

# Data Management Principles

---

## Principle 1

### **Data is an Asset**

Data is a core business asset that has value to TfL and is managed accordingly.

#### **Rationale:**

*Data is a valuable corporate resource; it has real, measurable value. This is especially so in the case of £multi-million contracts under PPP and PFI frameworks for running London Underground services. In simple terms, the purpose of data is to aid decision-making. Accurate, timely data is critical to accurate, timely decisions. Most corporate assets are carefully managed, and as up to 80% of an organisation's value rests in its data assets, data can be no exception. Data is the foundation of our decision-making, so we must also carefully manage data to ensure that we know what we've got, where it is, can rely upon its accuracy, and can obtain it when and where we need it.*

#### **Implications:**

*This is one of three closely-related principles regarding data: data is an asset; data is shared; and data is easily accessible. The implication is that there is an education task to ensure that all departments within TfL understand the relationship between value of data, sharing of data, and accessibility to data.*

*Data used in TfL must be classified into high-level groupings to which final accountability for quality can be assigned.*

*If any employee has responsibility for upkeep of data in any form, this must be clearly stated in their job description and identified as a measure in their Performance Review process. In practice this will apply to all office-based employees to varying degrees.*

*All employees with any kind of data responsibility must have the authority and means to manage the data for which they are accountable.*

*Procedures must be developed and used to prevent and correct errors in information and to improve those processes that produce flawed information. Data quality will need to be measured and steps taken to improve data quality - it is probable that policy and procedures will need to be developed for this as well.*

*A forum with comprehensive enterprise-wide representation should decide on process changes suggested by data stewards (or role with similar term implying custodianship or trusteeship).*

*Since data is an asset of value to the entire enterprise, data stewards accountable for creation, quality and retention/disposal of data must*

*be assigned at a suitably senior level.*

---

## **Principle 2**

### ***Data is Shared***

Users have access to the data necessary to perform their duties; therefore data is shared across enterprise functions and departments.

#### ***Rationale:***

*Timely access to accurate data is essential to improving the quality and efficiency of enterprise decision-making. It is less costly to maintain timely, accurate data in a single application, and then share it, than it is to maintain duplicative data in multiple applications. TfL holds a wealth of data, but it is stored in hundreds of incompatible stovepipe databases and file-servers. The speed of data collection, creation, transfer, and assimilation is driven by the ability of the departments to efficiently share these islands of data across TfL.*

*Shared data will result in improved decisions since we will rely on fewer sources of more accurate and timely managed data for all of our decision-making. Electronically shared data will result in increased efficiency when existing data entities can be re-used.*

*Additionally, barriers to the outside world must come down too. We work with more and more external business partners, and efficient sharing of information assets is essential for this to work. It is more effective to de-protect. Legacy departmental culture tends to be over-protectionist. In fact, all the evidence is that more value can be derived from sharing more.*

#### ***Implications:***

*This is one of three closely-related principles regarding data: data is an asset; data is shared; and data is easily accessible. The implication is that there is an education task to ensure that all departments within TfL understand the relationship between value of data, sharing of data, and accessibility to data.*

*To enable data sharing we must develop and abide by a common set of policies, procedures, and standards governing data management and access for both the short and the long term.*

*For the short term, to preserve our significant investment in legacy systems, we may have to invest in software capable of migrating legacy system data into a shared data environment.*

*We will also need to develop standard data models, data elements, and other metadata that defines this shared environment and develop a repository system for storing this metadata to make it accessible.*

*For the long term, as legacy systems are replaced, we must adopt and enforce common data access policies and guidelines for new application developers to ensure that data in new applications remains*

*available to the shared environment and that data in the shared environment can continue to be used by the new applications.*

*For both the short term and the long term we must adopt common methods and tools for creating, maintaining, and accessing the data shared across TfL.*

*Data sharing will require a significant cultural change.*

*This principle of data sharing will continually "bump up against" the principle of data security. Under no circumstances will the data sharing principle cause confidential data to be compromised.*

*Data made available for sharing will have to be relied upon by all users to execute their respective tasks. This will ensure that only the most accurate and timely data is relied upon for decision-making. Shared data will become a TfL-wide "virtual single source" of data.*

---

## **Principle 3**

### ***Data is Accessible***

Data is accessible for users to perform their functions.

#### ***Rationale:***

*Wide access to data leads to efficiency and effectiveness in decision-making, and affords timely response to information requests and service delivery. Using information must be considered from an enterprise perspective to allow access by a wide variety of users. Staff time is saved and consistency of data is improved.*

#### ***Implications:***

*This is one of three closely-related principles regarding data: data is an asset; data is shared; and data is easily accessible. The implication is that there is an education task to ensure that all departments within TfL understand the relationship between value of data, sharing of data, and accessibility to data.*

*Accessibility involves the ease with which users obtain information.*

*The way information is accessed and displayed must be sufficiently adaptable to meet a wide range of enterprise users and their corresponding methods of access.*

*Access to data does not constitute understanding of the data. Personnel should take caution not to misinterpret information.*

---

## Principle 4

### **Data Quality is Fit for Purpose**

Data quality is acceptable and meets the business need for which it is intended.

**Rationale:**

*Data produced and reported must be fit for purpose. That is, of sufficient accuracy and integrity proportional to its use and cost of collection and maintenance.*

*Data is used in all areas of decision-making, operations, planning and performance management in order that TfL achieves its objectives. It is increasingly being used externally by citizens and customers to inform their personal decisions, and by stakeholders to assess the aggregate performance of TfL. This reinforces the need to ensure that the quality of data held is sufficient to meet diverse needs.*

*Significant human and system resource is consumed in the collection, manipulation and dissemination of data whether of high quality or not, so it is essential that the most effective use of public funds is achieved through appropriately directed attention to data quality and the procedures to realise quality.*

**Implications:**

*Data should be sufficiently accurate for its intended purpose, representing clearly and in sufficient detail the activity which it represents*

*Disciplines and standards will need to be put in place and monitored so that accuracy and integrity of data is assured*

*Collection and manipulation of data should be reliable and should reflect consistent processes where needed between departments, to allow for meaningful comparison where appropriate.*

*Data collected should be complete and captured once 'right first time' such that it can be aggregated, analysed and manipulated for decision making purposes*

*Data should be timely, so that its usefulness for decision making can be maximised*

---

## Principle 5

### ***Data is Compliant with Law and Regulations***

TfL's information management processes comply with all relevant laws, policies, and regulations.

***Rationale:***

*There are a number of legal requirements that govern the use of data in the course of TfL business. As a public authority with a clear ethical imperative and world class reputation to maintain, compliance is an essential business driver.*

***Implications:***

*TfL must be mindful to comply with laws, regulations, and external policies regarding the collection, retention, and management of data.*

*Education and access to the rules, efficiency, need, and common sense are not the only drivers. Changes in the law and changes in regulations may drive changes in our processes or applications.*

---

## Principle 6

### ***Data is Secure***

Data is trustworthy and is safeguarded from unauthorized access, whether malicious, fraudulent or erroneous

***Rationale:***

*Open sharing of information and the release of information via relevant legislation must be balanced against the need to restrict the availability of classified, proprietary, and sensitive information.*

*Existing laws and regulations require the safeguarding of national security and the privacy of data, while permitting free and open access. Pre-decisional (work-in-progress, not yet authorized for release) information must be protected to avoid unwarranted speculation, misinterpretation, and inappropriate use. Integrity, confidentiality and availability are maintained as long as information is needed.*

***Implications:***

*TfL's technology infrastructure should move towards a single directory-based system that provides authentication services to each and every application, database, file-server and collaboration environment. Each of the latter should then manage access control appropriate to the business needs of each user identity.*

*Data security safeguards can be put in place to restrict access to "view only", "know only of its existence" or "not know of existence". Sensitivity labeling for information access will be based on uncustomised government-standard 'Protect' flag and protective marking schemes.*

*Security must be designed into data elements from the beginning; it cannot be added later. Systems, data, and technologies must be protected from unauthorized access and manipulation. Information must be safeguarded against inadvertent or unauthorized alteration, sabotage, disaster, or disclosure.*

---

## **Principle 7**

### ***There is a Common Vocabulary and Data Definition***

Data is defined consistently throughout TfL, and the definitions are understandable and available to all users.

***Rationale:***

*Both unstructured and structured data must have a common definition throughout TfL to enable sharing of data. A common vocabulary will facilitate communications, enable dialogue to be effective and facilitate interoperability of systems.*

***Implications:***

*TfL must establish a common vocabulary for the business. The definitions will be used uniformly throughout TfL.*

*Whenever a new data definition is required, the definition effort will be co-ordinated and reconciled with the corporate "glossary" of data descriptions. The TfL Group IM Data Authority will govern this co-ordination.*

*Ambiguities resulting from multiple parochial definitions of data must at worst be mapped as 'non-preferred terms' to a 'preferred term', and at best give way to accepted enterprise-wide definitions and understanding.*

*Multiple data standardization initiatives need to be co-ordinated.*

*Functional data administration responsibilities must be assigned.*

---

## **Principle 8**

### ***Data is Not Duplicated***

Development of information services (such as business applications, data warehouses, directory services etc) available across TfL is preferred over the

development of information silos which are only provided to a particular department or group of departments.

**Rationale:**

*Duplicative capability is expensive and propagates conflicting data. It also militates against a policy of sustainability in the use of infrastructure resources such as servers and data centre air conditioning.*

**Implications:**

*Departments will not be allowed to develop capabilities for their own use which are similar/duplicative of enterprise-wide capabilities. In this way, expenditures of scarce resources to develop essentially the same capability in marginally different ways will be reduced.*

*Departments which depend on a capability which does not serve the entire enterprise must change over to the replacement enterprise-wide capability as soon as practically possible, if it is available.*

*The design of business service capabilities to replace silo applications will be driven directly by the business processes they are designed to support.*

---

## **Principle 9**

### ***Data Management is Everybody's Business***

All departments in TfL participate in information management decisions needed to accomplish business objectives.

**Rationale:**

*Information users are the key stakeholders in the application of technology to address a business need. In order to ensure information management is aligned with the business, all departments in TfL must be involved in all aspects of the information environment. The business experts from across TfL and the technical staff responsible for developing and sustaining the information environment need to come together as a team to jointly define the goals and objectives of IT.*

**Implications:**

*To operate as a team, every stakeholder will need to accept responsibility for developing the information environment.*

*Commitment of resources in business departments will be required to implement this principle.*

---

## Principle 10

### ***Decisions Maximise the Benefit of Data to TfL***

A spirit and culture of collaboration and the sharing of data, information and knowledge for the greater corporate good shall pervade all decision-making, especially relating to the selection and prioritisation of programmes, projects and their approval points.

***Rationale:***

*This principle embodies "service above self". Decisions made from an enterprise-wide perspective have greater long-term value than decisions made from any particular departmental perspective. Maximum return on investment requires information management decisions to adhere to enterprise-wide drivers and priorities. No minority group will detract from the benefit of the whole. However, this principle will not preclude any minority group from getting its job done.*

***Implications:***

*Achieving maximum enterprise-wide benefit will require changes in the way we plan and manage information. Technology alone will not bring about this change.*

*Some departments may have to concede their own preferences for the greater benefit of the entire enterprise.*

*Application development priorities should be established by the entire enterprise for the entire enterprise.*

*Application components should be shared across departmental boundaries.*

*As needs arise, priorities must be adjusted. A forum with comprehensive enterprise representation should make these decisions.*