

# **New Zealand Government Data Management Standards**

version 1.1, July 2000

# **New Zealand Government Data Management Standards**

Version Notes	4
<b>INTRODUCTION</b>	<b>5</b>
<b>OWNERSHIP</b>	<b>6</b>
Reasons for collection/creation	7
Transfer of intellectual property	8
<b>CUSTODIANSHIP</b>	<b>10</b>
Physical custodian agreement	11
<b>ACCESS RULES</b>	<b>14</b>
Publishing	15
Security	17
The position, not the person	18
Secure electronic exchange	19
Individual privacy and confidentiality	20
Commercial sensitivity	21
Equity of access	22
<b>IDENTIFICATION</b>	<b>23</b>
Individual responsibilities	24
Discovery implications	25
Data catalogue and function map	26
Process maps	28
<b>CONTEXT</b>	<b>30</b>
Core metadata	31
Media-independent classification of documents	32
<b>SOURCE</b>	<b>34</b>
Prime authoritative data source location	35
Synchronise data and document stores and publishing systems	36
<b>AUTHENTICITY, INTEGRITY, RETRIEVABILITY</b>	<b>37</b>
Referential Integrity	38

<b>Integrity of Application Software</b>	<b>39</b>
<b>Integrity of configuration</b>	<b>40</b>
<b>Integrity of content</b>	<b>41</b>
<b>Integrity of Process</b>	<b>43</b>
<b>Skills and training</b>	<b>44</b>
<b>Version Control</b>	<b>45</b>
<b>Document Templates</b>	<b>46</b>
<b>Retrievability</b>	<b>47</b>
<b>AUDITABILITY</b>	<b>49</b>
<b>Change Control</b>	<b>50</b>
<b>Audit trail</b>	<b>51</b>
<b>INTERCHANGE, REPLICATION, INTERFACES</b>	<b>53</b>
<b>Interchange Agreement</b>	<b>54</b>
<b>Replication</b>	<b>55</b>
<b>Interfaces</b>	<b>56</b>
<b>Migration</b>	<b>57</b>
<b>RETENTION</b>	<b>59</b>
<b>Backup, Recovery &amp; Restore</b>	<b>60</b>
<b>Storage Media</b>	<b>62</b>
<b>Disaster Recovery</b>	<b>63</b>
<b>Retention requirements</b>	<b>64</b>
<b>Transfer between Agencies</b>	<b>66</b>
<b>Destruction Protocols</b>	<b>67</b>

## Version Notes

<b>Version Number</b>	<b>1.0a</b>
Summary of changes affecting meaning	Nil
Other changes	Correct titles used in internal cross-references
<b>Version Number</b>	<b>1.1</b>
Summary of changes affecting meaning	Changes resulting from input from Privacy Commission: <ol style="list-style-type: none"><li>1. Removed references to “private information” and replaced them with</li><li>2. Reasons for collection and creation: Re-worded reference to Privacy Act;</li><li>3. Individual privacy and confidentiality: Replaced “guidelines” with “codes of practice”, noted need to prevent ‘fishing’ for personal information;</li><li>4. Added warning that “inappropriate comment” mixed with business information will be discoverable as part of the record;</li><li>5. Data Catalogue and Function Map: Add “Reasons for collection of data” to table of suggested metadata elements;</li><li>6. Prime authoritative data source location: Added Privacy Commissioner requirement re providing information from replications to personal information access requests;</li><li>7. Destruction protocols: Noted applicability of Privacy Act.</li></ol>
Other changes	Nil

## **INTRODUCTION**

This standards document complements, and should be read in conjunction with, the “New Zealand Government Data Management Policies” document.

As noted in the Policies document, the standards emphasise best practice and minimum quality levels for implementation of these policies, rather than prescribing technological solutions which could be beyond the reach of smaller agencies. Government agencies who provided early input to the development advised that the policies and standards should avoid going beyond generic best practice into business, legal and technology specific areas (these are typically low-level, operational policies and standards).

The Policies and Standards should be regarded as living documents, to be improved by regular reviews based on feedback from the Public Service, legislative changes, developments in international standards etc. It is unlikely that any government agency will be fully compliant from day one, however every agency will be required to demonstrate reasonable progress towards compliance.

# Ownership

## **Standard**

### ***Reasons for collection/creation***

Criteria for collection or creation of data and documents must meet explicit legal or business requirements and be subject to scrutiny by the Business Custodian.

The supplier must be informed which data or document collection is compulsory, which is voluntary and the reason for collection.

### **Supports Policies:** *Ownership*

### **Scope and Interpretation**

Agencies must collect or create data and documents only in the execution of statutory or defined business requirements. Those requirements must be documented to a standard set by the Business Custodian, before collection or creation. Legislation must be quoted, where applicable. It is important to identify where the business requirements may include the collection on behalf of other agencies, eg for the Department of Statistics.

The Crown Data Steward will audit the defined requirements. Unless the criteria for collection are themselves classified, the public must be able to see those criteria (see [Publication](#)).

Any existing creation or collection must be reviewed, and where there is no requirement, creation or collection should cease. Agencies must inform the Crown Data Steward when and why they intend to stop collecting data. The Crown Data Steward will then review the ongoing need for creation/collection, with a review of any affected processes in agencies.

When collecting information, agencies must identify whether each data element requested is mandatory or whether the collection is voluntary and helpful for the organisation to conduct its business. Collection methods, e.g. paper or electronic forms, must make it easy to distinguish mandatory from voluntary supply. The consent of individuals will be needed where data is to be used for other government purposes e.g. public registers.

### **Rationale**

The purpose for collection must be well defined in order to prevent the collection or creation of irrelevant, unnecessarily intrusive or meaningless information. Trust and compliance of the supplier will be enhanced by above which should in turn improve quality of data. This is a demonstration of the Crown acting in good faith.

Provisions of the Privacy Act, the Statistics Act, and the Official Information Act apply when implementing this standard.

## **Standard**

### ***Transfer of intellectual property***

Where transfer of exclusive intellectual property to the Crown is by statute, collecting agencies must ensure that the supplier is informed of the transfer.

In the absence of statute, any transfer of intellectual property to the Crown must be by explicit agreement between the Crown and the supplier.

In the absence of statutory authority, agencies must have the approval of the Crown Data Steward to transfer out of Crown ownership either the Crown copy of collected data or the intellectual property of the Crown.

### **Supports Policies:** *Ownership*

#### **Scope and Interpretation**

Statements of statutory requirements for the transfer of intellectual property must be publicly and equitably available. For formally constituted groups, informing them of the transfer of intellectual property could mean a formal directive, or a formal communication directing their attention to statutory requirements. For individuals or groups supplying data or documents, there could be a formal statement on a data collection form advising the supplier of the transfer. Statutory requirements must also be available via the New Zealand Government website, NZGO.

Transfer of intellectual property rights may or may not include copyright (see [Glossary: Copyright](#) for more details).

If transfer to the Crown is by agreement then the current owner must be able to ultimately refuse the transaction if desired. If transfer is by statute, the original owner should at least have due notice, ideally before the data is collected or documents created.

The Crown Data Steward will refer proposed transfers of data or documents out of Crown ownership for assessment by the Chief Archivist.

#### **Rationale**

Transfer of intellectual property must be explicit, to preserve the rights of individuals and organisations exchanging data and documents with the Crown, and to preserve the assets of the Crown.

Similarly to the standard *Reasons for Collection/Creation*, this standard enhances trust and compliance by the supplier.



# Custodianship

## Standard

### ***Physical custodian agreement***

The Business Custodian for an agency will put in place and monitor explicit directives or agreements with the Physical Custodian(s) for managing data and document stores.

**Supports Policies:** *Custodianship*

#### **Scope and Interpretation**

The purpose of the agreement is to make the expectations of the Business Custodian explicit to the Physical Custodian, and to ensure that any misunderstandings, in particular relating to technology, are resolved. It should allow each agency to confidently report that its data and document assets are being well managed and that services can be audited against a specific agreement.

Where the physical custodian is an external service provider, the agreement will be part of a formal service contract. If the agency has an internal information technology group it will be an internal agreement between the business and technology areas. An agency may have datasets in the custody of different physical custodians. There must be agreements between the Business Custodian and each Physical Custodian.

In some agencies the roles of Business and Physical Custodians may fall to the same person for some categories of data or documents (see [Policies: Access rules](#) and related table *Indicative Categories of Data and Access rules*). In these cases the agreement will take the form of a clear definition of each role and the responsibilities involved.

As well as requirements specific to the agency, the following must be covered in the agreement:

- Adherence to these policies and standards, agency specific extensions, and any specific legal requirements
- Maintenance of equipment and system software required to access and maintain data or documents
- Maintenance of application software required to access and maintain data or documents if applicable
- Maintenance of the physical environment where equipment or physical media are stored to the required standard
- Ability to connect data and document stores to a government wide standards based technology infrastructure
- Enforcement of data and document security rules
- Explicit backup, off-line storage, and restoration provisions
- Explicit disaster recovery plans
- Maintenance of a current data catalogue for each operational system if required by the Business Custodian
- Maintenance of system documentation
- Readiness to be audited at any time on a request from the Business Custodian
- Management of the physical aspects of data or document conversion, or of reformatting as required by the Business Custodian
- Proper separation of the production environment from any other and adequate migration procedures between them
- Transmission and transportation of data either via electronic interface or secure physical media
- Regular performance reporting

**Rationale**

Explicit agreements are necessary so that both parties have clear common understanding of what is required.



# Access rules

## **Standard**

### ***Publishing***

All current ratified *internal* policy and standards documents will be published within the agency as a discrete set.

All current ratified policy and standards documents of direct relevance to the public will be available as a discrete set, via NZGO, and available through any reasonable delivery methods necessary for equitable access for the public.

Other publications of direct relevance to the public should increasingly be available electronically.

### **Supports Policies:** *Access rules*

#### **Scope and Interpretation**

The policy applies to all policies and standards approved through **official** processes as **agency** policies and standards.

Ability to retrieve the material as a discrete set means that a simple proven method exists to retrieve all relevant policies and standards. The method must be available to all users and be guaranteed to be complete without producing unnecessary clutter.

Implementation should be simple for small agencies e.g. an internal manual available in paper and/or electronic form, with public material available via the New Zealand Government On-line website (NZGO). "Via NZGO" can mean either published on the NZGO site, or linked via that site.

Other equitable delivery methods must be able to ensure access for the public without transferring costs from the publishing agency.

The agency must be able to demonstrate that:

- All current internal policies and standards are available to all staff
- Staff are trained on how to access current policies and standards
- All current public policy and standards are available to the public via NZGO
- Delivery methods must include electronic form
- Members of the public are made aware of how to access current public policies and standards
- Superseded versions of policies and standards are available, with the contextual information of when they were in effect
- Updates of internal policies and standards are brought to the attention of staff affected by them
- Updates of public policies and standards must be clearly identified, so that members of the public can observe them

Draft versions must be distinguished from current policy and standards documents. Superseded or withdrawn policies and standards will be available within the agency, but will be identified as no longer in effect. Any on-line publishing system must not only provide access to the current and complete policy, but also must also be quick to access, so that staff and/or members of the public will choose to access it.

### **Rationale**

Internal policies and standards are **key agency documents** that should be actively published to ensure that they are up to date, and that they are completely, consistently, and readily available to staff.

For these critical reference documents, extra care and effort will be taken to maximise the ease with which staff can retrieve information, otherwise staff will revert to less authenticated sources of policy and standards. This has special relevance in urgent situations, such as those encountered by operational staff referring to policy while a client is waiting.

Policy of direct relevance to the public must be truly available to the public.

## **Standard**

### ***Security***

Security systems that control access to document and data stores must be designed to implement access rules defined by the agency, regardless of the storage medium

**Supports Policies:** *Access rules*

### **Scope and Interpretation**

Each agency must develop, document and maintain access rules within legislative and business restrictions. These access rules will include compliance with the security provisions defined by the Department of Prime Minister and Cabinet. See also *Generic Business Security Policy for Government* published by the State Services Commission.

Any approved data or document stores must be capable of recording and enforcing these access rules, regardless of complexity. For a small office this may be simply using secure directories on a PC, and storing paper documents in a locked filing cabinet. Large organisations may require full-scale databases with several layers of security attributes.

Documents may exist in either electronic or physical format or both, and each agency must ensure that its systems can maintain access rules regardless of storage medium. All staff must be aware of their responsibilities in handling all information, particularly restricted, security-classified, personal and commercially sensitive information.

Ongoing maintenance of access rules is required to allow for changes in legislation and business requirements.

### **Rationale**

Data and business documents held by agencies in approved document stores must be available to all those with a legitimate requirement for access, and protected from unauthorised access. The agency needs to ensure and demonstrate that its information is stored and accessed in accordance with applicable privacy and confidentiality requirements.

Effective implementation of access rules mitigates the risks of breaches of security and/or privacy of information held by the agency. In the event of a breach of security or privacy, the agency should be able to identify its source and nature, recognising that some breaches may be accidental or inadvertent.

## **Standard**

### ***The position, not the person***

Security must be on a role basis so maintenance level control of data or documents goes with a position – not with a person.

**Supports Policies:** *Custodianship; Access Rules*

### **Scope and Interpretation**

Data elements in a database must be available to people currently holding an authorised access/position and not confined to nominated individuals. Documents must be retrievable by an agency regardless of whether their creators are still employed in the positions they held when they created the document.

The Business Custodian is responsible for ensuring that the standard is put into practice. Any list or database table relating staff and positions must be kept up-to-date.

If a position is dis-established its associated security role(s) must be linked to a current position to ensure continuity of access.

### **Rationale**

Implementation of this standard reduces the risk of lost or unavailable material. The agency needs to have access to all business data and documents:

- Staff members moving into positions need to have access to all data and documents created by their predecessors
- Managers need access to the data and documents created by departed staff
- Staff members moving out of a position should no longer retain access restricted to that previous position.

## **Standard**

### ***Secure electronic exchange***

When exchanging electronic data or documents, government agencies must transmit all security-classified material via secured electronic communications.

**Supports Policies:** *Access rules*

#### **Scope and Interpretation**

Each agency must have systems in place that protect security-classified data and documents, and must train users on the security classifications and their use.

Current examples of potentially unsecured communications are:

- Email transmitted via the internet, both messages and attachments, without encryption
- Using public switch networks, e.g. a public phone line from home without encryption
- Fax communications without encryption or where the receiving machine is not secured

When technology permits security, external partners may be given controlled and appropriate access to data or documents in approved stores. Technical solutions will be compliant with NZ Government standards for secure electronic exchange of information.

#### **Rationale**

Security-classified documents must be handled and stored to ensure their safety, and protect against loss or unauthorised disclosure.

## **Standard**

### ***Individual privacy and confidentiality***

To protect privacy and confidentiality of individuals, agencies will determine which types and instances of data and documents contain details about individuals. Provisions of the Privacy Act and Privacy Commissioner codes of practice must apply. In addition, access restrictions must define authorised user groups and the period for which the constraints apply.

**Supports Policies:** *Access rules*

### **Scope and Interpretation**

To assist in ensuring privacy and confidentiality, users will be able to identify whether any data set or document:

- Contains personal information on an identifiable individual who is the subject, whether or not the person is a member of the agency or external to it
- Contains personal information on identifiable third parties who are not directly the subject, whether or not those persons are members of the agency or external to it
- Contains aggregated summary information about individuals, and privacy of the individual is protected
- Does not contain information about individuals

To preserve privacy, for summary data any linkages must be removed between individual and summary data. In addition, unrestricted access to scan index listings without supplying identifiers associated with a particular individual must be avoided. While there may be a legitimate need for this type of facility, access rules must be applied to prevent unauthorised 'fishing' for personal information.

While the appropriate handling of all information about individuals is the responsibility of all staff, each agency also has a privacy officer legislated under the Privacy Act.

### **Rationale**

Assists the agency in ensuring only appropriate personal information is held and released. The agency can demonstrate the systems and processes it has in place to ensure the security of personal information held by it. The risk of unauthorised or inadvertent disclosure of personal information to third parties is managed. Provisions of the Privacy Act apply.

## **Standard**

### ***Commercial sensitivity***

To protect legitimate commercial interests, agencies will develop and maintain guidelines to determine which types and instances of data and documents contain commercially sensitive material. In addition, access restrictions must define authorised user groups and the period for which the constraints apply.

**Supports Policies:** *Access rules*

### **Scope and Interpretation**

Agencies will manage commercially sensitive information to ensure that information held is:

- Protected from unauthorised access
- Identifiable and accessible for audit of its storage and use
- Managed to comply with statutory and documented business requirements
- Auditable to provide a record of access
- Not retained unnecessarily

The appropriate handling of all commercial information is the responsibility of all staff. All commercially sensitive information must be clearly identified, guidelines must be in place for its use, and staff must be trained in those guidelines.

Each agency must be able to demonstrate the systems and processes it has in place to ensure the security of commercially sensitive information held by it. In some cases specific agreements with commercial entities may be required.

### **Rationale**

Assists agencies to ensure that commercially sensitive information is secure. Some agencies are such large players in specialist markets that their decisions can affect the viability of commercial companies.

The risk of unauthorised or inadvertent disclosure of commercially sensitive information to third parties must be managed.

## **Standard**

### ***Equity of access***

Each agency must develop a strategy and an implementation plan to deliver equity of access to data and business documents, within the agency, between itself and other agencies, and for the public.

**Supports Policies:** *Access rules*

### **Scope and Interpretation**

Equity of access differs from equality of access.

Within an agency, staff at any level and at all locations must have access to data and documents such that they can do their work efficiently and effectively.

Equity of access between government agencies could be an issue if agencies legitimately need to access information but encounter difficulties related to their systems, size, etc.

Equal access for the public might be that all government agencies offer the same information to all New Zealanders via the same delivery mechanism, e.g. the Internet. Equity of access might be that all New Zealanders can readily access appropriate government information whether or not they have a computer, phone and modem, and whether or not they live in a major centre. There is a trend for agencies to physically withdraw from small centres and rely on technology to provide services. However, small centres may be technology poor.

Issues of language, culture, age, disability etc may also impinge on equity of access. Agencies should consider the implications of their practices on access in relation to these sorts of issues.

Development of strategy and an implementation plan will ensure that the issue is considered and discussed, instead of using implicit assumptions. This process includes:

- Who needs to access the information
- What information they want to access, or could access if they knew it existed
- What delivery mechanisms they need/prefer to access it
- What delivery mechanisms exist
- Gap analysis, including reasons for gaps and strategies to minimise them
- Reasons for lack of action to address any identified gaps

### **Rationale**

The government drive to electronic government opens up new opportunities to access data and documents, and new access delivery mechanisms. This could both improve and diminish equity of access, depending on awareness and responsiveness of government agencies to issues about equity of access.

# Identification

## **Standard**

### ***Individual responsibilities***

Individuals have a set of responsibilities when creating or maintaining data and documents for an agency. Agencies must define those responsibilities and deliver standards, processes and systems to monitor and support them.

**Supports Policies:** *Data Identification; Document Identification and Capture*

### **Scope and Interpretation**

The emphasis in this standard is on individual responsibility, and applies to all media and all contributors whether inside or outside the agency. This includes reference to responsibilities defined outside the agency e.g. the Public Service code of conduct. Examples include:

- the need to be accurate and honest
- completeness –ensure all relevant data and documents are identified and captured either directly or by specific delegation
- using agency conventions and applying quality checks, etc. see [Standard: Skills and training](#).

For standards on processes and systems, see [Standards: Integrity of configuration](#), [Integrity of content](#), [Integrity of process](#), [Version control](#), [Document templates](#).

The Business Custodian is responsible for ensuring standards, processes, systems and a monitoring regime are in place, while individual users are responsible for compliance. Monitoring could include analysis of data element or document content to verify accuracy as well as reviewing individual work practices.

Each system must be designed so that

- Individuals find it facilitates compliance with business requirements
- Individuals find it beneficial to enter data or create documents correctly

### **Rationale**

Individual commitment is an essential prerequisite to the success of managing data and business documents in approved stores. Individuals must understand their responsibilities and competently process appropriate data and documents for their own immediate benefit and for that of the agency.

## **Standard**

### ***Discovery implications***

When creating, modifying or saving data or documents in any storage medium, staff must know the discovery requirements of external legislation, e.g. the Official Information Act, and court requirements for evidence.

**Supports Policies:** *Document Identification and Capture; Data Identification*

### **Scope and Interpretation**

Anyone creating or receiving information for an agency needs to be aware that of the potential for public or private release of information at some point in the future. For example, if combining business information and personal information or inappropriate comment e.g. in a business email, staff must bear in mind that all content will be discoverable as part of the record. Similarly, database notes must be robust enough to stand up to internal and external scrutiny for relevance and accuracy. This standard applies to all media, including paper based material.

The agency is at risk of “vicarious liability” for the actions of individuals unless it can demonstrate that there is nothing the employer could have done. Data or documents created by the individual can lead to corporate liability, even where the individual perceives the record as personal.

Each agency will need to focus not only on keeping data and documents, but also on timely destruction when retention periods expire. See [Standards: Retention Requirements](#), [Destruction Protocols](#).

Management has a responsibility to train staff to understand the implications of relevant legislation on data or documents they use and create, and to monitor compliance. Staff must be directed to internal or government codes of conduct and specific policies e.g. for Email use.

### **Rationale**

Information in any medium can be sought by the courts or under the Official Information Act. There is a need to have a quality record of a business transaction.

Agencies can minimise their vicarious liability by ensuring the appropriateness of what is captured.

At present most requests from outside an agency are for physical documents. Electronic data and documents, including casual emails, can be used as evidence.

Individuals have a responsibility to record information appropriately. See [Standards: Individual responsibilities](#), [Integrity of process](#), [Skills and training](#). For standards on destruction, see [Standards: Retention requirements](#), [Destruction Protocols](#).

## Standard

### **Data catalogue and function map**

The data catalogue and function map for each agency will be in an electronic format consistent with government metadata standards and applicable international standards.

### **Supports Policies:** *Data Identification*

### **Scope and Interpretation**

The data catalogue will list both individual data elements and the containing data stores or datasets held as a prime source by the agency. In addition, high level business functions must be identified and mapped to the main data stores that support them. Over time standard terminology across government for common administrative functions can be established. The catalogue must be published in a common electronic format to be determined by the Crown Data Steward. This is likely to be compatible with either standard *ANSI Z39.50 “ Information retrieval application service definition...for open systems* , or the Lightweight Directory Access Protocol (LDAP) standard.

The catalogue content is designed to identify and briefly describe data and document assets to users within an agency and to external users where appropriate. It is not intended to make content available other than that of the metadata elements described below. It is not intended as a tool to support system design or maintenance, although the content of the catalogue may be largely derived from tools that define database schemas or system models. The following tables illustrate the minimum useful content.

<b>Data store or dataset metadata</b>	<b>Mandatory</b>	<b>Dublin Core equivalent</b>	<b>ISO/IEC 11179-3</b>
Object type – e.g. Oracle database, SAS dataset etc.	Yes	Object type	N/A
Controlling agency/Business Custodian	Yes	Publisher	N/A
Physical format in which data can be made available	If applicable	Form	N/A
Identifier	Yes	Identifier	N/A
Name	Yes	Title	N/A
Content description	Yes	Description	N/A
High level business functions supported	Yes	N/A	N/A
Relationships to other stores or datasets	Yes	Relation	N/A
Copyright or ownership statement	If applicable	Rights management	N/A
Sources if derived or replicated from other prime data sources	If applicable	Source	N/A
Date of last update and/or version– for datasets maintained at set intervals	If applicable	Date	N/A
Spatial location or time periods covered by the data in the store or dataset	If applicable	Coverage	N/A
Reasons for collection of data e.g. reference to legislation	Yes	N/A	N/A

<b>Data element or container metadata</b>	<b>Mandatory</b>	<b>Dublin Core base equivalent</b>	<b>ISO/IEC 11179-3</b>
Object type – e.g. Oracle column	Yes	Object type	N/A
Controlling agency/Business Custodian	Yes	Publisher	Responsible organisation
Physical format of the data element – size, data type etc.	Yes – minimum requirement: data type, max size, valid values (if needed)	Form	Category Form Data type Max size Min size Layout Valid values
Identifier	Yes	Identifier	Identifier
Name	Yes	Title	Name
Content description in business terms	Yes	Description	Definition
Subject Key words	Optional	Subject	Key words
Synonym names	Optional	N/A	Synonyms
Version	Optional	N/A	Version
Relationships to other data elements, container objects	Yes if applicable	Relation	Related data reference Type of relationship
Primary source data store or dataset	Yes	Source	Context

### **Rationale**

The catalogue allows the agency to define its prime authoritative data sources both to staff and to authorised external agencies. Over time this will reduce duplication and improve data usage. The Crown Data Steward will use agency catalogues to create a high level government wide catalogue.

While a variety of tools may be used to produce catalogues as appropriate to each agency, a common publishing format ensures simple and cost effective information sharing.

## **Standard**

### ***Process maps***

For each of its functions, each agency will have a process map to determine where data and business documents are captured or created, and how these processes relate to any statutory or business requirements.

**Supports Policies:** *Document Identification and Capture, Data Identification*

### **Scope and Interpretation**

Detailed process analysis is always required when developing a database system to ensure it will operate effectively within the organisation. This is part of the cost of developing an information system.

Some agencies have detailed and elaborate legacy processes for handling paper documents, so that the responsibilities for creating and filing material are clear. Mostly these processes have not been translated for use with electronic documents as a paper copy is still often considered to be the “file” copy.

With the virtual universal use of electronic documents in offices it is now becoming essential to apply discipline to both document creation and to the process of managing the resulting document stores. To do this effectively managers and staff must understand the relationships between document creation and document management, and act to ensure consistent results across all storage media. See also standard *Individual responsibility*

Examination of processes could be done on a “Pareto” basis, where the agency determines where to put its best effort to maximise returns. This may involve identification of key functions, cross agency processes and agency/private organisation processes. It is also important to determine where and when new versions of data or documents are created (see also [Standards: Version control](#))

Many processes are generic across agencies and there are potential benefits from agencies collaborating to map them.

### **Rationale**

Process mapping assists the organisation to know where its data and business documents are being created and how they are being managed. It also improves “buy in” from staff, since the focus is on statutory or business needs. Process mapping also reduces risk of important information never being captured, or of being handled inappropriately, and increases transparency and assists in audit.



# Context

Standard

### **Core metadata**

Core metadata about a data element and its context, document, or data/document store will be the minimum needed to identify and retrieve the content in a usable format .

#### **Supports Policies:** *Context*

#### **Scope and Interpretation**

Core metadata requirements must be based on the definitions and structures identified in the New Zealand Government metadata standard. The agency will relate its data capture to that standard, and work towards conforming to it. (see also [Standard: Data catalogue and function map](#)) This will allow the agency to:

- Ensure data elements or documents can be retrieved by the right users in a timely fashion
- Comply with legislative requirements
- Establish relevant contextual information, including a unique agency identifier
- .

Core metadata must be defined at three main levels of detail:

- The properties of a data or document store which could be a database, dataset, or paper document repository.
- The properties of a container which provides an immediate context for data elements. This could be a tightly organised structure such as database table or object class, or something more variable and familiar such as a document
- The properties of a base level data element

Core metadata capture will be kept to a minimum, with as much system-generated data as possible. However, systems should permit users to add extra information beyond mandatory data capture, in order to meet particular business needs. Where metadata values are user supplied, the creator of the data element or document will normally be responsible for their input and accuracy.

These provisions are equally applicable to electronic and paper based systems, (eg the metadata for a printed document in a physical file could identify the creating or custodial agency, the document store and the document itself).

#### **Rationale**

A consistent metadata framework is essential for the effective organisation and retrieval of stored data and documents both within and across government agencies.

E-Government initiatives to provide a single entry point for government services will fail without such a standard.

Standard

### ***Media-independent classification of documents***

If an agency has physical file systems and an electronic classification systems, those systems should use the same rules for describing and classifying documents.

**Supports Policies:** *Context; Retention*

#### **Scope and Interpretation**

The concept of the traditional “file” may become that of a container for readily grouping material of common functions. This standard does not compel an agency to classify its documents, but does require consistent use of classification across an agency.

The Business Custodian must ensure the design of any classification scheme is appropriate, is kept current, and is used accurately and consistently.

Any classification scheme should be coherent (e.g. by subject, client, function) regardless of media. Linking related physical & electronic files will assist agencies to:

- Identify the location of documents in electronic and paper media
- Identify duplication in the system where it occurs
- Greatly simplify the development of retention schedules
- Link individual documents to the contextual information and business rules attached to a classification or physical file ( e.g. security, retention, owner, keywords etc).

Many business rules can be linked to a classification scheme. Documents can inherit the rules from the parent class, and so minimise data entry and improve consistency and accuracy.

#### **Rationale**

Consistent classification greatly increases the chances of finding documents by grouping them in a consistent manner, independent of storage medium. Linking documents to a classification scheme increases the probability of retention of documents for correct periods.



# Source

## **Standard**

### ***Prime authoritative data source location***

For data and documents created by business processes, the agency must clearly identify the prime authoritative data sources in which they can be found, which storage media are used, and how they can be retrieved.

#### **Supports Policies: *Source***

#### **Scope and Interpretation**

This standard relates to instances where multiple copies of data elements or documents exist in different locations. See also Standards: [Retrievability](#), [Replication](#), [Storage media](#). Knowledge of the physical storage media, location and available retrieval methods of prime data sources needs to be held by the Business Custodian. This is essential information to discover the true state of data source integrity.

Physical location may be a storage area, or a particular device or node on a computer network.

A prime source may be held outside the agency and be available under contract or legislation. In some cases a prime source may act as the source of replication to locations inside or outside the agency e.g. cabinet papers. In other cases multiple prime data sources, internal or external to the agency, may be used to produce a collated derived dataset.

When making an access request for personal information pursuant to the Privacy Act, the applicant should be provided with:

- information from the prime authoritative data source
- any replications of the prime authoritative data source within the agency that contain variations to the personal information (except where such disclosure is forbidden by law)
- certification that any non-supplied replications contain no variation from the original
- the name of any agency that has been provided with a replication of the data, or from whom the data was sourced.

#### **Rationale**

Identification of a prime authoritative source is a prerequisite for the application of all definition and integrity standards.

## **Standard**

### ***Synchronise data and document stores and publishing systems***

Agency publishing systems must manage any required linkage or replication from data or document stores, and ensure superseded material is available when required.

**Supports Policies:** *Source*

#### **Scope and Interpretation**

Agencies must have systems to manage:

- The synchronisation of published data elements and documents replicated between data and document stores and other publishing systems at regular intervals
- Maintenance of dynamic links from the publishing system to identified data elements or documents in data or document stores where this is a feature of the publishing system
- Retention of superseded published documents such as superseded public policy including the period for which they were current policy

Examples of electronic publishing systems are:

- Any agency intranet publishing system
- New Zealand Government On Line or an agency website linked to NZGO for policy available outside the agency
- Any web site that permits users to make enquiries of the agency's publicly available data elements or document stores

Synchronisation requires version control of current, superseded and draft documents, and of current and historic data elements. This also requires the retention of superseded documents. See also [Standard: Replication](#), [Retention Requirements](#)

Users go to the publishing system for the current versions of agency policies, and to either the publishing system or an approved document store to see the relationship of the current version to other versions of the policy. The Business Custodian must ensure systems are synchronised.

#### **Rationale**

Synchronisation ensures the publishing system is up to date, and is used as the authoritative source of official policy. Non-compliance with policy is minimised by maximising visibility and accessibility of policies and standards.

This also reduces the liability, and the risk of confusion over versions.

# Authenticity, Integrity, Retrievability

Standard

### ***Referential Integrity***

Agencies must put in place systems to:

- Maintain the context of data elements in database structures
- Maintain relationships between documents in a set

**Supports Policies:** *Authenticity, Integrity and Retrievability*

#### **Scope and Interpretation**

Within a database referential integrity implies that defined relationships between data elements and data structures are maintained when data content is added, updated or deleted. Rules to achieve this should be enforced by the database management system wherever possible. In addition, appropriate edit rules and data entry screens must be defined during system definition and enforced by the application system maintaining the data.

At a user level referential integrity is more likely to be maintained when data elements on screens are appropriately named. Staff training programmes must ensure that users understand the meaning of the data elements they work with as represented in the application system.

Documents in a set need to be retrievable as a group if required. This could be done through a classification scheme, metadata or simply by keeping the documents physically (or logically – say in a common directory) together.

In a document store referential integrity must be maintained within the metadata environment as with any database.

Aids to maintaining referential integrity include:

- Planned, thorough technical system and user testing of all new application code that could add or update data elements
- Adequate procedures to discover and recover from hardware or software data corruption
- Planned thorough technical evaluation of system and data or document management software upgrades before they enter production
- Protection of live data in the production environment through adequate security and migration procedures. The Physical Custodian must fully control access to production data and systems and the movement of data and software from testing environments to production.

#### **Rationale**

If referential integrity within a database breaks down then the data content quickly becomes unusable.

Loss of context within a document set or store will quickly lead to unreliable retrieval and the store will no longer be viable.

Any ambiguity about what a particular data element represents will result in unreliable data, since the meaning may change depending on who entered it.

## **Standard**

### ***Integrity of Application Software***

The integrity of any application software operating on approved data and document stores will be monitored at appropriate intervals, and action taken to repair and prevent defects.

**Supports Policies:** *Authenticity, Integrity and Retrievability*

### **Scope and Interpretation**

Agencies must be able to demonstrate that software operating on a data or document store can maintain the integrity of the contents including:

- Maintain data element content and interact with the database management system to ensure correct data-relationships are also maintained
- Implement business rules as defined by configuration ( e.g. security, retention)
- Retain profile data and history of the document object
- Retrieve data and documents accurately and consistently
- Report accurately and consistently

### **Rationale**

The agency must ensure that the software used to access and maintain prime authoritative data sources is adequately tested. Otherwise the integrity of the source will always be in doubt.

## **Standard**

### ***Integrity of configuration***

The configuration of each agency's approved data and document stores will be regularly reviewed and updated to ensure it still meets business needs.

**Supports Policies:** *Authenticity, Integrity and Retrievability*

### **Scope and Interpretation**

Configuration can apply to both manual and automated systems and includes:

- Adequate design for capturing metadata
- Consistency and relevance of classification
- Consistency and relevance of validation sets
- Business rules such as retention, disposal, access rights, security-classification and privacy
- Exceptions and any requirements for changes to business rules
- Verification at intervals as determined by the Business Custodian
- For automated system, adherence to industry best practice and government IT standards for technical configuration management of hardware and software

### **Rationale**

As business requirements evolve it is essential to regularly review the initial and ongoing configuration, both at business and technical levels, of the systems managing data and document stores. Otherwise the agency risks serious loss of integrity over time in its data and document assets.

## **Standard**

### ***Integrity of content***

To optimise the integrity of data and metadata:

- System data and metadata capture must be maximised, and user input must be minimised
- Where users enter data into a data store, or create metadata in a document store, validation at the time of input is required wherever practical
- Processes must be in place to monitor and correct errors in the data and metadata
- Any changes to the use of a data or metadata field design purpose must be agreed with the Business Custodian and documented and effects on downstream systems taken into account

**Supports Policies:** *Authenticity, Integrity and Retrievability*

### **Scope and Interpretation**

Systems must be designed to minimise user input, and maximise the validity of remaining user input. Examples include: use of data known to the system e.g. date and time of user input, use of standard pick lists, or data derived from user actions on the system.

National or international standard pick lists should be used where these exist e.g. list of countries, list of occupations etc. Currently recognised standards could be available via NZGO. Where an agency deviates from the standard list the additions and substitutions must be documented.

Document metadata validation should be by the use of pick lists or controlled vocabularies for metadata fields and by spell checking on fields where pick lists are not feasible.

The internal content of documents should be to standards required by the agency, e.g. any agency style guide, use of templates/layouts, correct spelling, use of terms and abbreviations etc. Users must spell check the contents of their documents, and agency workstations should all be set to a standard dictionary, ideally New Zealand English (or failing that, Australian English). Only official abbreviations or contractions should be used, from a discrete set maintained by the agency and/or an external government approved source. The gazetted list of New Zealand place names is available from Land Information New Zealand.

Care must be taken in the use of validation tools to ensure they are fit for use by those entering data. Public access via e-government initiatives may allow the public to enter data directly into some systems.

Quality standards must be regularly monitored for compliance, and updated as required. Each agency must develop or adopt data monitoring standards and obtain regular samples.

Changes to terminology should be recognised and incorporated into the data and document retrieval systems, so that retrieval requests are accurate and comprehensive.

### **Rationale**

Validation increases accuracy, reduces opportunities for errors, and improves retrieval and implementation of business rules such as security and retention. Unauthorised and undocumented alterations to the usage of data and metadata fields pose a high risk of corruption and loss of both current and historical data.

Consistent application of data quality standards will result in a high level of success when searching for and retrieving information from document stores. Similar benefits are gained when searching for or manipulating data in data stores, or interchanging data with another agency.

Use of standard pick lists ensures better opportunities for matching within and between agencies.

## **Standard**

### ***Integrity of Process***

Agencies must be able to demonstrate that their processes do capture required data elements and business documents, that business rules and standard operating procedures are in place for their management, and that they are implemented.

**Supports Policies:** *Authenticity, Integrity and Retrievability*

### **Scope and Interpretation**

This covers the whole broad area of defining standards, developing guidelines, providing training and demonstrating that staff members comply with those guidelines. See also [Standards: Skills and training](#), [Individual responsibilities](#), [Process maps](#).

### **Rationale**

In order to demonstrate the authenticity of data element or document content, agencies must be able to show that the processes involved in creating and maintaining content deliver to business and statutory requirements.

## **Standard**

### ***Skills and training***

Staff will be trained in their responsibilities when working with Crown data and document assets. These responsibilities will be written or referred to in key agency documents such as Job Descriptions and Performance Agreements, for staff at all levels.

**Supports Policies:** *Authenticity, Integrity and Retrievability; Document identification and capture*

### **Scope and Interpretation**

For each agency to implement effective data and document management, individual users must understand and accept responsibility for applying key principles, rather than just work on a “step-by-

Training on data and document management should be provided for existing staff and should be part of the orientation/induction process for new staff. Training includes:

- Reasons why data and business documents are important to the Crown
- Reasons why careful management is required
- The correct entry and usage of data in database systems
- The appropriate management of business documents
- Access to up-to-date manuals and guidelines
- Education in the importance of generic legislation e.g. the Official Information Act, the privacy Act, and legislation specific to the organisation
- Education in the importance of the Treaty of Waitangi, its relevance and practical application to the capture and management of data and documents.

The inclusion of data and document management responsibilities in documents such as Job Descriptions and Performance Agreements is designed to ensure all employees understand that this is a key component of their work. These responsibilities should also part of the “rules of the house”, meaning the core rules on how employees will conduct themselves in an agency. See also [Standards: Discovery implications](#).

### **Rationale**

Staff must have the necessary knowledge and skills to do the task. It is important to ensure that appropriate training and documentation are available when and where needed.

## **Standard**

### ***Version Control***

Agencies will determine business rules for version control of data elements, business documents, and data sets. Rules will be built into systems or expressed as guidelines for users.

**Supports Policies:** *Authenticity, Integrity and Retrievability; Auditability*

### **Scope and Interpretation**

Each agency must build guidelines on when versions should be created, and users must create versions according to those guidelines. The focus of the guidelines should be on capturing evidence of business transactions. See also [Standards: Integrity of process](#).

Versions of individual data elements are not necessarily held on operational systems however they may be retained in a time series within a data warehouse. At least some of the versions of a document that are generated during its development will normally need to be saved and identified as part of its history.

Data and document stores would not normally be versioned as a whole unless a migration to a different environment is required. Individual data sets may require a version history when used as a basis to generate other material e.g. a policy document.

Approved document stores must permit the user to generate documents that can be treated as versions of the same document. Each version will have its own unique identifier that might be generated or user entered free text. Users should create new versions of documents where there is substantive change to the document, typically linked to a defined process.

Some versions may become “read only”.

Guidelines will aid users in deciding when to create a new version. Guidelines will tell users whether the prime authoritative data source is to be kept in electronic or physical form.

Each agency must have available all required versions of a data element, business document or dataset, and the status of each version will be known. See also [Standard: Process maps](#).

### **Rationale:**

Version control assists historical evidence of business transactions to be maintained over time and is essential where this is a statutory requirement.

As data and documents are created and modified, the changes themselves provide additional evidence (e.g. development of policy, published version).

Reduces risks of wrong versions being referred to, and ensures correct versions can be identified and retained.

## **Standard**

### ***Document Templates***

Agencies will develop and maintain a set of standard document templates and styles for their document types.

**Supports Policies:** *Authenticity, Integrity and Retrievability*

### **Scope and Interpretation**

These templates will:

- Assist each agency to standardise information capture
- Be universally available and up to date
- Be clearly labelled

Each agency must ensure standard document templates are designed and made available, and audit their usefulness and usage. The focus is on frequently used documents like letters, memos, web pages. and distributed widely e.g. within the agency, to the public, to the web or intranet. The template can also address issues of suitability for use in different processes, eg the effect of shading/colour on scanning.

NZGO could be a repository for a set of generic optional templates which agencies could modify for their own use. Such templates would need to conform to open standards and not be restricted to proprietary software.

### **Rationale**

Templates ensure that standard business documents are quicker and easier to create and edit, whilst consistently conforming to agency standards. This and reduces the risk of inappropriately formatted documents in use, reduces the need for users to have advanced formatting skills, ensures consistent look and feel and use of styles, and improves recognition.

Templates ensure that documents are partially structured, and therefore improved for use by search engines and knowledge management tools.

## **Standard**

### ***Retrievability***

Data element content and business documents will be retrievable in formats that meet open international standards.

**Supports Policies:** *Authenticity, Integrity and Retrievability*

### **Scope and Interpretation**

Data elements and documents will be stored in an easily accessible format for their retention periods, and if necessary migrated to other storage media and/or software to maintain their accessibility.

Easily accessible formats conform to open standards and are software independent. Practically the choice will usually fall to current industry standards supported by multiple vendors to ensure contestability and future choice. See *E-Government IS Policies and Standards*.

Changes to low level formats controlled by operating systems and storage hardware are generally infrequent and can be managed through standard migration processes, provided open rather than proprietary technologies are used.

For electronic documents:

- Effective and efficient searching and retrieving tools will be incorporated as part of the approved document store design
- All documents saved into an approved document store can be retrieved through that document store, regardless of where they are currently stored (location on the network, on-line, near-line and off-line)
- Agency *rules* must control the retention of documents, even if processes external to the approved document store actually implement the deletion or transfer of documents to other storage media

Acceptable speed of response will be stated in agreements between the Business Custodian and the Physical Custodian.

### **Rationale**

There is little point in preserving data or documents if they cannot be easily retrieved in a usable format. The use of open industry standard formats for storage and retrieval mitigates the risk of expensive recovery being needed to retain the use of older information.



# Auditability

## **Standard**

### ***Change Control***

Change control procedures will be applied to the structure of data and document stores and the business processes that affect them, to ensure the contextual integrity of current content and that historical material maintains its integrity. Applications that create or maintain data or documents and interfaces to downstream systems e.g. a data warehouse, must be included in the change control process.

**Supports Policies:** *Auditability, Interchange, Replication, Interfaces*

### **Scope and Interpretation**

The change procedure must ensure that all parties involved, whether internal or external to the agency, are consulted appropriately. Analysis of system change must include all interfaces to both internal and external systems.

The Business Custodian will ensure that adequate change control procedures exist within the agency and that they are regularly audited for effectiveness. The procedures must cover both system and business process changes that affect agency data and document stores.

### **Rationale**

Uncontrolled changes to the structure of data and document stores or to systems and business processes that manipulate their content, has the potential to quickly destroy integrity. While change is often essential to meet business needs, it must be controlled to preserve existing investment.

## **Standard**

### ***Audit trail***

Agencies must have an audit trail where there is a statutory or business requirement for audit and monitoring of creation, update, deletion, and in some cases retrieval events, applied to data element content or document metadata and content.

**Supports Policies:** *Auditability*

### **Scope and Interpretation**

This standard is focussed primarily on monitoring permitted activity, so agencies must carefully analyse requirements to ensure all relevant areas are covered. See also [Policies: Access Rules](#) and related standards on the prevention of unauthorised access.

While changes to data or document content are usually the most critical events, in some cases agencies will also need to log the retrieval of designated sensitive content.

Logging systems must produce alerts when events outside the agency access rules occur. Over time, patterns of valid system use should be established so that significant variations can be recognised and investigated.

While this type of logging is easily applied to electronic data and documents, paper based systems would require elaborate security arrangements to achieve the same result. Costs for this could only be justified for highly confidential documents.

Staff need to know what activities are being monitored and regular reporting must be in place.

### **Rationale**

The existence of viable event logging, monitoring and reporting systems is a major deterrent to breaching security, and a major aid to determining accountability when it does occur. They deliver the ability to establish location, user and change/access event across the data and document asset base.



# Interchange, Replication, Interfaces

## **Standard**

### ***Interchange Agreement***

Agencies will define requirements and develop agreements before sharing data between primary authoritative data sources, whether those sources are internal or external to the agency.

**Supports Policies:** *Interchange, replication and interfaces*

### **Scope and Interpretation**

Interchange may involve data transfer/duplication or data access at a system level. It must involve two or more prime authoritative data sources and may be one or more data flows.

Where similar data is collected from various sources into a new combined data source, this is considered to be replication in the manner of a data warehouse. [See Standards: Replication.](#)

Ongoing interchange arrangements require a single agreement but regular reviews of its operation must be included. One time interchanges still require an agreement to ensure the process is documented.

Agreements will include a clear statement of objectives and requirements, and will detail:

- The legislative environment that permits and/or limits the interchange
- The purpose of the interchange
- What data is included in the interchange
- A review of how fit the data is for that purpose
- Quality requirements
- Security requirements
- A review schedule, if the interchange is repeated over time
- Notification of any changes to the either the primary authoritative data source or to the interchange use

Agreements between agencies require the approval of the Crown Data Steward. Agreements within an agency require the approval of the Business Custodian.

The Privacy Act applies to the interchange of personal information, particularly where the interchange involves data matching.

### **Rationale**

Interchange agreements within agencies, however brief, will ensure that all interfaces are known to the Business Custodian and documented accordingly. Externally the same applies, so that the Crown Data Steward is aware of all interchanges and all are adequately documented.

## **Standard**

### ***Replication***

Replication of data or documents will be controlled by the Business Custodian(s) involved and will only come from prime authoritative data sources. All replication arrangements will be auditable to ensure that a true replica is made.

**Supports Policies:** *Interchange, replication and interfaces*

### **Scope and Interpretation**

Permanent replication of data or document store content will most frequently take place for disaster recovery purposes or where data is to be included in a data warehouse.

In some cases data may be usefully replicated from many distributed sources into a single source, that can be used as an aggregate in the manner of a data warehouse.

Permanent replication of data to servers in different geographic areas may sometimes be justified as a technical solution where network bandwidth is short. This kind of solution should be transparent to system users. Data or document stores physically segmented and replicated on the basis of business organisational structures are to be avoided. The excessive complexity inherent in such systems puts government data and document assets at unnecessary risk.

Temporary replication may be used for data matching programs where a snapshot copy of the required data is used for comparison with another database, and then discarded. It may also be used to check out documents or data to a local system for off-line work. In these cases the data or documents in the prime authoritative data source must be locked from update until the changes are checked in.

Temporary replication is also commonly used to populate databases within test and development systems from the production base. This must not be undertaken unless access rules followed e.g. person data could be encrypted or scrambled. See [Policies: Access Rules](#)

Replication places responsibility on the Business Custodian to ensure checks are in place to monitor the accuracy of the replication. This may include the periodic restore of a database from tape backup, or regular comparison of databases where one is replicated online by the other. When two agencies are involved the Business Custodians will co-operate to ensure a consistent approach. See also [Standards: Interchange Agreement](#).

Wherever practical other methods of interchange should be used to avoid the build up of unnecessary duplicate data stores.

### **Rationale**

Replication dramatically increases the risks of: data corruption, confusion between data sources, use of incomplete data, security breaches, confusion over retention and consequent increased potential for data loss. Replication of data and document stores therefore requires constant monitoring to ensure an accurate result. It should only be undertaken to meet specific business requirements.

## **Standard**

### ***Interfaces***

Electronic interfaces between systems must use mechanisms based on open industry standards as specified in the government information technology policies and standards.

**Supports Policies:** *Interchange, replication and interfaces*

### **Scope and Interpretation**

Interfaces between systems need to be robust and exhibit the following characteristics:

- Use appropriate security for the data or documents being used
- Use standard communication protocols
- Use an encryption standard if a public network such as the internet is involved
- Not rely on creating duplicate data or document stores unless replication is the purpose of the interface
- Be transparent to system users
- Show that data and document integrity is retained

### **Rationale**

Interfaces based on closed proprietary standards are likely to be more expensive to run and are prone to becoming obsolete. Use of standard open protocols will normally allow a wide range of cheaper systems and equipment to be used.

Using standard mechanisms will increase the ability of individual agencies to respond to whole-of-government initiatives.

## **Standard**

### ***Migration***

Data and document stores will be constituted such that all content, structure and metadata can be migrated to a different environment without loss of integrity.

In the event of a migration or major upgrade, migration plans will be produced by the Physical Custodian and subject to approval by the Business Custodian.

**Supports Policies:** *Interchange, replication and interfaces*

### **Scope and Interpretation**

Each agency must retain the ability to upgrade or migrate to a new platform/operating system or new application, or even to remove its databases or business documents and their contextual information from an application/store, without loss of information. Where the material is held electronically, migration should operate on a copy of the original dataset, so that the original dataset remains intact.

Documents migrated out of the current document store or DMS must still be linked to their metadata, either stand-alone or in another application. An approved document store design must permit transfer of documents to another storage repository complete with their metadata, with or without the application managing the current document store or DMS.

For any upgrade, migration or transfer, the process report will include alterations to relationships and content etc, including:

- Original data
- What data has changed
- What data has been added
- What changes are agreed
- Who decides
- Back-up and roll-back plans
- Documentation on the original system
- What security has been applied

Each agency must have standards for data and document management, conversions, and migrations, that are sufficiently robust to ensure that integrity of data is maintained or enhanced through migration processes. Agencies should take all reasonable steps to ensure the future integrity and accessibility of data and business documents, whether or not those data or documents were created within an agency or inherited from another agency.

Agencies must assess and select software to meet these standards, and ensure that any data migrations meet these standards. This standard also applies to manual systems.

### **Rationale**

Migration of data carries with it the risk of loss of contextual information and/or change of some data or metadata and security permissions. Each agency needs to have in place standards and procedures to ensure that the integrity of data is maintained through system and/or platform migration.

This strategy assists an agency to be independent of any software provider. It reduces the risk of inability to upgrade software and of inability to access data or documents.



# Retention

## **Standard**

### ***Backup, Recovery & Restore***

All agencies will have a backup regime for data and document stores to insure against system failure or human error. Backup operations will be regularly monitored for completeness and tested for retrievability. The regime will be developed and actioned by the Physical Custodian, and subject to approval by the Business Custodian.

#### **Support Policies:** *Retention*

#### **Scope and Interpretation**

This standard applies primarily to electronic data and document stores. Paper based document stores could be backed-up by copying documents and transferring them to another site. This would only be practical in special circumstances.

#### ***Back-up***

Back-up schedules must be determined and applied to all data and electronic business documents held by government agencies. This will normally be done by specifying schedules for the various network database and application servers, and by ensuring that all data elements and documents fall to at least one of these.

The following sample schedule could be used for a network server:

- A full system and application software back-up to be stored securely off-site and updated at every system/application software change. The back-up must be re-done at least annually if there are no changes.
- Incremental data back-ups at least daily and more frequently if possible - normally stored on-site
- Full data back-ups at least weekly to be sent off-site maintaining at least 4 generations

With this scenario the maximum amount of data that can be lost by machine failure is 1 day. The maximum amount that can be lost by the destruction or loss of the site housing the system hardware is 1 week.

Where the prime source of agency data is held on an isolated PC e.g. field workers using portable computers, the user must back-up data and documents at least daily to a second drive or separate medium for as long as it remains the only copy.

The exact details of the back-up schedule for any one machine will be determined by the most sensitive and important data held on it.

The requirements for data back-up should be part of the agreement with the Physical Custodian.

#### ***Recovery***

Data available on-line and being constantly updated should not be totally reliant on recovery from back-up, except in cases of catastrophic failure. Restoring a back-up is not the same as recovery. Restoring a back-up will result in data loss and cause expensive re-work as well as errors. Database management software that allows automatic recovery from temporary system or software failures should be used wherever possible.

Back-up of network file systems should operate at an appropriate frequency and level of granularity to meet business requirements for recovery of single files, Emails etc. with minimum data loss.

### *Restore*

The actual recoverability of each grade of back-up must be periodically tested in case it turns out to be unrecoverable when needed. Both the data and necessary support systems must be fully recoverable at an independent site from that housing the normal production systems. The Business Custodian will need to determine how much data input or document development work the agency can afford to lose in a disaster that would require a full restore. One day is a typical choice for cost effective tape based systems. See also [Standards: Disaster recovery](#).

### Rationale

Electronic storage has become more and more reliable as technology has been refined, however the cost of back-up is far outweighed by the cost of losing data and documents essential to the functions of the agency.

## **Standard**

### ***Storage Media***

Data and document storage will conform to the following standards:

- Storage conditions for electronic data or documents will conform with government Information Technology standards
- Physical storage conditions will conform with National Archives Storage Standard NAS 9901

### **Support Policies:** *Retention*

#### **Scope and Interpretation**

Storage technology is constantly changing. Examples include: magnetic storage – video, audio and data tape, disk etc., CD, microfilm, paper with chemical inks or fused toner etc. While the agency may reproduce data or documents in a variety of media for operational purposes, it must actively choose media for retention. The decision to use a particular technology will normally be a trade off between price and requirements, but it must be fit for the business purpose:

- Online and near line storage must meet system performance requirements in terms of access, integrity, and failure rates laid down by the Business Custodian
- Off line storage media must meet business requirements for access and retention
- Where a medium, e.g. magnetic tape, is chosen for back-up or off line storage, rules must be in place to ensure that data is refreshed at a predetermined frequency. In the case of magnetic tape archives an annual refresh is usually advised.
- Agencies must monitor the processes that place data and documents into the medium chosen for the prime source and ensure that they are working. For example, processes that rely on individuals to “print and file” are especially risky and need close monitoring.

#### **Rationale**

Adequate storage provisions are essential to protect all Crown data and business document assets, whether it be specialised storage for paper documents or software and hardware for material in electronic format.

There is a risk that context and content can be lost in transfer to another storage medium, unless the transfer is correctly managed and audited.

Standard

### ***Disaster Recovery***

All agencies will have a fully tested disaster recovery plan to reconstitute data and document stores to ensure timely re-establishment of the business. Plans will be produced by the Business Custodian and subject to agreement by the Crown Data Steward.

#### **Support Policies:** *Retention*

#### **Scope and Interpretation**

Disaster recovery planning is a component of business continuity planning and includes strategies to ensure:

- Back up of all electronic data and document stores to a standard that permits their reconstitution. See also [Standards: Backup Recovery and Restore](#).
- Restoration of data and document stores and their operating software to established time-frames in the event of a disaster
- Vital records (see [Glossary: Vital Records](#)) will be identified in approved data or document stores.
- Vital data and vital document stores to be restored in time-frames to permit the rapid re-establishment of the business of the agency. Where data and documents are stored only in a physical medium such as paper, consideration must be given to duplicating vital records.
- Minimal damage to or loss of irretrievable physical material. See also [Standards: Storage Media](#).
- Users and technical staff are trained to required levels of expertise and available to respond in the event of a disaster
- Hardware and software required to support a restored system are in place or will be available
- Testing schedule to ensure plans are regularly exercised and assumptions are challenged. A full test must include full restoration at the disaster recovery site and simulation of total expected usage at that site.

#### **Rationale**

In the event of a system failure each agency must be able to retrieve or recreate data and documents relating to its core activities.

Many agencies are now almost totally dependant on computer systems to continue everyday business. Without a tested disaster recovery plan, government agencies place their functions at risk of very serious disruption during an emergency.

Identification of vital data and documents permits an agency to focus its main effort.

## **Standard**

### ***Retention requirements***

Agencies must identify, describe and comply with their retention requirements for data elements and business documents.

### **Support Policies:** *Retention*

#### **Scope and Interpretation**

This applies to all data elements and business documents owned by the Crown.

Each agency must:

- Clearly identify retention requirements, including access rules over time as stated in government legislation
- Clearly identify any additional business needs for retention
- Document the retention rules for the types of data elements, documents and data stores created in the course of an agency's business activities and work with National Archives to develop retention and disposal schedules.
- Monitor compliance with identified retention requirements
- Review retention requirements when changes to the agency functions, structure, classification, technology or legislation occur.

Each agency must have a clear record of its official data and documents, and must apply policies and business rules to retain them for defined periods, without duplication of effort across systems. Access restrictions may change over time, and this information must be included in the retention rules.

The Business Custodian will define relationships between retrieval requirements and systems for off-line or remote storage, and will negotiate these with the Physical Custodian. Practically, items will usually be moved from operational systems in a related group e.g. a client record, project documents etc. However it may be found that sub-groups of data elements may be usefully moved separately - for example parts of a client record relating to a specific business area may be required on-line for a limited period only, while other parts may need much longer retention.

Retention and disposal schedules must meet the criteria stipulated by National Archives. These criteria address three main issues:

- How long to keep what material
- Who has custody
- Who has access
- Retention and disposal schedules must be authorised by the Chief Archivist. Data and documents may only be disposed of with such approval

#### **Rationale**

Agencies must comply with the statutory requirements in the Archives Act that only the Chief Archivist can approve the destruction of Crown material. The most efficient way to achieve this is via a Retention and Disposal Schedule.

Rules based retention ensures:

- Data and documents are kept in a systematic way and are available for appropriate periods
- Timely access to accurate information
- Data and documents are managed in accordance with their statutory, evidential, and business value
- Access restrictions are administered over the retention period

- The appropriate policies, business rules and operation standards are demonstrably applied to official data and documents. Removes the risk of breaching statutory requirements including the Archives Act
- Enables systematic identification and prompt destruction for those no longer needed.

## **Standard**

### ***Transfer between Agencies***

Where all or part of a data or document store is to be transferred between agencies, those agencies must develop an explicit agreement for that transfer. With the exception of transfers between an agency and National Archives, the explicit agreements will be between agency business custodians and will be ratified by the Crown Data Steward.

#### **Support Policies:** *Retention*

#### **Scope and Interpretation**

Transfer in this context means the complete removal of material from one agency to another.

As agencies acquire or discontinue functions, datasets related to those functions may be transferred to another agency. The agencies concerned must draw up an agreement detailing:

- Which agencies are involved in the transfer
- What is to be transferred – all or part of the dataset
- What if any part of the dataset is not to be transferred, and which agency has responsibility for it
- Any existing access restrictions including security-classifications
- Any existing Retention and Disposal Schedules
- Any applications required to maintain the integrity of dataset

The transferring Business Custodian will also inform National Archives of the transfer.

Where datasets are no longer in operational use, the agency may negotiate either retention within the agency or transfer to National Archives. Only material approved by the Chief Archivist as requiring retention may be transferred to National Archives. Transfers to National Archives must be to standards set by National Archives.

In the past, most agencies have transferred paper documents to the physical custody of National Archives when several years have passed since they were in operational use, but with electronic material that is readily reproduced this paradigm can be reviewed. Data and documents do not have to be transferred to the custody of National Archives, and each agency can negotiate to retain custody to an agreed standard of storage and access. Agencies may wish to negotiate which agency has the responsibility for retention, whichever medium is used for storage

#### **Rationale**

Ensures that the Crown is aware of where its datasets are and who has responsibility for them.  
Minimises the risk of orphan datasets.

Data and documents will be held in the physical custody of whichever agencies are best able to meet the business needs and regulatory requirements for their retention. Each agency must negotiate retention with National Archives.

## **Standard**

### ***Destruction Protocols***

No data or business documents will be destroyed while they are needed to fulfil the statutory or business requirements of the crown:

Any deletion or destruction process must be secure, deliberate, authorised and auditable.

### **Support Policies:** *Retention*

#### **Scope and Interpretation**

Each agency must at all times be able to identify the existence and status of documents, and, where applicable, demonstrate accountability in the deletion or transfer of those documents. The deletion or destruction process will be auditable, and there must be no unauthorised deletion of data or documents.

- For any processes in an automated system that permit the deletion or overwriting of data elements, agencies must validate those processes against retention and disposal requirements and against the risk of accidental destruction
- Agencies must provide advice to staff on what document types do not require retention of non-substantive versions
- Data and documents may only be deleted or destroyed in accordance with Retention and Disposal Schedules agreed with National Archives and according to business rules. See [Standards: Retention requirements](#).
- Responsibility for deletion of data and documents rests with the Business Custodian
- Data and documents must not be automatically destroyed without review
- Staff searching for documents that have been destroyed should be able to identify the fact of and the reason for deletion of documents from the system
- For electronic documents, deletion will comply with agency policies and standards for secure destruction of electronic data
- The Privacy Act applies to data used in authorised data matching programmes.

#### **Rationale**

Agencies may not destroy Crown data or document assets without agreement from National Archives. A formal destruction process will ensure legislative requirements are met, e.g. ensure that the reason and authority for destruction are available.

Auditable destruction of data and business documents:

- Demonstrates accountability in managing Crown assets
- Saves time wasted searching for documents which no longer exist
- Satisfies Ombudsman's Office requirements for declining requests for official information