Data Strategy Template

***This template is an amalgam of many different projects, and is therefore appropriately anonymised. However, it is for internal use only and not to be copied, redistributed or repurposed without permission.***

# Contents

[Contents 1](#_Toc482673578)

[Our vision for data 3](#_Toc482673579)

[Context 3](#_Toc482673580)

[Details 3](#_Toc482673581)

[Context 3](#_Toc482673582)

[Section 1: Defining the business context 5](#_Toc482673583)

[Section 2: Defining business success 6](#_Toc482673584)

[Section 3: Defining business capabilities 7](#_Toc482673585)

[Delivering data governance 7](#_Toc482673586)

[Defining data architecture 7](#_Toc482673587)

[Implementing data management 8](#_Toc482673588)

[Implementing business intelligence (BI) 8](#_Toc482673589)

[Section 4: Data principles 9](#_Toc482673590)

[Draft Data Principles 9](#_Toc482673591)

[Information is a valued asset 9](#_Toc482673592)

[Information is managed 9](#_Toc482673593)

[Information is fit for purpose 9](#_Toc482673594)

[Information is standardised and linkable 10](#_Toc482673595)

[Information is reused 10](#_Toc482673596)

[Information is accessible 11](#_Toc482673597)

[Shared sector information is published 11](#_Toc482673598)

[Section 5: Delivering data governance 12](#_Toc482673599)

[Section 6: Defining data architecture 14](#_Toc482673600)

[Corporate data model 14](#_Toc482673601)

[Conceptual data model 14](#_Toc482673602)

[Logical data model 14](#_Toc482673603)

[Physical data model 14](#_Toc482673604)

[Business glossary/data dictionary 15](#_Toc482673605)

[Architectural Governance 15](#_Toc482673606)

[Section 7: Implementing data management 16](#_Toc482673607)

[Information Lifecycle Management (ILM) 16](#_Toc482673608)

[Master Data Management (MDM) 16](#_Toc482673609)

[Data quality 16](#_Toc482673610)

[Section 8: Implementing business intelligence 17](#_Toc482673611)

[Section 9: Organisation and culture 18](#_Toc482673612)

[Organisation 18](#_Toc482673613)

[Data governance and control 18](#_Toc482673614)

[Data management 18](#_Toc482673615)

[Culture 18](#_Toc482673616)

# Our vision for data

## Context

[To include]

* what the institution does
* why data is a vital to the success of this
  + What is collected
  + What is processed
  + What is presented
  + Etc.
* How data must be managed as an asset
  + Value
  + Governance
  + Outputs
  + Opportunities

[Much of this content is generated in the toolkit process]

## Details

[Sample]

Basics – doing the things right. Governance/best practice for the operational data. Regulatory and legal obligations. Support for efficiencies and better decision making

Culture – Doing the right things. Exploiting the opportunities of doing the things right. Institutional wide view of value/power data and education of responsibilities. How to drive culture to see services as data driven activities and look to build on owned data with external sources.

Opportunity – How the vision will support the aspirations/public strategy of the institution. Why it’s important to link the two. What might that look like.

## Context

[Sample]

“This Data Strategy provides the foundation for achieving our vision for data. It defines the relationships between data and the business context in which we operate, the outcomes we aim to achieve from successful implementation of the strategy, and the capabilities and culture we need to develop to realise these outcomes”

It will provide this through:

* Delivering data governance
* Defining data architecture
* Implementing data management
* Implementing business intelligence

“This strategy will inform the development and implementation strategy programme and projects and investment decisions including IT infrastructure. It will inform the development of new products and services and operational planning, decision-making and delivery across the business. More specifically, the strategy provides the foundation for the development and implementation of a range of underpinning data and information policies and procedures.”

# Section 1: Defining the business context

This Data Strategy should be an enabler for the corporate Strategy.

[List Strategy/Aspirations here –use objectives tab to help]

[Now list how a data strategy would support this. Begin with something like this]

An effective Data Strategy helps to exploit new opportunities to generate income, minimise risks from competitors, and ultimately thrive in different and diversifying education and HE markets.

[Add detail. Objectives/benefits links work well here]

# Section 2: Defining business success

[What does success look like, sample text could include]

“To achieve this aim we govern and manage our data using a defined set of policies and procedures based on pragmatic standards and best practice. We develop a consolidated data model and use proven data architecture and external data standards. We manage our data effectively in integrated data repositories collecting data items only once, maintaining them in only one place, and making them available for use in multiple applications across the business and beyond – providing all customers with “a single version of the truth”.

Successful implementation of the Data Strategy enables us to deliver benefits to our customers and the business by:”

[List of benefits as defined in the process. Always around either ‘reducing the cost of processing data’ or ‘increasing the value of data we hold/link’]

[Summary could include]

“The realisation of the benefits in the Data Strategy is dependent on the effective alignment of data governance and management with technology and IT infrastructure development, coupled with a clear articulation of business needs.”

# Section 3: Defining business capabilities

[Sample]

We realise our vision for data through a four layer model:

## Delivering data governance

We implement a standardised approach to data governance, based on external standards and best practice, which is integrated into the wider institutional framework.

This is based on a set of data principles aligned to our corporate objectives.

Adopting a standardised approach to data governance provides a defined accountability framework, policies and procedures. It creates the functions of Senior Information Risk Owner, Head of Data Governance, and Data Steward and places responsibility for the ownership of data assets with the part of the business best placed to take decisions about the collection, management, use and disposal of the data asset, and to identify and manage the risks associated with each data asset.

To deliver this new data governance framework we develop and grow data governance and data management knowledge and skills within the business.

## Defining data architecture

We implement a consistent view of the organisations data landscape by developing and implementing a consolidated data model using industry standard approaches. We establish a common vocabulary and set of data definitions and standards (aligned to external standards) to create a clearly understood “single version of the truth” with regards our data assets.

We use a standard architectural approach in line with IT strategy and/or roadmaps, including standards, principles and models.

## Implementing data management

We implement an exemplary approach to data management, given the organisational imperative to excel in this area. The three core capabilities developed and implemented are:

* Information Lifecycle Management (ILM)
* Master Data Management (MDM)
* Measuring and improving data quality

These capabilities support better data integration, data management and quality and will be realised through the introduction of policies, processes and automated tools. Performance is measured by a set of metrics and KPIs to determine the quality of data assets.

## Implementing business intelligence (BI)

[This may not be relevant for every intuition but it’s important to consider at a distinct theme not part of standard data management activity – lots of specialist skills and a reasonable level of data maturity required. It is also absolutely tied in with visible data governance activity]

We can introduce a BI environment which gives the institution the capability to use its data for operational monitoring and reporting, impact assessment and better evaluation of change requests, whilst also providing opportunities to offer our customers and stakeholders better data products and services around decision support.

A BI environment can provide near real-time and historical data underpinned by a range of BI and analytical tools. This is based on industry best practice and can be supported by the establishment of a virtual BI competency centre in the business.

# Section 4: Data principles

[Samples taken/modified from the Cabinet Office Data Standards work. These would need to be cut down but are included for completeness]

## Draft Data Principles

* Original source ‘Government Design Principles’ for 2012 – published by the Cabinet Office
* Note ‘Information’ is used rather than ‘Data’ to shape the principle around the value of the data
* This is a generic set of principles, suggested context and focus for your organisation are highlighted in [ ]
* Items in *italics* are those I consider to be the most important part of the principle.

## Information is a valued asset

* Information should be understood and valued as much – and sometimes more - *as other organisational assets such as buildings, people or money*. [Add your own most valuable balance sheet assets here if they are different]
* This principle is the foundation for what follows and highlights the need for information to be valued in the same way as these other types of asset. It is important to note that the full value of information lies not just in its original purpose but *in its potential to be reused for other purposes.*

## Information is managed

* Information should be managed – stored, protected and exploited – *according to its value.*
* Information owners/stewards [these may not be how these roles are named in your organisation] need to consider the whole lifecycle of the information, from identification of need, creation, quality assurance, maintenance, reuse and ultimately to archiving or destruction once the *information has ceased to be useful*. [consider this in the context of architectural debt]
* A range of best practices need to be in place, for example to ensure appropriate availability and integrity, avoid loss and *ensure continuity across technology upgrades*. It is particularly important that personal data is adequately protected. [Focus on your primary datasets / transaction data]
* Information also needs to be governed as it moves through its lifecycle, for example to make *sure it’s always clear who is responsible for it* [this is the heart of information asset ownership. Ensure your organisation has a plan to implement this based on your current or new structure] and to comply with relevant legislation and regulation. The consistent *assessment and ownership* of these information risks is another important consideration when managing data and information. [This is where you should highlight your data governance function]
* The organisational culture must s*upport best practice in information management*, and make sure everyone responsible for processing these business assets is professionally qualified and appropriately skilled. This principle therefore also includes the *processes, roles, responsibilities, training, and organisational structure and culture* needed to ensure the effective and efficient use of information. [These are the four dimensions of the assessment and are all equally important. Consider how your organisation will make progress on all these front to support this principle]

## Information is fit for purpose

* Information must be good quality and fit for both its primary purpose and *potential secondary/tertiary uses*. It will not always *be possible for the owner/originator to foresee these addition uses*, so it is important that the quality of the information is communicated consistently so future users can decide if it is suitable. [This is the value of data architecture/shared models/shared business vocab/reference data/taxonomy/etc]
* Quality includes factors such as accuracy, validity, reliability, timeliness, relevance and completeness [Consider what is most important for the operational datasets, don’t try and create quality factors that have no value]. The quality of information *should also be regularly monitored to ensure that they at least meet the levels that have been assessed as necessary*
* *for their purposes* [do not set Quality metrics you are not able/willing to monitor AND intervene on breach]
* This principle *doesn’t require information to be perfect,* only that it is the right quality for its intended use and that its quality characteristics are pragmatic, appropriate and transparent

## Information is standardised and linkable

* Standardisation is important for structured information such as dataset definitions, and unstructured information such as Meta and reference data applied to documents. [There will be many more opportunities for exploiting information if it is available in standardised and linkable forms.]
* Standardisation within an organisation is important for staff to fully exploit the information; if an organisation uses widely accepted open standards [the HEDIIP Data Landscape will form part of these open standards but also consider industry/sector wide initiatives] it will unlock even more value
* Standardisation is important both for the way information is recorded and in the way concepts are defined:
  + Format, eg date always being entered as dd-mm-yyyy
  + Content, eg forename, surname, address, etc.
  + Concepts, eg defining roles such as student, course, module, certificate, etc
* [Even further value can be unlocked if information can be linked. A good example is document references and citations that allow the reader to draw on a wealth of associated information (this is the basis of the ‘world wide web’). A similar concept can be applied to structured data, based on an understanding of the relationships between items and the use of consistent identifiers to reference authoritative sources (the basis of the ‘semantic web’). For example, tagging funding information with an authoritative code for the organisation involved would allow it to be unambiguously linked with details of the organisation itself and third-party information about that organisation]

## Information is reused

* Information is even more valuable *if it can be used more than once or for more than one purpose*. A good data steward will proactively look for opportunities for reuse. These could include:
  + Internal reuse – making the most of information for its primary purpose and identifying secondary uses. For example, operational data can sometimes be reused to support performance improvement or research [student data with absolutely everything else. Student data is the ‘operational hub’ but has many forms of reuse. This principle MUST be applied early on. If not a lack of reconciliation is the outcome]
  + External reuse – sharing information with other organisations, either within the sector or with private businesses and individuals [Collections, Student Experience, etc. are obvious candidates here]
  + Holding master reference data – ensuring an organisation’s data is the only authoritative source for business information (eg an authoritative list of organisation codes), which is nominated, maintained and promoted as such. [This is difficult across the sector but will in some way be addressed by the landscape. JACS was a classic example held by UCAS but this eventually became unsustainable. HESA codes are another]
* Whilst this principle strongly encourages reuse, it is important to appreciate that reuse does require a *careful risk-based judgment* to be made with regard to exploiting versus protecting information, as well as consideration to the costs and benefits involved [some reuse scenarios do not make technical or economic sense, but the principle should be applied and only discarded if this is clearly the case]
* Information which initially *appears unsuitable may be reusable if it can be reformatted*. For example, operational information that identifies individuals can be ‘anonymised’ or aggregated and then be of wider value. Also, in cases where the partner organisation is known beforehand, concerns over security or privacy can sometimes be mitigated by means of negotiation, joint-working and data-sharing agreements [this can be applied to internal data sharing agreements as well]

## Information is accessible

* Individuals and organisations must be able to access information about themselves, along with an *explanation of how that information is used by others*. This may be either on request or, preferably, by making it available by default. In effect, such information should be considered as belonging to the individual, although entrusted to the care of a public body [This should match the FOI/DP policies of the organisation, however it is more than just compliance, see below]
* This principle goes beyond minimum legal requirements. It advocates a proactive approach which makes it easy for individuals to access information about themselves, without having to make a request and even when access is not mandated in legislation. This might be achieved, for example, by making it securely available online. Information owners need to consider how this can work in practice, to enable users to view information and perform transactions, *for example correcting inaccuracies*. [Much of the data collected is self certified and that may be the quality metric. Making information accessible has mutual benefits for the individual and the organisation]
* Clearly the desire to publish information does need to be balanced against constraints which may prevent this. Exclusions would include, for example, legally privileged information, information that is required to maintain security and information that is competitively sensitive.

## Shared sector information is published

* This principle goes beyond adhering to minimum legal / regulatory requirements and *advocates a proactive approach to presenting, formatting and promoting information in useful formats for wider consumption*, without it needing to be specifically requested or mandated in legislation. [This principle is in line with the HEDIIP new landscape/shared sector language]
* Consider the different channels available to publish information to the public. This includes internal publication processes, the use of publication hubs and relationships with third-party ‘information intermediaries’ such as commercial / academic publishers.
* The benefits of publishing information should be balanced against possible risks and sensitivities, such as information which could compromise individuals’ privacy, commercial and legally privileged information, and information that is required to maintain security.

# Section 5: Delivering data governance

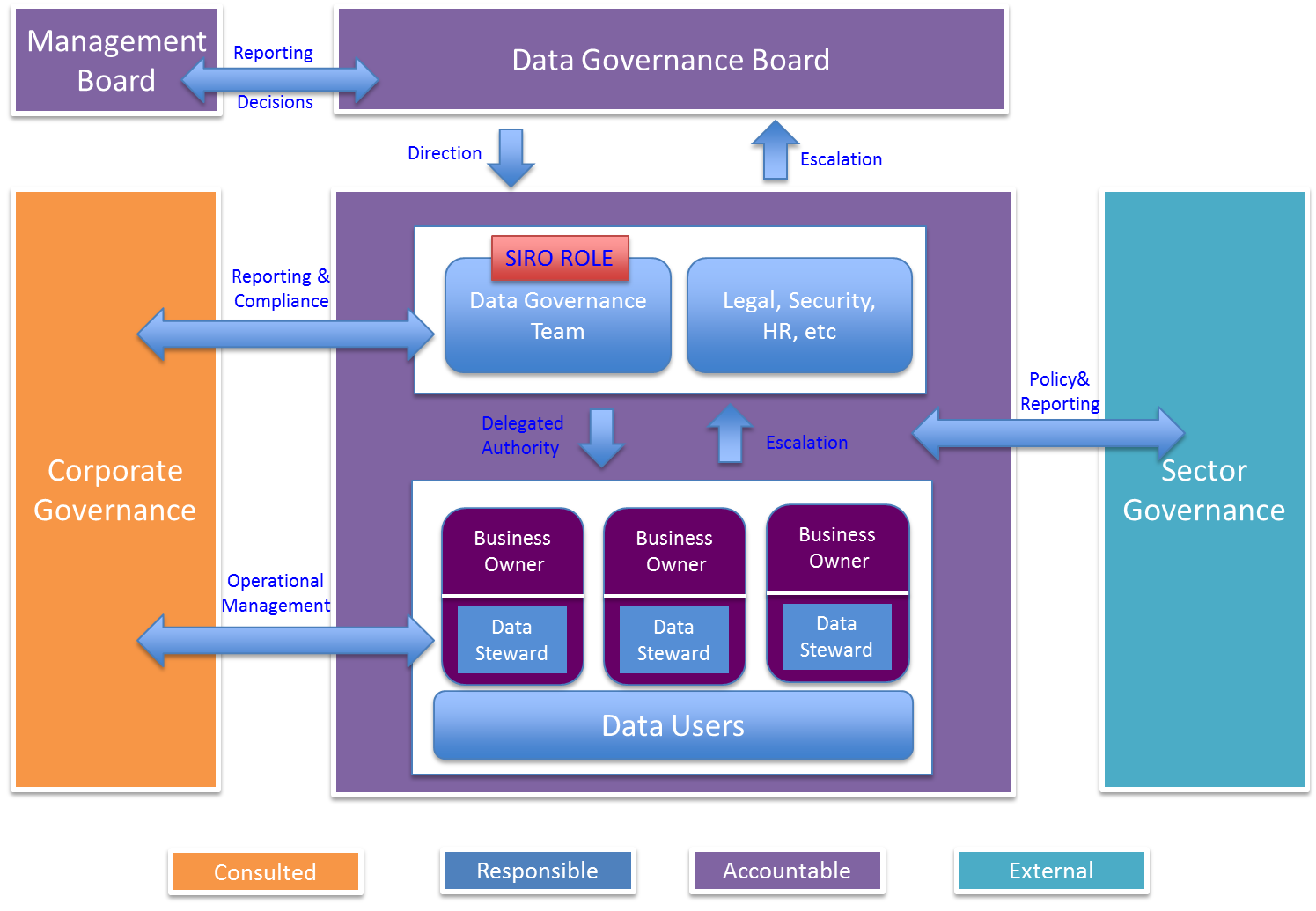
[Sample]

Data governance provides the policies, procedures, functions, and the roles and responsibilities that are needed to support data architecture, data management and business intelligence in the institution.

Specifically this includes:

* ensure data is secure and well managed, and risks to the organisation are contained
* prevent and correct errors in data
* measure data quality and provide a framework for data quality improvement that is monitored and assessed
* define the standards to document data and its usage within institution

[Sample governance model]



The main governance body is the Data Governing Board (including any Information Security Forum). It is responsible for direction and decision-making in support of data governance.

On a day-to-day basis the Data Governance Office is responsible for implementation of data governance and data assurance.

The main governance functions include:

* Executive accountable officer for data – includes the role of senior information risk owner and is ultimately accountable for the security and quality of data
* Information asset owner – accountable and responsible for data governance and data management for information within each business unit, ensuring all policies and procedures are implemented and data quality initiatives are implemented
* Data steward – responsible for the application of data governance, data management and data quality plans on behalf of information assets owners within each business unit

# Section 6: Defining data architecture

[Sample]

The data architecture is a key enabler in implementing data management and business intelligence.

The data architecture is based on an enterprise model including:

* enterprise data architecture model
* corporate data model consisting of:
  + conceptual data model
  + logical data model
  + physical data model
* business glossary/data dictionary
* data flow model
* master data/metadata definitions

[Note some of these functions may be based in other areas, especially the reference data/physical layer models. Virtual Data Architecture groups are a possibility but need managing with care]

## Corporate data model

The corporate data model is developed and maintained pragmatically, as attempting to create a complete model in one go is an onerous and complex process.

## Conceptual data model

The conceptual data model is an unambiguous view of the organisation that provides a high level technology independent view of the key institutional data assets and represents a complete view of the data landscape.

## Logical data model

The logical data model provides a more detailed technology independent view of the data landscape, and is maintained by the business to capture the data requirements for business processes to gain a full understanding of the organisation’s data. The logical model is constantly evolving due to new and changing requirements and is developed on an ‘as needed’ basis to prioritise requirements and resources to meet business objectives. A detailed and consistent data model is the basis for the physical model.

## Physical data model

The physical data model provides a detailed and unambiguous definition of how data is stored within the institutions systems. Similar to the logical model, physical models are developed on an ‘as needed’ basis to deliver the business objectives, and are used to generate the data definitions needed to create and maintain the physical data repositories.

The enterprise model is documented using recognised industry standards and using a set of appropriate modelling tools.

## Business glossary/data dictionary

Deploying a common set of data definitions across the organisation is a critical component in ensuring data quality, maintaining the integrity of data, reducing the cost of cleaning data and reducing the complexity of data integration by keeping the duplication of data to a minimum.

|  |
| --- |
| The benefits of detailing the data architecture are:   1. It provides an unambiguous view of the data in the institution that can be used to determine the impact on the organisation resulting from changes to the data. It can be used to define cost, effort and business benefit to ensure changes are affordable and architecturally sound 2. It provides a view of data that can be used to determine business risk in terms of:    * impact on service due to loss of parts of the data infrastructure    * impact on information lifecycle management (ILM) from data change, with particular reference to interfaces (both internal and external) between systems    * risk resulting from data quality issues    * risk from external threats, i.e. security 3. It aids interoperability with our key suppliers and customers of data |

## Architectural Governance

Data changes can have organisation wide impacts and therefore are governed by the wider architecture group; all data changes will be aligned to the enterprise model.

The following key principles are adhered to in developing and maintaining this model:

* data is documented in the enterprise model
* data conforms to the institution ILM process
* data duplication is kept to a minimum
* all data has a responsible data owner who is responsible for quality and conformance of data

All data in transit (all internal and external interfaces) is subject to data governance and data quality policies.

# Section 7: Implementing data management

[Sample]

The key areas included in data management are:

* an Information Lifecycle Management (ILM) process
* data quality (DQ)
* master data management (MDM)

Data management is supported by the data architecture and data governance.

## Information Lifecycle Management (ILM)

The ILM capability covers the policies and procedures that support data quality, data logistics and data integration covering the following five stages:

* creation or acquisition of data
* distribution
* usage
* maintenance
* retention, archiving, disposal

## Master Data Management (MDM)

The MDM capability provides the policies, procedures and solutions to manage and maintain master data within the ILM, and ensure that consistent versions of the same data are used throughout the institution. The processes within MDM include:

* capture and de-duplication of data
* conflict resolution
* distribution of master data.

## Data quality

A capability is provided that monitors and measures the quality of data as it is captured and transformed by the institutions processes and is supported by automated toolsets where appropriate.

The key capabilities provided by data quality management are:

* measuring data quality
* maintaining data integrity
* data cleansing
* minimising data duplication

The core capabilities are delivered by a small team of people who also have responsibility to manage and administer the enterprise data architecture and corporate data models.

# Section 8: Implementing business intelligence

[Sample]

The deployment of a BI environment provides the institution with a consistent, trustworthy and appropriate capability to support the needs of its internal and external customers.

Business intelligence is supported by data governance, data architecture and data management and is a key contributor to data quality in its own right by delivering the ‘single version of the truth’.

[Need to add details of any BI Strategy here]

# Section 9: Organisation and culture

[Sample]

## Organisation

Two core organisational functions deliver the Data Strategy:

* Data governance and control
* Data management

## Data governance and control

Data governance comprises the data governance “organisation”, plus policies, procedures, processes and standards. [List all parties and reference Data Governance Charter]

Compliance with data governance is a corporate as well as individual responsibility and compliance is monitored and evaluated in the first instance through the Data Governing Board and in turn through the corporate governance framework.

## Data management

To ensure that the institution is successful in delivering timely, robust, consistent, high-quality data it has the organisational capability to develop and maintain data assets in line with governance policies, procedures and standards.

Responsibility for the ownership and management of individual data assets is delegated to the appropriate business function, who are supported by data stewards. In addition, the data management team supports data quality through data management and master data management capabilities.

## Culture

Our data assets are at the heart of the business. We aim to unlock the power of the data assets by engendering more understanding within the business about how data can be employed to improve operational performance, evaluate options and make better sustainable decisions.

It is essential that everyone who works for the institution understands their responsibilities for maintaining the integrity and quality of our data assets, complying with data legislation and regulations, and keeping the data assets safe and secure.

Training covering data protection, information security and Freedom of Information forms a mandatory part of the induction programme for all institution staff and contractors and refresher training and communications are undertaken regularly.