Information and Data Management Principles

UPR IM02 version 05.0

Policies superseded by this document

This document replaces version 04.0 of UPR IM02, with effect from 3 January 2023.

Summary of significant changes to the previous version

This document has been amended substantially and should be read in full.

Glossary

A glossary of approved University terminology can be found in **UPR GV08**.

Table of contents

1 Introduction and Scope

1.1 These Information and Data Management Principles apply to information in all its forms and provide a corporate framework for all the University's information-related activities. Within this overall framework there may be a number of subsidiary policies and procedures.

1.2 The Principles apply to all University of Hertfordshire activities, to individual Members of the University as defined in UPR GV06[1], to the University's wholly-owned subsidiaries and their wholly-owned subsidiaries to companies in which the University has an interest subject to necessary consultation and agreement with the respective Board of Directors, and to collaborative activities undertaken with Partner Organisations.

1.3 In an environment of increased market competition, and increased regulation and risk introduced by the Office for Students and the General Data Protection Regulation 2018 (GDPR), the University has a need for a clear vision for the management and utilisation of data. Data and information are used interchangeably throughout this strategy, but have distinct definitions as defined in UPRs IM11 'Records Management' and IM16 'Data Management', and in the glossary below (see section 3).

[1] UPR GV06 'Membership of the University'

1.4 This document sets out ten key principles for the processing of data. These principles have been developed by the Data Management Group, in consultation with the Data Protection Officer, University Records Manager, Library and Computing Services, Internal Audit and business stakeholders.

1.5 The Information and Data Management Principles are based on industry best practice and reflect the University's strategic objectives, which are implemented through policies, processes, and standards. The Principles can be used to help benchmark project and process change, and represent the high standard toward which all Members[1] should aspire.

1.6 This document clearly defines the strategic principles, outlining the following for each:

- Statement - summary of the principle;
- Rationale - business benefits of adhering to the principle;
- Implications - the requirements, both for the business and IT, in applying the Principles.

2 Data and Information Principles

2.1 Data is treated as a valuable University asset

2.1.1 Statement

Data is an asset that has value to the institution and is managed accordingly.

2.1.2 Rationale

Information is a valuable resource; it has a cost to collect, maintain and dispose. Information also has a value in its funding and use. Other organisational assets such as buildings, staff and finances are carefully managed, and information assets should be no exception. It would be impossible for the University to function without access to a variety of information and data. The efficient operation of the University is linked to the accuracy and availability of information. There are growing demands on the higher education sector to manage data properly and invest in it. Recent threats include the General Data Protection Regulations (GDPR), Office for Students (OfS), Data Futures and Graduate Outcomes. In a competitive market, good decision making is based on good data. Therefore, information should be treated as a valuable asset, with appropriate investment. Good data means meeting a pre-defined standard of data quality, which cannot exist without both investment and due care and attention.

2.1.3 Implications

i The University invests in information management.

ii Investment means having the proper strategy, policies, procedures and processes in place to manage data, as well as an investment of staff time and technology to support those processes.

iii The University ensures its tools make the most of information.

iv The University develops metrics to show the cost and value of information.

2.2 Data is everyone's responsibility

2.2.1 Statement

All Members[1] have a responsibility for the secure and appropriate management of data and should be involved in decisions and discussions to achieve business objectives.

2.2.2 Rationale

All Members[1] have a responsibility to ensure that data and information are aligned with business needs that underpin statutory and customer experience objectives.

2.2.3 Implications

i Staff take accountability for the data they collect, ensuring it is fit for purpose for their own business needs, and the needs of anyone else who makes use of that data.

ii Staff are trained to understand the implications and importance of data, and with the skills needed to manage that data appropriately.

iii Managers allocate staff sufficient time to meet their responsibilities with data.

iv Responsibilities and expectations around the management of data are written into job / role descriptions.

2.3 Data is owned at the authoritative system

2.3.1 Statement

All data has a defined authoritative system (see section 3.18) and a responsible Information Asset Owner (see section 3.26).

2.3.2 Rationale

Where data has a clearly identified system, this enables data quality to be managed efficiently, in a single location. An authoritative source for data ensures auditability and control over data flow and protection. This helps to ensure a single version of the truth. Ownership of systems and data ensures accountability for compliance with legislation and policy resides with the most appropriate role.

2.3.3 Implications

i Data is only ever amended at the authoritative system.

ii The University maintains a catalogue of authoritative systems and information assets.

iii The Information Asset Owner is accountable for the nominated data/ information assets, and for ensuring those assets are properly managed in line with policy.

iv The Information Asset Owner may delegate responsibility for the proper management of their data assets.

v The Information Asset Owner is ultimately responsible for determining the appropriate use of their data assets, and should give authority for data sharing.

vi Information Asset Owners are responsible for the appropriate security of data within their area, but also recognise that data is a University asset to be shared appropriately.

vii Information Asset Owners ensure that data is fit for purpose for use within their area, and fit to be shared with any subsequent area which requires that data.

viii Systems and processes facilitate the sign-off of access to data (and the systems which hold that data), by the Information Asset Owner or nominated responsible officer (not by the technical system administrators).

2.4 Data is processed, retained and shared appropriately

2.4.1 Statement

Management of data conforms to legislation, the University Policies and Regulations (UPRs) and any external requirements.

2.4.2 Rationale

With Office for Students regulation, significantly increased data protection fines from the Information Commissioners Office (ICO) (see section 3.32), and the opportunities for compensation under new legislation, there is now a much greater risk of financial damage. There is also a growing public and media interest in the use of personal data which presents a greater risk of reputational damage. In a highly competitive and regulated market, data which is not processed in an efficient and effective manner brings with it the cost of business inefficiency. In this environment, the University needs to embed appropriate practice into all business activities.

2.4.3 Implications

i The University respects the rights of data subjects, and the legal basis for processing.

ii Personal information is processed only for the purpose stated when it was captured.

iii Privacy By Design (see section 3.40) is implemented as a key deliverable of any projects or change involving personal data, as part of the proper completion and approval of a Data Privacy Impact Assessment (see section 3.6).

iv Data minimisation (see section 3.22) is ensured for any integration or sharing of data.

v Systems, databases, records and information conform to the University Retention Schedule (Appendix I, UPR IM11[2]), and information is retained only as long as there is a valid legal or business justification.

vi The University Retention Schedule[2] is fit for purpose, up to date, properly resourced and embedded.

vii Non-personal business information is open and accessible by default, unless there is a good business justification to restrict access.

viii Data Sharing Agreements, Data Processing Contracts, or additional contract clauses are in place for all sharing of data outside of the University.

ix The sharing and use of data with suppliers, partners and other third parties is clear within the contract.

x Internal sharing between systems within the University is documented, approved and managed. Oversight and authorisation of the sharing of data internally is given by the Data Management Group.

xi The University ~~We~~ will continuously develop and improve ~~our~~ its policies in line with legislation, delivering a more efficient business and reducing information risks.

xii The University will make use of tools to classify and protect information appropriately.

2.5 Data is structured and interoperable

2.5.1 Statement

Data is structured and pre-defined (see section 3.3). Systems and processes promote interoperability (see section 3.35) and automation (see section 3.19) in data and information exchange.

2.5.2 Rationale

Structured data is essential for interoperability and business efficiency. Structured data also helps the institution to comply with its regulatory responsibilities. Where information is stored in unstructured formats, such as documents and emails, this makes it more difficult to find and automate, and increases administrative double handling and inefficiency. Structured, machine readable, information aids discovery and compliance with information requests. Unstructured information cannot be easily used for good data-based decision making, as used in analytics and visualisation. Where information is held in silos, disconnected from the authoritative source, it is likely to be inaccurate and out of date, leading to bad decision making. Good data and information management must be designed and accounted for as part of all projects, and points of change within the University, in order to be built into processes.

[2] Appendix I, UPR IM11 ‘Records Management and the Archiving and Retention of Prime Documents and Business Records

2.5.3 Implications

i Data is stored and presented in a digitally accessible, structured format.

ii Projects and business change give precedence in design to structured data over un/semi-structured (see section 3.5) data.

iii Design gives precedence to automation and integration, over manual keying.

iv Design gives precedence to digital by default i.e. creating information electronically, rather than creating paper and scanning / re-keying.

v Design makes use of industry best practice tools and methodologies for management of data, including federation (see section 3.29) and master data management.

vi Structured data must be portable i.e. it must be exportable in structured, commonly used, and machine-readable format.

vii Data must not be duplicated, held, or used in isolation, without regular refreshes from the authoritative system.

viii Data formats and processing follows international standards as appropriate.

ix Where unstructured information (e.g. paper) is processed, preference is given to automatically capturing the contents as structured data, but must as a minimum be digitised and stored with appropriate, structured metadata.

x Preference is given in the design or procurement of systems, to integration without the need to buy or develop middleware (i.e. interoperable) as long as that integration gives sufficient control and audit of the data flowing between systems.

2.6 **Change that affects data is managed**

2.6.1 Statement

All changes that affect or could potentially affect data are managed.

2.6.2 Rationale

Core University data is used throughout the organisation and underpins the majority of systems and processes. Changes to the format, structure or content of data, or to the business processes which underpin data, can significantly impact on the delivery of business activities unless this change is managed. Changes to data can prevent the proper operation of University systems creating an unforeseen but ongoing management cost.

2.6.3 Implications

i Where there is:

a a new University system,
b a planned upgrade or downtime on a University system,
c a new requirement for data sharing (internally or externally),

d a new data feed / integration,
e a change to an existing data feed / integration, or
f a change to formats, data structures or types where the data is shared between 2 or more systems or supplied in data returns, or a change in any authoritative system,

then this must be raised at the Data Management Group as early as possible in the project / development or, in the case of an urgent request, with Library and Computing Services.

ii Any data changes where there is no integration or data sharing, it is recommended but not essential to consult the Data Management Group.

iii Data management and change control processes will be continuously developed and must be followed.

2.7 Data is defined in context and consistent

2.7.1 Statement

Data is defined consistently across the institution, and in the contexts in which it is used.

2.7.2 Rationale

There are different interpretations of data within different contexts. It is recognised that a student within the context of a non-credit bearing short course, or accommodation, is different to a student within a HESA return. Likewise, there are different interpretations of staff, user, researcher and name. Clear definitions of data, within context, enable accurate and appropriate use, and reduce the risk of breaches.

2.7.3 Implications

i The organisation must establish a common vocabulary for describing information assets and associated metadata which will be facilitated by the Data Management Group.

ii Data is defined with descriptors and standards.

iii Data is defined within the specific contexts in which it is used.

iv Definitions and contexts are accessible to relevant Members[1].

v Standards and definitions are owned by the Information Asset Owner for the relevant data set.

2.8 Data is correct at the point of creation

2.8.1 Statement

Data is correct and checked at the earliest point of its lifecycle.

2.8.2 Rationale

Data quality issues can undermine efficient delivery of, or entirely halt, business activities. As data is used in more areas than it is collected, correcting data quality issues after the point of collection causes duplication of effort and inconsistency. Without a standardised management process, solutions may tackle the symptom, but not the underlying cause of quality issues.

2.8.3 Implications

i The University ensures information is accurate and fit for purpose.

ii Data conforms to data management standards at the point of creation.

iii Where data quality issues are identified, these are corrected at the point of collection / creation and pushed automatically to integrated systems.

iv Data quality issues identified outside of the authoritative system must be notified to the responsible Information Asset Owner as soon as possible.

v Data quality management conforms to a standard process, as defined by the Data Management Group.

vi Quality is defined according to data quality dimensions (see section 3.8) appropriate to the data.

vii Data validation is implemented on systems which collect data, and the University procures systems which enable Members[1] to undertake data validation.

2.9 **Data quality is measured and continuously improved**

2.9.1 Statement

The management of data, information and the underlying systems are sustainable.

2.9.2 Rationale

The Office for Students, HESA and the ICO demand a high level of data quality. More generally, data quality issues can undermine efficient delivery of, or entirely halt, business activities. Without ongoing measurement and improvement of data quality, the University cannot comply with regulatory expectations, or run its business efficiently.

2.9.3 Implications

i Robust and innovative analysis is conducted on all data.

ii Data quality issues are reported to the appropriate Information Asset Owner or delegated authority and reviewed.

iii Data quality issues are categorised and assigned a level of risk / impact.

iv Resource is allocated to the improvement of data quality, subject to the level of risk / impact.

v The University uses visualisation tools to highlight the scale and impact of data quality issues.

2.10 **Data is sustainable**

2.2.1 Statement

The management of data, information and the underlying systems are sustainable.

2.2.2 Rationale

Systems and data contained within them are not sustainable if sufficient resource is not allocated to their management. Where the data supports business processes locally, and across the University, there is a reliance on that data to be available, maintained and of sufficient quality for the lifetime of the wider processes it supports.

2.2.3 Implications

i The wider context of data and connected systems must be considered for any change in resourcing, business processes or management of the system.

ii Interdependencies between data and systems are catalogued.

iii Total cost of ownership and breadth of stakeholders of systems are understood.

iv It is recognised that reducing the resource in the management of systems and data, reduces the overall quality of systems, data and the services they support. This reduction in quality also introduces and increases information risks for the University.

3 Glossary

3.1 **Data**:

Distinct units of information such as facts, numbers, letters, symbols, usually formatted in a specific way; often stored in a database and suitable for processing by a computer.

Data is a form of representation of individual facts about the world or a representation of values or details attributed to objects and concepts. Data can be the relationships between objects and concepts. Data is a unit of information.

3.2 **Information**:

Data combined and processed into a meaningful form.

Data which is processed, organised, structured or presented in a given context so as to make it useful, it is called information.

3.3 **Structured Data**:

Data in a standardised, pre-defined, structure, model or format such that it can be processed in a controlled, automated manner. Structured data, by its nature is classified or categorised. Structured data is typically found in database systems. Examples of structured data include financial data such as accounting transactions, address details, demographic information, login details, online survey responses, and location data from smart phones.

3.4 **Unstructured Data**:

Data which does not have a standardised, pre-defined, structure, model or format. This is typical of documents which contain natural language content such as correspondence or reports. You cannot know, or be certain of, the format or structure of the content. Unstructured data can also include photos, video, audio, text files, social media content and presentations. By its nature, it is sometimes possible to process unstructured data automatically, but requires a significant investment of time, money or other resources.

3.5 **Semi-Structured Data**:

Semi-structured data is unstructured data, with additional classification or categorisation applied to it. For example, a document contains content which is not pre-defined, but may have been classified as an invoice, and have additional metadata stored against it such as the author and the financial year, by use of a document management system. Email is a form of semi-structured data, as the email contents are generally unstructured, but the email comes with additional metadata about the sender and recipient.

3.6 **Data Privacy Impact Assessment**:

Also known as **Data Protection Impact Assessment**, DPIA, PIA.

A process to help you identify and minimise the data protection risks of a project, as mandated by the General Data Protection Regulations (GDPR). You must do a DPIA for processing that is likely to result in a high risk to individuals. It is also good practice to do a DPIA for any other major project which requires the processing of personal data. The DPIA process is defined and managed by the University's Data Protection Officer.

3.7 **Data Quality**:

Data quality refers to the state of qualitative or quantitative pieces of information. There are many definitions of data quality, but data is generally considered high quality if it is "fit for its intended uses in operations, decision making and planning". Data is deemed of high quality if it correctly represents the real-world construct to which it refers. As the number of sources of data increases, the question of internal data consistency becomes significant, regardless of fitness for use for any particular external purpose.

3.8 **Data Quality Dimensions**:

Data Quality Dimensions describe desirable aspects of data, and are a useful way to measure, compare or assess data quality levels across different systems. While there is no universal definition of the dimensions, six core dimensions are normally considered when looking at data. These are completeness, validity, timeliness, uniqueness, accuracy and consistency.

3.9 **Consistency**:

Data across all systems, entities and attributes do not conflict and do not contradict. For example, employee status is terminated but pay status is active.

3.10 **Conformity**:

The data is following the set of standard data definitions like data type, size and format. For example, the date of birth of a customer is in the format "dd/mm/yyyy".

3.11 **Validity**:

Data conforms to expected rules. For example, in surveys, items such as age and nationality are typically limited to a set of options and open answers are not permitted.

3.12 **Integrity**:

Data which relies on references to other data to derive meaning is correctly referenced according to expected rules and avoids expected constraints. For example, an employee has one and only one manager, and is correctly linked to that manager.

3.13 **Accuracy**:

Accuracy is the degree to which data correctly reflects the real- world object or event being described. For example, the address listed is correct and up to date.

3.14 **Precision**:

The data has sufficiently detailed information. For example, a payroll system should record pounds and pence, whereas a currency conversation system may need to show the conversion rate down to a thousandth of a penny.

3.15 **Interpretability**:

The extent to which data is clear, unambiguous and can be comprehended. Interpretability is an important component of quality as it enables the information to be understood and utilised appropriately.

3.16 **Timeliness**:

Timeliness is whether data is available when it is expected and needed. Timeliness depends on user expectation. Online, real time availability of data could be required for a meeting room booking system, but an overnight data feed could be perfectly acceptable for a billing system.

3.17 **Attribute**:

A quality or feature regarded as a characteristic or inherent part of someone or something. A single piece or entity of data.

3.18 **Authoritative System**:

The system which is recognised as being the source of, and maintains the version of, the truth for a particular data item or set of data. Where there are multiple information systems as sources for information, they may disagree about this same piece of information. These disagreements may stem from differences in interpretation, use of different sources, differences in the timing of the data exchange, or may simply be the result of bugs. The integrity and validity of any data set is open to question when there is no traceable connection to a good source.

3.19 **Automation**:

Automation is use of technology or software to perform a process or procedure with minimal to no human assistance.

3.20 **Change Control**:

Change control is a formal process used to ensure that changes to a product or system are introduced in a controlled and coordinated manner. It reduces the possibility that unnecessary changes will be introduced to a system without forethought, introducing faults into the system itself, or unintended consequences for other systems which are dependent upon it. The goals of a change control procedure include minimal disruption to services, reduction in back-out activities, and cost-effective utilisation of resources involved in implementing change.

3.21 **Completeness**:

The data meets the expected level of comprehensiveness. Data can be complete even if optional data is missing. As long as the data meets the expectations then the data is considered complete.

For example, a customer's first name and last name are mandatory but middle name is optional; so a record can be considered complete even if a middle name is not available.

3.22 **Data Minimisation**:

Data Minimisation is a principle that states that data collected and processed should not be held or further used unless this is essential for reasons that were clearly stated in advance to support data privacy.

You must ensure the personal data you are processing is:

- adequate – sufficient to properly fulfil your stated purpose;
- relevant – has a rational link to that purpose; and
- limited to what is necessary – you do not hold more than you need for that purpose.

Standards for the length that data should be retained are held within the University Retention Schedule[2].

3.23 **Data Quality Issue**:

A data quality issue can be defined as a matter that causes the high quality of the data to be in dispute. Data quality is concerned with the accuracy and completeness of the data among other key factors, and it needs to be fit for its intended uses.

3.24 **Data Sharing Agreement**:

Also know as **Data Processing Contract**, DSA, DPC.

A legal document which clearly documents what data is being shared by the University with another organisation or external party, and how that data should be processed. The agreement should include:

- the period of the agreement / length of processing;
- intended use of the data;
- confidentiality and security;
- whether the data can be further shared with other parties
- liabilities;
- constraints; and
- what should happen with the data at the end of the relationship.

3.25 **Data Standards**:

Rules by which the data is described, collected and amended. The rules should define the format of data and make reference to expectations according to the Data Quality Dimensions.

3.26 **Information Asset Owner** (formerly Data Steward):

A role within an organisation responsible for utilising an organisation's data governance processes to ensure fitness, security and appropriateness of data, both in processing and retention or disposal.

3.27 **Data Subject**:

A data subject is any living person whose personal data is being collected, held or processed. A person who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person.

3.28 **Double Handling**:

Double handling is the action of completing processes more times than necessary i.e. redoing work.

3.29 **Federation**:

The processing of data in multiple locations or sources, but with agreed or negotiated control.

3.30 **Information Asset**:

An information asset is a body of knowledge that is organized and managed as a single entity. Like any other corporate asset, an organisation's information assets have value.

3.31 **Information Asset Management**:

The governance of information assets by policies, procedures and processes. Ensuring information assets are properly described, categorised, secured and risk managed.

3.32 **Information Commissioners Office**:

Also **Information Commissioner**, ICO.

The UK's independent regulatory office (national data protection authority) dealing with the Data Protection Act 2018 and the General Data Protection Regulation. The ICO can issue enforcement notices and fines of up to 4% of global turnover, or €20 million (whichever is greater).

3.33 **Information Exchange**:

The act of giving and receiving information, typically electronically, via integrated IT systems.

3.34 **Information Management**:

Also **Data Management**.

Information management concerns a cycle of organisational activity: the acquisition of information from one or more sources, the custodianship and the distribution of that information to those who need it, and its ultimate disposition through archiving or deletion. This cycle of organisational involvement with information involves a variety of stakeholders, including those who are responsible for assuring the quality, accessibility and utility of acquired information; those who are responsible for its safe storage and disposal; and those who need it for decision making. Stakeholders might have rights to originate, change, distribute or delete information according to organisational information management policies.

3.35 **Interoperability**:

The degree to which a system or process is connected.

3.36 **Integrated, Interoperable, Connected**:

The ability of computer systems or software to exchange data in real-time or on a schedule. Contrasted with manual keying of data from one system to another by manual human intervention. Interoperable implies integration between systems, without the use of middleware.

3.37 **Master Data Management (MDM)**:

Master data management (MDM) is a method used to define and manage the critical data of an organisation to provide, with data integration, a single point of reference. The data that is mastered may include reference data - the set of permissible values, and the analytical data that supports decision making.

3.38 **Middleware**:

Middleware is software that provides a service to allow multiple systems / software applications to communicate. It can be thought of as a middleman, or "software glue".

3.39 **Personal Data**:

Also **Personal Information**.

Any information relating to a person (a 'data subject') who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person.

3.40 **Privacy by Design**:

Also **Data Protection by Design and Default**.

The design process of ensuring that systems and processes only make available or share the minimal necessary data. Ensures protection from inappropriate sharing or access. The GDPR requires you to put in place appropriate technical and organisational measures to implement the data protection principles and safeguard individual rights. This is 'data protection by design and by default'.

Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, the controller will, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects.

3.41 **Retention Schedule**:

Also **Retention Policy**.

A retention schedule is a control, document or policy which lists the organisational information types, records, or series of information in a manner which facilitates the understanding and application of the identified and approved retention period, and other information retention aspects. The retention period specifies the minimum amount of time for which the data / information must be retained, and then should be securely destroyed once no longer necessary for its purpose.

The University's Retention Schedule can be found in Appendix I, UPR IM11[2].

3.42 **Stakeholder**:

An individual, team, group, organisation, member, or system that affects or can be affected by an organization's actions. In the context of a project, stakeholders are those who may affect, be affected by, or perceive themselves to be affected by a decision, activity, or outcome of a project.

3.43 **Total Cost of Ownership (TCO)**:

Total cost of ownership (TCO) is a financial estimate intended to determine the direct and indirect costs of a product or system. TCO takes into account more than just the costs of software, but the greater costs of maintaining and supporting that software by way of infrastructure, staffing and other indirect costs.

Sharon Harrison-Barker
Secretary and Registrar
Signed: **3 January 2023**

Alternative format
If you need this document in an alternative format, please email us at governanceservices@herts.ac.uk or telephone us on +44 (0)1707 28 6006.