

DATA OFFICE

Data Governance at Ordnance Survey

Jonathan Holmes

Date :

5th December 2018



The DAMA Body of knowledge

To establish the big picture in terms of the current effectiveness and maturity of data management capabilities at OS, we performed an assessment against a standard industry view of data management. For this we use the Data Management Body of Knowledge (DMBoK) published by the Data Management Association (DAMA) International.



The DAMA-DMBoK is an internationally recognised framework that describes the commonly accepted good practices, widely adopted methods and techniques (without reference to specific technology vendors or their products) that can be used to: guide discussions on each of the data management capabilities; and understand and diagnose an organisation's data management challenges.

While the DAMA-DMBoK defines the requirements and activities for effective data management, it is not prescriptive about how an organisation should achieve these capabilities.

The framework is comprised of 11 required capabilities to manage data:

1. **DATA GOVERNANCE:** Planning, supervision and control over data management and use.
2. **DATA ARCHITECTURE MANAGEMENT:** Defining the blueprint for managing data assets.
3. **DATA MODELLING & DESIGN:** Analysis, design, implementation, testing, deployment, maintenance.
4. **DATA STORAGE & OPERATIONS:** Providing support from data acquisition to purging.
5. **DATA SECURITY MANAGEMENT:** Insuring privacy, confidentiality and appropriate access.
6. **DATA INTEGRATION AND INTEROPERABILITY:** Managing the combination or association of different sources of data together to produce useful business information.
7. **DOCUMENT AND CONTENT MANAGEMENT:** Managing data found outside of databases.
8. **REFERENCE AND MASTER DATA MANAGEMENT:** Managing golden versions and replicas.
9. **DATA WAREHOUSING AND BUSINESS INTELLIGENCE MANAGEMENT:** Enabling reporting and analysis.
10. **META-DATA MANAGEMENT:** Integrating, controlling and providing meta-data.
11. **DATA QUALITY MANAGEMENT:** Defining, monitoring and improving data quality.

Maturity levels

Capability	Insufficient	Sufficient	Differentiating		
Data Governance	<ul style="list-style-type: none"> No clear data ownership assigned. Data Owners, if any, evolve on their own during project rollouts (i.e. self-appointed data owners). No standard tools or documentation available for use across the whole enterprise. 	<ul style="list-style-type: none"> Data Ownership Model does not exist. Owners commissioned in the short-term for specific projects & initiatives. Often department or silo focused leading to ownership by "Data Teams" or "Super Users" that manage "all" data. 	<ul style="list-style-type: none"> Defined Data Ownership Model exists. Ownership Model is loosely applied to key data entities. Limited collaboration. Not fully 'bought in' to data ownership at an enterprise level. 	<ul style="list-style-type: none"> Data Ownership Model is implemented for the key data entities. Collaboration between stakeholders in place. Governance process regularly reviews this model and its application, updating and improving as needed. Benefits begin to be realised. 	<ul style="list-style-type: none"> Data Ownership Model has been extended such that the majority of data assets are under active stewardship. Effective governance process employed by stakeholders & stewards. Well defined standards adopted.
Data Architecture	<ul style="list-style-type: none"> Limited awareness of the information needs of the enterprise as a whole. No enterprise or conceptual data models. Frequent disagreements between business users and IT due to no common language. Data Integration and BI architecture where it exists, is developed in silos for individual projects. 	<ul style="list-style-type: none"> Enterprise recognises the need for common terms and definitions to be used in business models across the enterprise. Initial attempts to define a conceptual or enterprise model for the entire business. A level of consistency develops between projects and how they access and integrate data. Design patterns start to emerge but there is no formal mechanism for ensuring these are followed. 	<ul style="list-style-type: none"> An EDM captures and defines enterprise information needs and data requirements. Information value-chain is understood. Data definitions are standardised across the enterprise. Best-practice architecture patterns are used as the basis for data integration, BI and MDM architecture. Standardisation of tools, technology standards and protocols takes place. 	<ul style="list-style-type: none"> Business models (e.g. process and organisation) align with the EDM. Enterprise understands how data problems in key processes will adversely affect it. Data architecture is developing into an enterprise wide initiative that encompasses data integration, DW/BI, MDM and Meta-Data. Enterprise Taxonomies, XML Namespaces and Content Management standards have been defined and align to the EDM. 	<ul style="list-style-type: none"> Stable and mature EDM is adopted as the authoritative source of information items and definitions. Information elements shown in business and system architecture models are directly traceable to the EDM. A mature enterprise data architecture exists that encompasses data integration, DW/BI, MDM and Meta-Data. There is a strategic approach to managing the EDM and enterprise data architecture with effective Architecture governance forums in place.
Data Modelling & Design	<ul style="list-style-type: none"> Models, where they exist, are developed in silos for individual projects. Limited re-usable data access or integration services. No standard tools or processes for data migration and conversion. 	<ul style="list-style-type: none"> High level data requirements have been established. Data models that exist are reasonably complete in terms of entity and attribute definitions. Re-usable data services are starting to be developed and tested with an attempt to develop common schemas. 	<ul style="list-style-type: none"> Data requirements are fully documented and business rules are trusted. Modelling standards and modelling tool of choice have been defined. Some data models have been extended to include additional metadata. Processes are introduced to control development and test of data and systems. 	<ul style="list-style-type: none"> Conceptual Data Models have been defined and are used as reliable references. Well populated data models exist at the logical and physical levels. Data access and integration services are formally defined. Formal design processes are used for data models and information products. 	<ul style="list-style-type: none"> Models and database designs are controlled against published model and DB design standards. Standard additional metadata is present on models. Formal and effective design, review and change control processes are in place. Data modelling competencies recognised and training / mentoring established.
Data Storage & Operations	<ul style="list-style-type: none"> Limited or no active management of data in operational or long term storage. Data managed on ad-hoc basis in spreadsheets or databases with no clearly defined storage and operations policies or roles defined. Business critical data exists in storage without oversight from DBAs. The operations to be performed on data to manage it through its lifecycle are not defined or implemented. 	<ul style="list-style-type: none"> DBAs may be responsible for some business critical data, but with fragmented and inconsistent responsibilities, standards and policies in place. Awareness of the need for data operations management gains recognition within the enterprise. Data storage, availability or performance issues are commonplace. Some production databases are managed separately from development and test databases, but not consistently 	<ul style="list-style-type: none"> Data storage and operations standards, policies and SLAs are introduced for business critical data and are being adopted within new projects. DBA roles and responsibilities are defined and in place for all business critical data stores. All production databases are managed separately from development and test databases. Monitoring of data in storage routinely takes place, with issues addressed reactively based on business impact. 	<ul style="list-style-type: none"> Database technologies are aligned with enterprise-wide technology architecture. Standards and policies are adhered to across the enterprise. DBA roles are actively demonstrating and promoting the value of data storage and operations management. 	<ul style="list-style-type: none"> All data is managed throughout its lifecycle in a consistent and appropriate way from acquisition to disposal. Automated processes are in place for defined data storage and operations processes. The value of data operations management to the enterprise is recognised. Use of abstracted, reusable data objects and virtualised data is routine, increasing agility and shortening development time.

Maturity levels

Capability	Insufficient	Sufficient	Differentiating		
Data Security	<ul style="list-style-type: none"> Security policies are not clearly defined or communicated. No standard approach is used for marking documents that contain sensitive data. No enterprise wide approach or processes to control access to sensitive data 	<ul style="list-style-type: none"> Data security policies and requirements are outlined and communicated. User authentication and access control is in place, e.g. through use of strong passwords. Protective marking schemes are in place for documents and data assets. 	<ul style="list-style-type: none"> Requirements for data security and data protection are well understood and documented. Access to sensitive data is controlled based on user authentication and not granted to non-essential users. Audits of data security are conducted on a regular basis 	<ul style="list-style-type: none"> Data Security policy has been published, derived from data security requirements and risk assessments. All personnel appropriately aware and trained in Data Security. Advanced authentication and access controls are in place for sensitive data assets. Unauthorised attempts to access data are recorded and monitored. 	<ul style="list-style-type: none"> Mature processes exist to discover inappropriate access to or usage of enterprise data and to ensure that security measures provide adequate protection. Ownership, Accountability and Responsibilities for data security clearly defined and understood. Clear management buy-in for data security.
Data Integration & Interoperability	<ul style="list-style-type: none"> Data is extracted from operational systems, or the DW if it exists, and integrated offline in Excel or Access. No common vocabulary or reference data exists; each source is independently mapped. 	<ul style="list-style-type: none"> Data integration effected in silos driven by individual projects / initiatives. Limited standards for integration and interoperability exist in pockets across the organisation; no enterprise-wide standards. 	<ul style="list-style-type: none"> Enterprise-wide standards and best practice are being adopted and applied within new projects; remedial integration projects are being planned and funded where priorities exist. Operational systems are increasingly integrated. 	<ul style="list-style-type: none"> All core data repositories are integrated using enterprise wide standards and tools. Enterprise-wide data integration standards are delivering trusted, complete, accurate and timely data for all those who need to consume it. 	<ul style="list-style-type: none"> Data from disparate sources across the enterprise is available at the point of need in a fit-for-purpose format. New data repositories are integrated quickly and efficiently following strategic patterns and standards approved and enforced through strong data governance and enterprise architecture capabilities.
Documents & Content	<ul style="list-style-type: none"> Documents and content are stored on network drives, local PCs and in physical cabinets within departments. Users rely on folder structures, document names and operating system / email client search functionality. Records management policies are defined at departmental level, and may only be applied to specific types of document. The creator of a document determines access and usage. 	<ul style="list-style-type: none"> Content management systems are used within some business functions. Other functions continue to use generic storage capabilities. Documents and content are shared / made available between department by specific request. Many business processes rely on document exchange by email. Document collaboration, version control, template maintenance & standardisation are challenging. Access controls are local and limited. 	<ul style="list-style-type: none"> A common document and content management architecture has been implemented within the enterprise. Document access and usage is determined through data governance. Document and content meta data standards are being applied. Those documents which have been indexed and classified can be searched and located quickly and efficiently using an enterprise search engine. Standard document templates have been defined and are used consistently across the organisation. 	<ul style="list-style-type: none"> Applications and workflow tools are integrated with content management. Document classification tools utilising content analysis and character recognition automatically categorise and index documents with rich meta data. Enterprise-wide standards exist for document disposal and retention under the stewardship of the appropriate governance and compliance teams. Electronic records are automatically archived and disposed. 	<ul style="list-style-type: none"> Electronic and physical documents and content are managed throughout their lifecycle; they are identified, classified, stored, retrieved, tracked, permanently preserved or destroyed in a manner compliant with regulation and legislation. All data in unstructured sources (electronic files and physical records) across the enterprise are available for integration and interoperability with structured (database) data. Structured and unstructured data are routinely and seamlessly mined for value and insight.
Reference & Master Data	<ul style="list-style-type: none"> Limited awareness of MDM. Master Data domains have not been defined across the enterprise. Silo based approach to data models means multiple definitions of potential master data entities, such as customer, exist. 	<ul style="list-style-type: none"> The impact of master data issues gain recognition within the enterprise. Limited scope for managing master data due to lack of Data Ownership Model. Project or department based initiatives attempt to understand the enterprise's master data. No MDM strategy defined. 	<ul style="list-style-type: none"> Definition of an MDM strategy is in progress. Master data domains have been identified. Several domains are targeted for delivering master data to specific applications or projects. Differing products may be adopted in these silos for MDM. Senior management support for MDM grows. 	<ul style="list-style-type: none"> A complete MDM strategy has been defined and adopted. MDM joined up with data governance and data quality initiatives. Robust business rules defined for master data domains. Data cleansing and standardisation performed in the MDM hub. Specific products adopted for MDM. Master data models defined. 	<ul style="list-style-type: none"> A full integrated MDM hub exists and has been adopted across the enterprise for all key master data domains. The hub controls access to master data entities. Many applications access the MDM Hub through a service layer. Business users are fully responsible for master data.

Maturity levels

Capability	Insufficient	Sufficient	Differentiating		
Data Warehousing & Business Intelligence	<ul style="list-style-type: none"> No BI, or reporting directly from operational systems. Limited flexibility of reports results in teams or departments extracting data and manipulating in Excel or Access to derive the information they require. Reports in no way linked to the business critical success factors. 	<ul style="list-style-type: none"> A number of data marts exist to service individual department or project needs. Data extracted from multiple operational systems to feed the data marts. Limited standards in place. BI tools and dashboards adopted but Excel and Access use still prevalent. 	<ul style="list-style-type: none"> Many operational systems feeding into data marts. Overall approach becomes centred around a data warehouse. Standards and best practice begin to be adopted. Strategy developing around the enterprise's information requirements. Data held in spreadsheets consolidated. 	<ul style="list-style-type: none"> The availability of an Enterprise Data Model aligned with the business sees the evolution of the Data Warehouse into an enterprise wide initiative. Critical success factors for the business are aligned with performance indicators and key performance indicators. 	<ul style="list-style-type: none"> With a stable EDW, providing consistent, quality data the enterprise can make best use of predictive analytics and reintroduce flexibility and agility based on defined processes and standards. Data mining and advanced analytics tools adopted. Data is fully exploited to support rapidly changing business models.
Meta-data	<ul style="list-style-type: none"> Data repository(s) does not exist. Some information may have been collected by specific projects and initiatives in the form of Excel spreadsheets / Word documents. 	<ul style="list-style-type: none"> Disparate set of data repositories exist as a result of specific projects & initiatives. Little or no synchronization / communication across these tools. 	<ul style="list-style-type: none"> Multiple data repositories that synchronize and/or communicate via bespoke interfaces. 	<ul style="list-style-type: none"> A single integrated data repository houses the "record of reference" (single version of the truth). Other systems access the "record of reference" from the central integrated repository. 	<ul style="list-style-type: none"> Central data repository is optimized via standard data collection & distribution mechanisms. Data accessible to other programs, projects and users in secure manner.
Data Quality	<ul style="list-style-type: none"> Limited awareness within the enterprise of the importance of information quality. Very few, if any, processes in place to measure quality of information. Data is often not trusted by business users. 	<ul style="list-style-type: none"> The quality of few data sources is measured in an ad hoc manner. A number of different tools used to measure quality. The activity is driven by a projects or departments. Limited understanding of good versus bad quality. Identified issues are not consistently managed. 	<ul style="list-style-type: none"> Quality measures have been defined for some key data sources. Specific tools adopted to measure quality with some standards in place. The processes for measuring quality are applied at consistent intervals. Data issues are addressed where critical. 	<ul style="list-style-type: none"> Data quality is measured for all key data sources on a regular basis. Quality metrics information is published via dashboards etc. Active management of data issues through the data ownership model ensures issues are often resolved. Quality considerations baked into the SDLC. 	<ul style="list-style-type: none"> The measurement of data quality is embedded in many business processes across the enterprise. Data quality issues addressed through the data ownership model. Data quality issues fed back to be fixed at source.

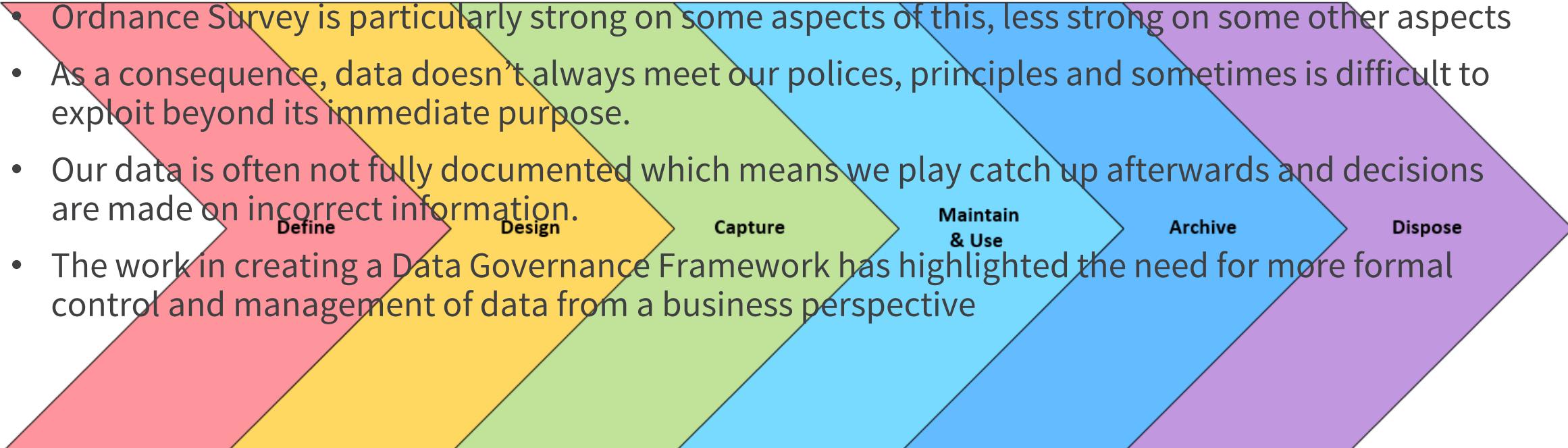
What happened afterwards?

- Appointment of new CDO
- Creation of a new team “the Data Office”
- Started to look at a Data Governance Framework

Background

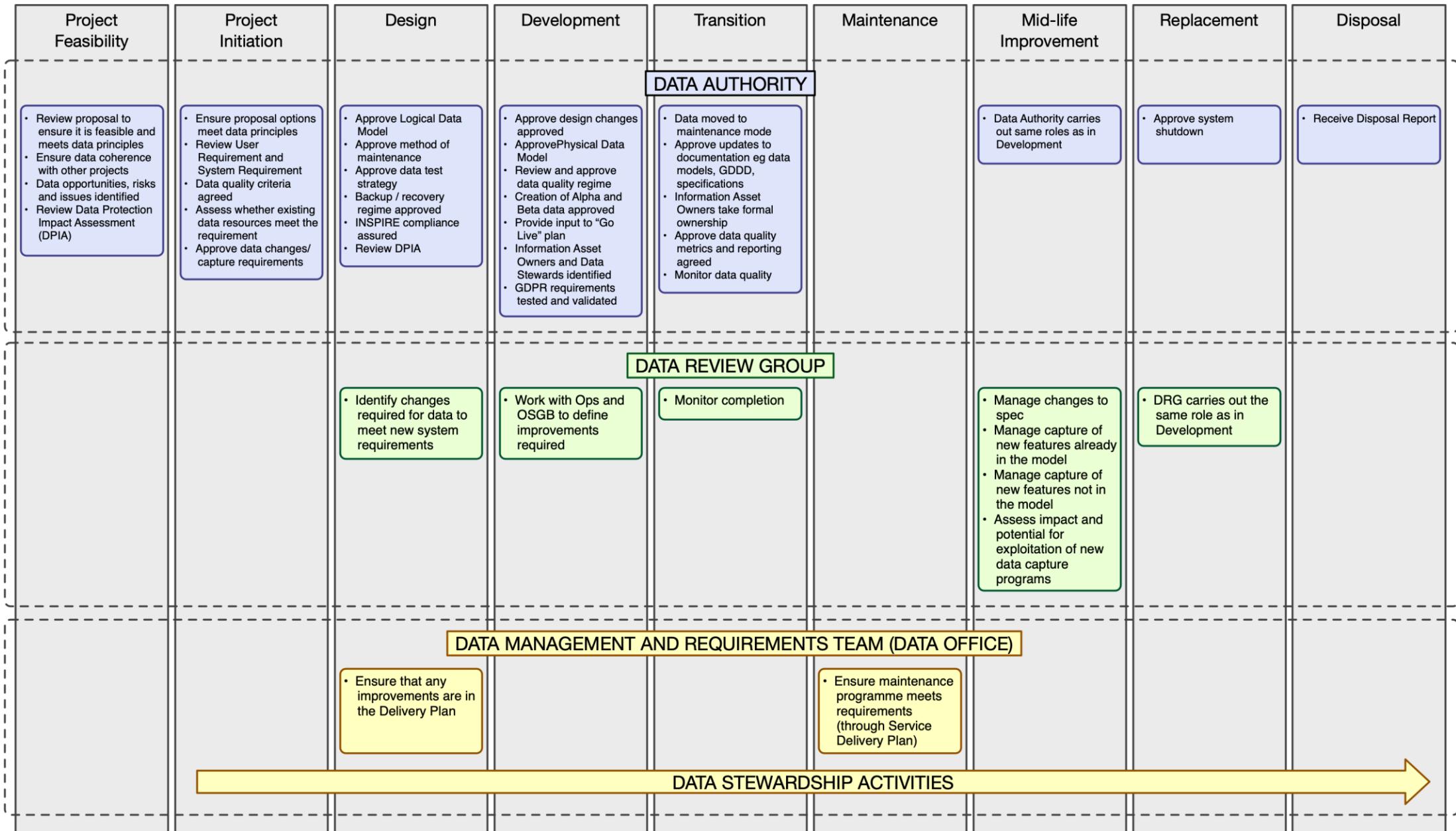
Data Lifecycle

- All data goes through the data lifecycle.
- Ordnance Survey is particularly strong on some aspects of this, less strong on some other aspects
- As a consequence, data doesn't always meet our policies, principles and sometimes is difficult to exploit beyond its immediate purpose.
- Our data is often not fully documented which means we play catch up afterwards and decisions are made on incorrect information.
- The work in creating a Data Governance Framework has highlighted the need for more formal control and management of data from a business perspective



Proposal

- That we create a Data Authority (DA) to provide governance oversight over the development of data.
- Its role would be
 - To govern the data wherever new data is being created or existing data is being changed.
 - To ensure all relevant data policies are adhered to during development.
 - To approve any exceptions to policy.
 - To approve all stages of project (from a data perspective).
 - To ensure that on project “Go-live” the data is fully transitioned to the next stage of the data lifecycle.
 - To ensure that any data related issues are corrected.
 - To advise CDO on the strategic fit of the project (wrt data) before it goes to Commissioning Authority.
 - To keep CDO informed of any major issues that arise during development and the remedial action being taken.
- The DA would do a similar job for data to the one that ARB does for the architecture of systems



Current Situation

- Data Authority about to be established
- Creation of an on line Glossary
- Creation of document outlining all our policies
- Next step – to get underway and hold our first meeting.