives - Capturing Behaviour Data  
      2.5.8 Strategic Initiative 8: Developing the Data Warehouse  
      2.5.9 Strategic Initiative 9: Reducing Technical Debt  
   2.6 A Summary of our roadmap  
   2.7 Our Non-Functional Requirements to support these initiatives  

Appendix A - Action Plans
1. Overview

1.1. Context
The University of Waikato, like most organisations, has a significant reliance on data to be able to operate effectively, make good decisions and achieve a strategic advantage over competing institutions.

Society has become conditioned to expect services in real time, of high quality, and with immediate results. Anything less leads to dissatisfaction and risks to brand and reputation. Education establishments are not exempt for this, and therefore it is more important than ever that data is managed and treated with the urgency that is required. This is particularly important for data supporting our core services in teaching and learning, research and business operations.

The University cannot afford to be complacent and do things “our own way” when dealing with data. The organisation needs to standardise its approach to managing data and follow good practice in data governance, management and design.

The University Strategy, published in February 2017, identifies priority areas that need attention, and provides a roadmap to align activities to, with a clear set of outcomes in mind.

The ICT Strategy directly identifies data and information management as an essential capability that is critical to the delivery of the ICT strategic objectives. The KPMG current-state analysis (undertaken in late 2016 as part of the ICT Strategy development) also identified systems architecture and information management as an area where immediate focus was needed to ensure it is fit for purpose going forward.

“Systems’ architecture and information management - There also was a strong theme, in the current-state feedback, regarding the opportunity to establish an enterprise-wide strategy and approach, to systems’ architecture and information management. This is, again, multi-faceted, and includes the opportunities to drive improved process integration through ICT - promoting integrated systems with consistent data, to ensure access to accurate/relevant information. There are related issues with respect to managing the University’s intellectual property more effectively, and, again, directly supporting improved outcomes for teaching and research, in particular.” [KPMG report, January 2017]

This document is designed to provide a framework for data practices and procedures, and operate effectively within the boundaries of the ICT Strategy. It applies to all technology projects implementing change, and existing business as usual (BAU) technical support that implements change. Whenever data is changed, or the opportunity for change and simplification and standardisation arises, this framework should be leveraged.
1.2. Current State Analysis

Data within the University of Waikato is a key asset in the traditional data sense; in that it assists to determine whether the organisation is performing effectively in terms of decision making and delivery of operational services.

Data is also an asset in the form of research knowledge produced for internal or external consumption as Open Data, and Data as a Service to key stakeholders and industry sectors. There is a need to partner with external stakeholders (e.g. NZ Police), to enable joint research on shared data.

An assessment of the current state of data management at the University has identified it as ‘fairly immature’, and the following pain points are commonly experienced across the organisation:

- **Key business processes are at risk of disruption**
  There are no resilient measures or designs in place to deal with unavailability of target systems. E.g. Lack of caching could result in fees payments being disrupted because of maintenance of downstream systems.

- **Multiple sources or versions of data**
  Duplicate data is at risk of being created or changed in several applications. Data quality is an issue. e.g. Student data in JADE, Raiser’s Edge, Moodle etc; as is replicated data (in different states) in different data stores e.g. Students in Interface Repository, OraJasper etc. Students can be duplicated in our UniAccess/Identity Management System which is responsible for giving out Usernames, and the vast number of active members in there (150 000 currently) can cause a lot of distraction and complication in filtering out and affecting BAU process efficiency.

- **The quality of data is unknown**
  Data is not fixed at its source. This are little or no reconciliation processes, data quality checks, or auditing of exceptions in a consistent manner. The first indicator of an issue is in many cases, the communication from the business/student users.

In very high level scans it has been discovered that there can multiple versions of the same student in the same systems, some with different student IDs, some with the same ID. Also the volume of students and staff (people) in our peripheral systems far exceed our total number of people at present - Identity Manage for example holds 150 000 people, all active with user names that makes processes and filtering complex.

- **It takes a long time before data can be used**
  Data goes on a long journey before it is available for consumption by other systems, sometimes taking many days, when it is needed more urgently. Most integration is performed by overnight feeds as a default solution e.g. student registration and printing of IDs cards, PO Approval, Student Library updates.

- **There is a lot of useful, but unused technical data**
There is a strong need to utilise less formal and less structured data from various technology logs or external sources that could provide sufficient answers. There is no common approach for collecting and analysing data, or making it available and usable as part of the business process.

- **Data is not available**
The University is rich with innovation, which can create new insights and a truly unique status. However, this is constrained by most of the transactional data being “locked and/or protected” in information systems not accessible to everyone.

Cross system reporting, particularly for transactional data for students and research, and data for support services reporting is extremely limited. There is a need for importing and storing open data sets or data from external organisations in a consistent manner that is accessible by everyone (as appropriate). Currently there is a lot of duplication of effort in this area. For example, educational count statistics open data set, Interfaces Repository or OraJasper were designed as an operational data store, but very little of the data is transactional data. There is a high demand for Moodle data to have student behaviour patterns, but access to the database through design and security is not possible.

- **No University-wide data intellectual property approach**
The design, transportation and usage of data is in silos, both within the business/student areas as well as the ICT areas. The same data is shared between all these areas but there is no consistency of definition, rules, who has responsibility for owning it, defining it, or classifying it (such as secure or not). Therefore, pragmatic decisions cannot be made on the usage of data, or the impact, or the ability to leverage existing standards.

**1.3. Looking forward - Where do we want to be?**
Data is acknowledged to be a key enabler to the success of the University.

The ICT Strategy specifically identifies ‘Unlock the value of information’ as a direct opportunity for the University. The development of the Data Framework provides the structure, standards and practices that will ensure data is of a high quality, well managed, and accessible for insightful problem solving, performance monitoring and decision-making.
2. The Data Framework

2.1. High-level UoW Data from 30,000ft

The diagram below is a high-level view of the University of Waikato data architecture.

There is a core domain where business data is mainly created and distributed to the rest of the organisation (identified in the centre of the diagram as Research Management, Student Management and Support Capability). This core is described as the Shared Enterprise Data core (SHED). This area has a strong need for robust master data management and data integration practices. It is effectively the heart of the organisation, pumping data through the arteries of the University. If data becomes corrupt here, it affects the whole organisation.

The peripheral zones of Community, Learning Delivery, and Research Execution have a more niche need for systems, and consume master data, but do not provide it back to the core.

The yellow arrows are the services that are delivered, and are the end point of our business processes. Services are delivered either directly from the Shared Enterprise Data core (SHED) or from the niche areas, all of which are heavily dependent on master data for identifying the correct students, staff, papers, lecture rooms, partners, and so on.

2.2 Data Goals

It is recommended that the following Data-driven goals be applied to the University as a whole:

1. Shared Enterprise Data is distributed in real time, of high quality, when and where it is needed
2. An enterprise-wide perspective of data is available; in one place, to all approved users who need it, and is not constrained or protected by access to specific systems.

3. A master data management approach is adopted, in terms of accountability, ownership, stewardship and governance of data that is consumed organisation wide.

4. Development of an informal data innovation platform to answer one-off questions.

5. Data management (University-wide) follows standards and procedures aligned with the ICT Strategy, so that the organisation as a whole benefits from consistency of skills, toolsets and practices, and work can be collegial and innovation can occur off those interactions.

6. Data security and compliance is defined, available to all, and adhered to.

2.3 Data Architecture Principles
ITS have adopted a set of Enterprise Architecture Principles that can easily be applied to Data Architecture.

- “There can be only one Master” - Only one application can create and modify master data.
- “Simplify” - Reduce complexity in technology and design for interfaces.
- “Reuse before Buy before Build” - Create interfaces and data stores that can be re-used.
- “Reduce Risk” - Don’t lose data, don’t modify data, ensure data transport is resilient.
- “Partnership with Business” - Deliver data value in WODS, and accountability in MDM.
- “Application [Data] is for Life” - SHED/EMD is not specific to an application or solution - think of its longevity.

2.4 Strategic Initiatives - Medium term initiatives to realise data goals
The approach is to implement a data centric framework, within the context of the University Strategy and the ICT Strategy, where there is a focus on data, dealing with quality and urgency.

The following initiatives have been identified to enable the University to achieve its desired future state for data management within the context of the Enterprise Architecture framework and are directly traceable to the ICT strategic opportunities of:

- Alignment
- Efficiency
- Partnerships
- Innovation, and
- Emerging Technologies.

The strategic initiatives promote a culture change in terms of data ownership and responsibility, and will provide the foundation on which the University can deliver core data services to students, research and internal operations in a manner that is consistent with good practice and modern organisations. These initiatives will also help develop positive student, research and staff experiences, and contribute toward the University’s strategic differentiation.

1. Consolidate technologies and complex techniques for data transportation / integration
This would be replaced by three basic patterns, with an agreed technology to implement each pattern, developing our technical skills to design, implement and support these patterns.
2. **Introduce Master Data Management (MDM)**
   This is not only a technical solution, but leverages off the three basic patterns that we establish. It needs business commitment and resolve to solve the problems at the root of our business pain. It needs architectural guidance and governance to ensure that both the technical and business architecture is in place for the solutions. It requires business and ITS partnership.

3. **Develop a Data Services Capability**
   We have no one person dedicated to the wellbeing of data in the organisation at an enterprise level. This initiative will correct that and create a dedicated group to be responsible for the governance, design and delivery of data to various consumers, whether it be applications, digital delivery channels, or reporting. NOTE: They will not own the data; the business owns the data.

4. **Expanding the scope of the Business Information and Analytics**
   This team will provide reporting, reporting support, reporting layers, self-service training, and business intelligence presentation components across the organisation, whether it be operational reporting, or strategic reporting. They will work closely with developing the Data Services.

5. **Developing an Enterprise Wide, Operational Data Store**
   That is a pool of all operational data whether it be master, transactional or reference. This will provide organization wide context reporting on how we are performing with minimum latency from source systems. This will be a pool of READ ONLY data for information queries.

6. **Future Initiatives: An Innovation Data Environment for Academia & Research (IDEAR)**
   A platform for one off questions and research either by ITS or by the University as a whole that share technology and storage, with a specific focus on learning and research insights, capabilities and skills. This will accommodate multiple sources of different formatted data, where traditional data architectures would be cumbersome and expensive to implement for one off questions.

7. **Future Initiatives: Behaviour Data**
   Providing a data behaviour store that extracts and consolidates simple, high volume data from various technical resources throughout the university such as technology audit log files, WiFi, account and video usage. This is a rich information source that can give very specific behaviour patterns for usage of University digital learning material, popularity of papers, behaviour of students, and use of University resources by students. This initiative would directly support innovation and insight around students.

8. **Develop a Data Warehouse**
   Extension of the existing data warehouse to meet the needs of the entire organisation and external open data sources, as well as fix existing anomalies in the current proof of concept schema.

9. **Reduce technical debt**
Decommissioning and simplification of our existing data architecture. We have left over interfaces and stores, and much complexity with the currently deployed architecture. These will gradually be removed and replaced by the previous initiatives.

2.5. Detailed Approach for Strategic Initiatives
This section drills down on the previously defined Strategic Initiatives.

Within the context of this document, an information system is considered a COTS or custom built application that either creates, modifies, deletes, archives or holds business data that is used by a business process. In other words, it does not include systems that include unstructured data such as Email, Google Drive, Word processing, etc.

The following diagram would make the data vision tangible in terms of the BAU perspective of operations.

2.5.1 Strategic Initiative 1: Consolidate technologies and complex techniques for Data Transportation/Integration
Integration is to be simplified and standardised, and made reusable as much as possible. To deliver all of the previous stated innovations, our complex point to point, batch architecture needs to be standardised and commoditised. It needs to support real time data transfer so that business process efficiencies can be gained, while not compromising data quality.

Almost all of our data movement currently is realized through “feeds” (i.e. bulk data transfers) at scheduled times, usually daily.
We need consistency of data integration for our master data strategy and reporting initiatives and also need to narrow our skill set from a broad range of technologies and standards, to a smaller, simpler, easier to maintain, easier to find skill set of mainstream technologies.

Integration Initiatives
- **Enterprise Scheduler** - Adoption of an enterprise scheduler so that all jobs are run from one single place. This provides tremendous simplification of the existing topography where there are lots of different methods and distributions for scheduled jobs. It also reduces risk as everything is centralised from one technology, can be monitored in once place, and alerts support to any failure of scheduled processes.

- **Integration Patterns** - The needs of the University fall into the following categories:

  - **Bulk/batch ETL movement of data at periodic intervals.** Typically this is transaction data, and would move large loads of data point to point between systems. This would typically be implemented (though not exclusively) through an intermediate ODS (WODS) depending on circumstances (if the data would be useful beyond integration for example). This would be realised by Microsoft SSIS as the ETL tool.

  - **Publish/Subscription mechanisms.** Near real time master data transmission. Master data is modified in the source system, validated and (if considered of worthy quality), published to the organisation. The data would be sent to the WODS so that master data is always current and sent to any necessary information systems that are capable of taking and updating the data in real time.

    The technology realisation of the publish subscription mechanism is through:

    1. **Subscriber** - MuleSoft ESB for Subscriptions, where subscribers to a published real time message can be drawn up quickly.
2. **Publisher** - There are 3 key questions when considering a publisher:
   a. Does the source master application have a hook or listener? - Most applications do that allow an event like data change to be captured and published.
   b. Can a service be run that polls for changes and then publishes these changes into the JMS queue?
   c. Can a trigger be added to pick up delta changes in data, which are then fed to a staging table and polled?

- **Queue** - Apache MQseries using JMS industry standard for queues and topics. The usage of queues will reduce the risk of losing data when systems are not available, as they will be cached instead.

### 2.5.2 Strategic Initiative 2: Introduce Master Data Management (MDM)

There are several patterns for master data management, these include a “Hub” approach and a “Synchronization” approach to name a couple.

The University has agreed to implement a simpler approach that does not require sophisticated technology, where only one application is the master (i.e. where it is created and maintained), and this “mastering” results in the propagation of data to other consuming systems.

The benefit of this “lighter” approach is that one area is responsible for maintenance & ownership of the data, instead of change being slowed down by several “governors”, and technology implementation is a lot easier.

**Master Data Initiatives**

- **Master Data Business Architecture** - Business Involvement in defining rules, ownership and governance of data and cleaning. These tasks may be done now in some cases, but this process is to make it more formal. The underlying principle is that the University function that creates the master data owns it, and is responsible for its quality and definition - not
ITS. The implication therefore is close business partnership to implement this initiative and this should be ingrained in any new project introducing information systems.

- **Conceptual Data Model** - All master data entities as a minimum need to be defined, so that everyone agrees (at least at a high level) what they entail, so they can be allocated to be ‘mastered’ somewhere. We are working closely to the CAUDIT reference model for business reference, however we have moved away from CAUDIT in terms of data entities as we are only interested in master data items, and business object terminology that we are familiar with.

- **Roles & Responsibilities** - The business needs a business owner to consult to make decisions and provide feedback to make the process more efficient, and a data steward to take responsibility for entering the data as it should be (note this is not considered any more of a burden than normal BAU related activities). ITS would have responsibility for the definition, design and intellectual property around the data, and would be key stakeholders and consultants.

- **Data Dictionary of Master data** - Creating a definition of all master data used in the organisation (only that which is Shared Enterprise Data (SHED)), that can be quickly and readily accessed, is accurate, and defines business rules and owners. There should be one place of contact for any questions about how the data should be treated. The data dictionary will also have security classifications that will generate “cookie cutter” patterns depending on if the data being transported is considered (for example) “sensitive”.

- **Data Quality checking and Verification** - Data while it is being created or exported to other systems, will be validated against business rules, if it does not comply, it will immediately be sent back to the business area that owns it for correction, probably via E-Mail. It would only be allowed to be consumed by other systems once it is fixed.

### 2.5.3 Strategic Initiative 3: Develop a Data Services Capability

The governance and design of data would be consolidated into two data centric teams whose primary focus would be on data - defining, designing, transportation and Information delivery. These teams would be Enterprise wide, and not specific to any one domain.

These technical capabilities would be provided by the following teams:

- Data Services team
- Business Information and Analytics team.
Data Services Team - Developing a “One stop shop” for data

This team will be a core foundation of ITS and provide data services to the University of Waikato. It can be “scaled out” using external or internal staff, ensuring that everyone follows the same policies, guidelines and approaches when dealing with data.

*External staff may also include temporary partnerships in faculty areas where innovation can be harnessed and standardised for the benefit of the whole of the University of Waikato by working in partnership, adopting the same practices, technology and standards as the DST for one off projects.*

The team will be part of ITS, and provide the following functional capabilities:

- Data design (data modelling & real time messaging, 3NF, star schemas etc.) in WODS and the data warehouse
- Data transport (ETL ownership)
- Data definition, classification
- Cube definition
- Setup and ownership of future initiatives including governance and technology
- Master data governance stakeholder
- Business partnership and liaison.

The roles would include:

- Senior Data Designer/Modeller - Managing and Designing data at a Logical and Physical Level, and working collaboratively on conceptual data with the Enterprise Architect
- Data developer (ETL, Pub/SUB/Messages and Schema) - Implementing data design and transport
- Business Data Analyst - Business engagement
- Enterprise Data Architect (Part of EA role) providing data leadership & governance.
2.5.4 Strategic Initiative 4: Business Information and Analytics Team

The team provides an enterprise wide reporting capability for OPS/reporting and strategic reporting from the WODS and data warehouse, leveraging of the data availability provided by the Data Services team, and also training staff members for self service.

The functional capabilities would typically include:

- Report models
- Pivot table technology
- Dashboard
- Report analysis and training
- Data quality reports - with the emergence of master data, we need data quality and correlation reports to monitor the integrity of the data
- Technology used for data presentation
  - Microsoft SSRS & self-service business modelling layers
  - Microsoft SQL Server
  - Oracle (to be retired)
  - Microsoft SSAS
  - Microsoft PowerPivot
  - Microsoft Excel
  - Native Applications

The roles would include:

- Reporting Analyst - A semi-technical person who solicits reporting, self service and BI information needs, engaging with the business, and providing specifications as to what is needed by leveraging of currently available reporting model, making requests for new, and extending existing reports and functionality
- Lead Report Developer
- Data Experience Manager (not just BI)

**BAU Data Operations**
A sub team that may be independent or part of Data Services, or can be internal or outsourced. Looking after the daily operation of the database.

Duties of a general Database Administration role would typically include (but not restricted to):
- Scheduling of database backups and transaction logs
- Patching/ database upgrades etc.
- Across all faculties
- Performance tuning
- Providing access to databases and database objects
- Weekly/monthly housekeeping schedule jobs
- Creation/destroying of database instances.

**2.5.5 Strategic Initiative 5: Develop an Enterprise-wide Operational Data Store (WODS)**
Making the right information available to everyone with the proper authorisation.

**WODS Initiatives**
The below diagram shows the Technology Architecture for WODS.
The aim is to consolidate all of the various reporting areas we have into a single, reliable, verified enterprise-wide operational data store (WODS), that provides integrated reporting from data sources that would otherwise not be connected.

Note: this would be operational data only, by definition supporting current process therefore high latency (running overnight batches) will try and be avoided where possible.

The aim is to make data accessible to authorised staff and students who need it.

Features

- A UoW focussed, application neutral logical data model
- READ ONLY, transaction and master data
- Self service reporting to leverage off the innovation and capabilities of our highly skilled and highly motivated staff in order to satisfy “one off, ad-hoc transactions”
- “Packaged reports” that support business operations and strategy
- Used as a facility for integration between systems
- Providing Data as a Service to external channels for read only/information e.g. web sites. This will be initiated by creating a portfolio of key services such as GetStudent, GetPaper, GetStaff etc. that are accessible to the whole organisation for data consumption purposes
- Providing the main source of integrated, validated, quality data as a source to business intelligence/strategic reporting
- Working with faculty - Faculties have their own innovation data ideas that can be shared with the rest of the organisation. The WODS will allow them to contribute to the data model using architecture, technology, standards, patterns and practices, and provide something for the University to benefit from.
**Strategic Reporting Initiatives**

This would be provided by a University data warehouse, which is more analytics, aggregated and self-service driven in nature and will traditionally provide the following capabilities:

- Dedicated data stores with Star & Snowflake Schema
- Canned reports
- Self-service for cubes and reports.

---

**2.5.6 Strategic Initiative 6: Future Initiatives - An Innovation Data Environment for Academia and Research (IDEAR)**

As part of the development and supporting of innovation within the ITS and University strategy, ITS would be supporting the idea of individual or collective innovation within data.

All of the previous initiatives are focussed around dealing with current problems and pain points of data, these next two initiatives are focussed around adding something different to the data radar to deal with existing pain points of wanting ad-hoc answers to information we just don’t currently or formally manage.

The plan in the medium term is to leverage off the innovation that our institution specialises in, and instead of everyone working in discrete pockets of technology and skills, with discrete data sets, a collaborative environment would be set up with technology, skills, operations and infrastructure enterprise standards.

A cloud based or in-premise platform would be set up that allows internal innovators access, where multiple datasets can be brought together to support research, or ask previously unanswerable business questions.

The platform will not have a fixed data schema like WODS or business intelligence, instead it will consist of a lot of queries on different data sources types such as CSV files, open data sources, open APIs etc. and allow them to be correlated with internal business data for one off queries.
The benefits of having one platform include:

- A common skill mode, technology mode, licensing model, infrastructure mode, support model etc that the whole University can leverage and benefit from cost savings. This focus is on innovation rather than standing up environments.
- A common data where data can be made available through published notebooks with context, as open data files, able to access and consume APIs from a central point, instead of localised security configurations.
- A single published “point of truth” to see the context and answers of previously asked questions.
- A possible new service that can provide data science as a service to the community.

This platform would resemble a data lake in Big Data terms. Except the data is not expected to be Big Data (although can be scaled up through Cloudera etc.) in future if we decided to.

This may require future development of the Data Services team’s capability and need for skills in future.

The vision of the IDEAR is illustrated in the below diagram.

2.5.7 Strategic Initiative 7: Future Initiatives - Capturing Behaviour Data

There is a need to capture large volumes of event data from technology and application usage such as Learning Management Systems, Internet, Active Directory usage, etc. to capture what students are doing and how successful they are at doing it.
This large volume of typically unstructured data (i.e. no formal database definition of data) needs to be accessed and queried randomly, often by ‘canned’ queries or reports in a rapid manner. This is typically the features offered by a NoSQL database, where the data definition is defined during the query not in advance.

Unlike the IDEAR initiative, this solution would focus on creating re-usable, re-queryable data, rather than random data brought together for a specific one-off problem to solve. This solution would involve repetitive reports and models that can be re-run, and that may consume other business information (WODS or data warehouse) or indeed populate them with its findings such as getting student usernames, and providing back summaries or KPIs on their behaviour. This type of initiative is very close to various student insight models that are available commercially.

2.5.8 Strategic Initiative 8: Developing the Data Warehouse

The data warehouse was set up as a proof of concept by Altis to demonstrate the use of business intelligence technology and was very focussed around students in terms of finances.

There is potentially some re-design work to be applied however (for example there is no student fact, only a dimension, and the student dimension that does exist contains far too much operational data).

Therefore the expansion of Star Schema (facts, dimensions and re-design of some of the existing model) is required, plus expansion of the schema to include enterprise wide needs and open data sources that would be maintained and collected to provide context.

In additional the development of new data cubes and presentation analysis in SQL Server as our future database road map.
2.5.9 Strategic Initiative 9: Reducing Technical Debt

For integration, there will be a slow movement from old to new. All new projects that are initiated will adopt these integration techniques, and can then apply master data practices and leverage of WODS as a constraint to each new project (including small projects).

A decommissioning strategy does have to be adopted to look at decommissioning existing integration particularly with reference to technologies such as:

- Hyperion reports
- Interfaces repository
- WMS - ODS
- IDM
- MyWeb
- WMS
- ORAISG Jasper ODS
- Oracle SOA Suite
- Oracle Service Bus
- Oracle BPM
- Current data warehousing technology

A plan should be put in place that defines the exact current state of these components (this is mainly fulfilled now in SPARX Enterprise Architect but parts are inaccurate and may be incomplete). Their precise dependencies on LMS and SMS projects need to be determined, and the information used to help steer a decommissioning project, absorbing any old decommissioned interfaces or applications that are still deployed.

There is a lot of risk in decommissioning old repositories and data stores because they are involved in daily processes and integration, and they have data sources in various applications. Therefore, a carefully laid out plan of dependencies needs to be created for decommissioning and associated action plans.
2.6. A Summary of our roadmap

Operational Data Integration

- Source System
- Application
- Instant Informing
- Bulk Data (Transaction Data)
- Publish Change
- Real Time Consumption
- Real Time Digital Channels
- Target Systems
- Consistent Delivery Channels
- Used By Applications
- Bulk Data Transport
- Strategic reporting
- BI (Data Warehouse)
- Web Sites etc.
- Batch transactions (if needed)

Data Information Availability

How Are we performing Strategically?

How are we behaving?

How Are we performing Operationally?

That is an unusual question?
2.7. Our Non-Functional Requirements to support these initiatives

The following are a list of “cookie cutter” non-functional requirements that can be distributed to any project, or vendor, in supporting this roadmap:

<table>
<thead>
<tr>
<th>ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>#10</td>
<td>No data can be modified in transit - it can only be modified and “corrected” at the point of source</td>
</tr>
<tr>
<td>#20</td>
<td>Everybody knows exactly what our Master data is in tangible terms, that they can access and verify, and gain the context.</td>
</tr>
<tr>
<td>#30</td>
<td>There can be only one Master Data source, where the data is created and modified.</td>
</tr>
<tr>
<td>#40</td>
<td>All defined Enterprise Master Data/SHED MUST be owned by a Business User, and managed by a Data Steward</td>
</tr>
<tr>
<td>#50</td>
<td>Dirty Master Data does not leave the Master Application, if data is deemed invalid, or incomplete by business defined rules, the Data Steward is notified that it needs correction.</td>
</tr>
<tr>
<td>#60</td>
<td>The Business Own the Data. They own decisions on its meaning and changing definitions. ITS ensure its packing, delivery and receipt, and risk mitigation in this process.</td>
</tr>
<tr>
<td>#70</td>
<td>No data shall be ignored or rejected while in transit, without notification to a person. No data shall be lost without recording. Data shall be resilient in transit and any outages and issues should not result in loss of data.</td>
</tr>
<tr>
<td>#80</td>
<td>Data should be delivered at the time it is needed by the business, either scheduled or real time.</td>
</tr>
<tr>
<td>#90</td>
<td>Any application we procure or build, MUST support:</td>
</tr>
<tr>
<td></td>
<td>Data extraction for ETL (Either direct connection to the database, or a method of extracting data, or available APIs if a Cloud based product)</td>
</tr>
<tr>
<td></td>
<td>Having hooks or listeners within code to write Publication Mechanisms (On premise or Cloud)</td>
</tr>
<tr>
<td>#100</td>
<td>Any COTS System should have a configurable User Interface to allow Business Data validation rules to be applied on entry. Where this is not possible in existing legacy systems, greater emphasis is placed upon the Business rules validation on the Publisher.</td>
</tr>
<tr>
<td>#110</td>
<td>There should be no schema changes whatsoever, or removal of referential integrity on data models. DBA tuning and performance excluded from this constraint.</td>
</tr>
</tbody>
</table>
## Appendix A - Action Plans

**Action Plan from an Enterprise Initiatives context**

<table>
<thead>
<tr>
<th>ID</th>
<th>Action</th>
<th>Description</th>
<th>Owner</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>EA#10</td>
<td>Set up Data Services Group</td>
<td><strong>Below would need to confirm numbers with investigation of demands and workloads</strong> Setup Data Services Team</td>
<td>Eion?</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 Data BA 1 Senior Data Designer 2 Developers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EA#20</td>
<td>Align capability and capacity BI&amp;A with framework</td>
<td><strong>Setup BI/Data Experience/Presentation Team</strong> 2 Report/Cubes Developers (or 1 Report Developer &amp; 1 Cube Developer) 1 Reporting Analyst/trainer 1 Manager Develop and train capabilities.</td>
<td>Campbell?</td>
<td></td>
</tr>
<tr>
<td>EA#30</td>
<td>Deploy Infrastructure</td>
<td>Deploy Integration Infrastructure for SSIS, Apache MQ, Mule ESB - Physical servers for Dev/Test/Production &amp; Technology</td>
<td>Ian?</td>
<td></td>
</tr>
<tr>
<td>EA#40</td>
<td>Operational Architecture</td>
<td>Defined a model for Technology Support - outsourcing or in-house for data integration technology.</td>
<td>Ian/Eion?</td>
<td></td>
</tr>
<tr>
<td>EA#50</td>
<td>Operational Architecture</td>
<td>Evaluate the need for an Enterprise scheduler, and requisition one if needed - Open source vs COTS to centralise data integration.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| EA#60| Consolidate HR                                  | Options Paper for consolidation or simplification of HR into a single staff based System :  
|      |                                                  | • Alesco HR  
|      |                                                  | • IDM  
|      |                                                  | • UniAccess  
|      |                                                  | • Oracle HR | Gary?  |
| EA#70| Decommission Plan                               | Devise a roadmap to decommission Repositories and Databases, and decommission interfaces & applications on these databases. If systems cannot be decommissioned, look at alternatives to simplify. | Gary?  |          |
EA#80 | Decommission ESB | **NOT SMS/LMS DEPENDENT**
| | | Devise a roadmap to decommission existing integration ESBs - Oracle ESB, Oracle SOA
| | | Ian?

EA#90 | Decommission Hyperion | Decommission Hyperion - Divide all Hyperion into
| | | • Reporting - Migrate to MS SSRS
| | | • Cubes - Migrate to MS
| | | • ETL - Migrate to SSRS
| | | Ian?

EA#100 | Pub/PUBs Policy and patterns based on POC | Complete POC process.
| | | Develop design patterns & policies.
| | | Socialize with projects.
| | | All Architects

EA#110 | Enterprise Web Content Options | Put together an options paper on the advantages/disadvantages of Enterprise Web Content vs current model of local initiatives, technology, and discontinuous user experience.
| | | Gary

**Action Plan from a SMS/LMS Initiatives context**
The following are a list of key tasks that have to be performed applying the strategic approach to the LMS/SMS context:

<table>
<thead>
<tr>
<th>ID</th>
<th>Action</th>
<th>Description</th>
<th>Owner</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>SL#10</td>
<td>High level architecture</td>
<td>Provide high level Approach to define Technical roadmap for transition from current state to future state for SMS/LMS including dependencies and application architecture.</td>
<td>Simon?</td>
<td>2 Weeks</td>
</tr>
<tr>
<td>SL#20</td>
<td>SITS SA Artefacts</td>
<td>Develop a Solution Architecture roadmap of artefacts for SITS. e.g. Project too big and too agile to have one, development work already started without one.</td>
<td>Gary, David, Vanessa</td>
<td>1 Day</td>
</tr>
<tr>
<td>SL#30</td>
<td>Review of MyWeb/Edlink</td>
<td>Detail review of MyWeb and Edlink design &amp; develop a migration strategy.</td>
<td>Simon?</td>
<td>2 weeks</td>
</tr>
</tbody>
</table>
| SL#40 | WODS | Develop WODS Data Model with student perspective that takes all areas into considerations:
  - Integration Needs
  - Enterprise Reporting | Data group in collaboration with SITS Data specialists | 2 months in total for all domains |
<table>
<thead>
<tr>
<th>SL#</th>
<th>Task</th>
<th>Description</th>
<th>Responsible</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>Student Insights</td>
<td>Provide an options paper on Student insights &amp; alternatives</td>
<td>Gary</td>
<td>3 days</td>
</tr>
<tr>
<td>60</td>
<td>Business Architecture</td>
<td>Business Engineering. Review of proposed application capabilities, process reengineering at a programme level to see if there are any missed opportunities, and defining cross application processes.</td>
<td>Programme BA</td>
<td>4-6 weeks</td>
</tr>
<tr>
<td>70</td>
<td>BI Review</td>
<td>Review BI Data Warehouse, ensure the architecture and design is fit for purpose for SITS. There are numerous issues - very transactional operational data store based. Contains lots of redundant information like address, telephone numbers etc which is not strategic. High Level Review <em>Contains sensitive data such as Passport numbers of students.</em> Students are not facts, so students cannot be reported upon unless in a financial context. Dimensions are not hierarchy, and are dependent on other dimensions (bad practice). There are over 250 000 students in the system. Summary - Not designed for Cubes or OLAP analysis, only OPS reporting</td>
<td>Gary</td>
<td>5 days</td>
</tr>
<tr>
<td>80</td>
<td>Data Quality</td>
<td>We have large volumes of data shared between systems, i.e. Staff and Students, of numbers varying between 12,000 and 200,000 in IR, Jade, BI, Orajasper, IR, IDM. Also we have many duplicates.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A strategy is needed to determine how to clean duplicates and bad data, and do we need to reconcile numbers and introduce enterprise logic to determine current vs past vs future?

The SITS program of work contains a core of LMS & SMS and both have respective Solution Architects. Outside of this however, is a peripheral of systems that affect SITS, that have to be decommissioned in a controlled manner, these are more Enterprise Architecture perspectives collaborating closely with the core SITS Solution Architects, hence some of the tasks above.