the national archives

Generic requirements for sustaining electronic information over time:

3. Sustaining authentic and reliable records: technical requirements

TABLE OF CONTENTS

1 Introduction

1.1.1 Sustainable records are defined as those electronic objects and their concomitant metadata which defines them as records, which require continued retention by the creating or owning organisation until such time as the records can be destroyed or, where that is warranted, passed to a specialist archive for permanent archiving. If records are to be sustained there must be confidence that the maintained records possess authenticity, reliability, integrity and usability.

1.1.2 This document is intended to provide the key technical requirements needed to specify and implement a sustainable solution for electronic records. This section of the generic requirements will define technical requirements as opposed to the management requirements which are described in Volume *2* of the *Generic requirements for sustaining electronic information over time - Sustaining authentic and reliable records: management requirements.*

1.1.3 Its initial focus is on electronic objects in document form, which will normally be located within folders displayed within a corporate classification system. It is assumed that such objects will either be imported from an unstructured environment into an electronic document and records management system (EDRMS) or will have been created and captured within such an environment.

1.1.4 This *document* should be read in conjunction with the other three accompanying volumes in the series on *Generic requirements for sustaining electronic information over time*. The titles of the other 3 volumes in the series are:

1. *Defining the characteristics for authentic records*
2. *Sustaining authentic and reliable records: management requirements*
4. *Guidance for categorising records to identify sustainable requirements*

1.1.5 Volume 1 provides a summation of the principles and attributes, which would support an attestation of authenticity and integrity and which need to be maintained as part of the electronic record in accordance with *BS ISO 15489 Information and documentation – Records management standard*. Volume 4 provides high-level guidance for departments seeking to categorise their records to scope the specific nature of the requirements needed to sustain these record categories as authentic records.

1.1.6 These generic requirements are not a full specification. They form a baseline, which sets out, - the minimum necessary to maintain credible electronic records which will continue to possess the attributes of authenticity and integrity over time. They also should be read as an accompaniment to the Functional Requirements for Electronic Records Management Systems 2002 revision: final version which are available at:

http://www.pro.gov.uk/recordsmanagement/erecords/2002reqs/default.htm

1.1.7 The guidance is intended primarily for those working in central government; the principles will also be relevant in local government and throughout the

public sector. Throughout this document the term ‘department’ should be taken to apply to any public sector organisation, including all departments, agencies and other organisations across government. Familiarity with the concepts of records as used in central government is assumed.

1.1.8 Each government organisation wishing to make use of these requirements, as a baseline or benchmark, will always need to consider its own specific business needs and context in determining its own requirements. These generic requirements must be tailored by: adding specialist business needs which are not covered at this generic level, selecting from alternative requirements according to corporate policy and practice, assessing whether any requirements listed in these volumes are highly desirable as opposed to mandatory for their own context

1.1.9 The generic functions described in this document may also be relevant to a permanent archive but the needs of archival preservation are considered as distinct from those operations required to maintain electronic records for continuing business needs even where the overall retention period may last for some decades.

2 Ingest - Importing across platforms

2.1.1 Electronic documents and records will require to be exported and imported across platforms. In order to achieve this effectively the system must ensure that objects are uncorrupted copies and where appropriate the pre-existing access permissions and other record management metadata applied to the folders and documents, contained within the exporting application, are mapped to provide the same level of functionality upon ingest into the importing application. Any failure to copy objects or map such functionality should automatically display a warning on screen and generate an exception report detailing the specific exceptions. The points listed below breaks the required information into specific categories that need to be considered when planning an export/import exercise.

- Content Information - the information that requires preservation.
- Preservation Description Information (PDI) - any information that will allow the understanding of the Content Information over an indefinite period of time.
- Packaging Information - the information that binds all other components into a specific medium.
- Descriptive Information - information that helps users to locate and access information of potential interest.
- Import requirement into the new platform
    - Acquisition and capture
    - Record aggregate requirement (as defined in the RM metadata standard

This will define how the original order of records is to be respected in the physical or logical structuring of sets or archival aggregates of records, and how they are to be presented for use.

2.1.2 The system must also be able to ingest or import records:

- in their native format, or a current format to which they been migrated and in order of preference;
    - an XML format which falls within the UK e-GIF framework, where possible
    - a rendition which is consistent with the range of formats currently specified in the e-GIF set, where an XML format is not available.

2.1.3 The system must also support the storage and management of schemas and style sheets required for rendering into the required format.

2.1.4 Upon import the system must accept copies of imported digital records together with, or separately from, their metadata, in “as-received” form – that is, unprocessed save for the allocation of a simple unique temporary component identifier (e.g. a sequential number).

2.1.5 Upon request the system must generate reports of records received, listing as-received components received for storage for periods of time defined by the administrator.

2.1.6 Ideally the system should have the capability to apply controls sufficient to guarantee an uncorrectable object error rate of no more than 0.000001% per year (1 error per 100 million objects per year). However it is not practical to be so specific about the maximum uncorrectable error rate as the error rate is entirely a factor of the error correction system used by a particular media type, and so varies considerably between types. For example, SuperDLT provides much lower uncorrectable bit error rates than CD-R. The custodian must determine whether the maximum uncorrectable error rate available is acceptable given the nature of the information being ingested.

2.1.7 Where non-standard metadata is present that is not defined in the XML schema (e.g. user defined), the system must be able to import, in bulk, electronic records in their existing format and their associated metadata by providing facilities to map the non-standard imported metadata to appropriate new elements.

2.1.8 The system must be able to import, in bulk, existing electronic documents that have no associated metadata presented separately from the document content. This should be achieved by automatically extracting metadata from the document where possible.

2.1.9 The system must provide facilities for managing the addition of missing metadata and the assignment of documents to folders by placing documents for further processing in queues and supporting the subsequent declaration of documents from processing queues by either a manual or an automated process.

2.2 Input Reconciliation

2.2.1 Input reconciliation is critical to the verification of the validity of electronic records. The requirements need to identify the key elements and performance measures for the underlying organisational policy, which determines the basis for sustainability. It will include standards and specifications for acceptable and unacceptable deviations from standards, such as when records that should be in an imported transfer into the sustained environment are missing or when information that should accompany the transfer is missing, inappropriate or unclear.

2.2.2 If required the system must be able to accept electronic objects separately from their metadata; associate these; reconcile and report any inconsistencies (e.g. missing or repeated objects or metadata).

2.2.3 The system must include further controls to ensure that all electronic objects and metadata expected are received and successfully imported.

For example, checking against transmittal a notice or manual checks of number of physical media.

2.2.4 The system must include facilities to allow the service provider's operational staff to remedy any inconsistencies reported as a result of 2.2.1 and 2.2.3 above.

2.2.5 The system must automatically generate a report of any uncorrectable error within one working day of its discovery.

2.2.6 The system must notify the administrator that any error has been detected and that a report has been generated.

3 Storage management

3.1.1 Storage management will need to reference existing standards and best practice but the critical factors, which will require definition, are as follows:

- determine how to choose the appropriate storage method (e.g. magnetic, optical, type and format of tape, or disk, on-line, near-line or off-line etc.)
- clarify appropriate environmental storage conditions and methods of carriage when transported or transmitted
- determine appropriate regimes and triggers for media migration to avoid obsolescence or degradation
- define monitoring mechanisms, timescales and performance measures to assess if data is still readable
- determine minimum information levels to be captured within the management audit trail for each process
- define the elements for disaster contingency management and recovery
- determine the scale and nature of the back-up regime
- monitoring strategies and mechanisms
- hardware/software performance monitoring criteria
- error correction standards (see also paragraph 2.1.6)
- monitoring of media to identify potential degradation
- media refreshment regimes
- media migration regimes
- modes of retrieval
  - render options
  - render management
- Evaluation of execution of strategies

3.1.2 It should be noted that retrieval can be further decomposed by transformation of the bit stream into a rendered object and evaluation mechanisms and performance measures for rendered objects should be determined and monitored.

3.1.3 As-received records must be protected against media degradation, according to the requirements in section 3.3 below.

3.1.4 The system must be able to retrieve components (and any metadata associated with them), using the unique temporary component identifier. This must include the ability to retrieve individual components or components with unique temporary component identifiers falling in a given range.

3.1.5 Deletion must not occur without a specific auditable instruction.

3.1.6 The system should provide on request reports listing as-received record components received, but not deleted, older than a specified threshold.

3.1.7 The system must allow users to delete objects, subject to the controls in 3.1.9 below. Deletion in this context includes deletion of all on-site and off-site copies.

3.1.8 If an object is stored on write-once media, deletion of the index for an object will be acceptable as deletion of the object.

3.1.9 The following control must apply to changes to unique identifiers and deletion of electronic objects:

- this action will not be initiated without confirmation from an authorised administrator
- the system will retain these a record of the confirming instruction
- the system will generate a report the changes and deletions at intervals to be determined by the system administrator
- the system will keep a copy of its records of changes and deletions, in a manner and for a period to be determined by the client

3.2 Management of Back-up and security copies

3.2.1 Where required the system must have the capacity to maintain at least two copies of each object and its metadata.

3.2.2 The system must have the capability to maintain at least two further copies of each object in secure off-site storage.

3.2.3 The system must routinely back up index data for the objects being preserved, and must have the capability to maintain at least two backups of this index data in secure off-site storage.

3.2.4 The system must check each object stored on-site for accuracy (against one of the other stored copies) at least once every six months or at a shorter interval defined by the system administrator.

3.2.5 The back-up procedure should establish a regime to rotate off-site copies to ensure consistency in checking to the level described in paragraph 3.2.4 above. The interval between such checks should not be less than that advocated in paragraph 3.2.9 below.

3.2.6 If the check detects an error, the system must replace the erroneous copy with a correct copy, using the off-site copies as necessary, while ensuring that at least one copy remains in the secure off-site store at all times.

3.2.7 If a new copy of an object is produced, in response to an error detection the system must update the new object's metadata to reflect its creation; and must update any objects, which referred to the earlier manifestation.

3.2.8 It should be noted that the metadata elements to be updated when a new copy is produced include history elements and elements which relate different manifestations of the objects.

3.2.9 The system must check every on-site and off-site storage volume for readability at least once every year.

3.2.10 If the check for readability detects an error, the system must:

- create a new volume containing correct copies of the information on the failing volume
- verify the correctness of information on the new volume
- dispose of the failing volume

3.2.11 Backup copies must be kept on a separate site, located to minimise the risk that the backups could be endangered by any disaster, which would endanger the main system.

3.2.12 Ideally the backup copies should be kept on a site at least fifty kilometres from the main system but where this is not practicable the back-up copies should be stored on a geographically separate site from the on-line system.

3.2.13 The backup copies must be stored, and moved between the system and the separate site, so as to protect them to the highest security classification applied to the records stored on the back-up copies and comply with data protection legislation.

3.3 Avoidance of Media Degradation the role of Media Refreshment and Migration

3.3.1 To avoid loss or corruption of the records through degradation of the storage media over time it will be necessary to establish a media refreshment regime which will involve re-writing the records to the same media type required by the storage strategy to ensure continued readability. This needs to be undertaken at regular intervals in accordance with the timescales determined in the storage strategy. These intervals should not however exceed the periods recommended by the manufacturers of the media for the refreshment of that type of media

3.3.2 When it is determined that the storage media currently used by the custodian to hold the records is no longer an appropriate storage medium (e.g. the existing media is considered to potentially obsolescent) a media migration should be established. Media migration differs from media refreshment in that the records are re-written to a different storage media from the one they were previously stored on. The new media type will have been identified as an appropriate replacement by the storage strategy. Following a successful media

migration a new media refreshment scheme must be established and maintained as described in the above paragraph.

3.3.3 The system should monitor the use of storage media, flag reminders to the system administrator and copy objects from any media, which are approaching the end of their anticipated life to fresh media.

3.3.4 The system should also monitor media for degradation to identify any deterioration that may have arisen during its active live.

3.3.5 When identifying the appropriate new media format, the following factors should be considered:

- longevity:- the media should have a proven life span of at least 10 years. Greater longevity is not necessarily an advantage, since technological obsolescence usually precedes physical deterioration of the medium.
- capacity: - the media should provide a storage capacity appropriate for the quantity of data to be stored, and the physical size of the storage facilities available.
- viability: - the media should support robust error-detection methods for both reading and writing data. Proven data recovery techniques should also be available in case of data loss. Media should be read-only, or have a reliable write-protect mechanism, to prevent erasure and maintain the evidential integrity of the data.
- obsolescence: - the media and its supporting hardware and software should be based on mature, rather than leading-edge technology, and must be well established in the market place, widely available, and based upon open standards.
- cost: - the total cost of ownership should be affordable. This should include not only the cost of the actual media (calculated as a price per MB), but also of purchasing and maintaining the necessary hardware and software, and of any storage equipment required.
- susceptibility: - the media should have a low susceptibility to physical damage, and be tolerant of a wide range of environmental conditions without data loss. Magnetic media should have a high coercivity value (preferably in excess of 1000 Oersteds), to minimise the chances of accidental erasure. Any measures required to counter known susceptibilities should be affordable and achievable.

3.3.6 The refreshment mechanism must allow verification of the copying process through a bit-level comparison between the source and target versions of each file copied.

3.3.7 The system must dispose of any failing or ageing media in a managed and secure fashion so that:

- the media will not normally come into the possession of any unauthorised third party;

- in the event that they should come into the possession of any unauthorised third party, the media should be overwritten so that no information can be retrieved.
- a record of the event and of the method used to overwrite the media to be disposed should continue to be held on the system

3.3.8 Where the information is considered to be particularly sensitive or it is subject to a national security classification over-writing of the media is not considered an adequate solution in itself. The media should be physically destroyed as part of a controlled process. The Ministry of Defence has established procedures for the destruction of such media and departments are advised to seek the advice of their departmental security officers before any decision is taken. Guidance on when to apply this process can be obtained from departmental security officers or from the Communications-Electronics Security Group (CESG)

4 Software File Format Obsolescence

4.1.1 In future it is likely that the software formats of some or all objects being preserved will become obsolete. As the sustainable systems evolve through new generations of hardware and system software, it may become impossible, or undesirable, to retain the objects in their original formats; in this event, it will be necessary to take other steps to ensure their preservation.

4.1.2 The techniques for such medium-long term preservation are not mature. It is likely that they will include mass format migration; they may also include the insertion of emulation software, and they may also include a move to "bundling." Given this uncertainty, supporting requirements cannot be specified in detail. This section is therefore limited to an overall requirement.

4.1.3 The system must contain no features, which would compromise maintenance of stored objects without changes for an indefinite period.

4.1.4 The system must not discard records after migration to a new format without a specific authorisation from an administrator possessing the appropriate access permissions. The reasons for such actions citing the appropriate authority to proceed should also be formally recorded.

4.1.5 In some situations, the organisation will wish to preserve the records in both the newly-migrated format and the older or original format in order to demonstrate and track the level of information loss in the newly migrated format. The system must document and retain details of any information loss incurred by any process such as migration.

4.1.6 Whenever any action is taken which changes an object in any way (such as a migration), the system must record this change in the appropriate metadata element(s).

4.1.7 The metadata elements referred to in 4.1.6 include both history elements and elements which relate different manifestations of objects.

4.2 Management of format conversion and renditions

4.2.1 The system must be able to convert objects into a preferred sustainable or interoperable format at any point in time after importation importing if they are not already in the designated format.

For example,

- *converting thousands of single-page TIFF images making up hundreds of inquiry documents into multi-page TIFF or PDF format;*
- *converting a file of mixed Word documents, Excel spreadsheets, e-mail messages into XML format.*

4.2.2 The system must associate copies of different formats of the same object, preserving each separately while retaining the association between them.

For example:

- *some surrogate images may be preserved as both TIFF and JPEG images or MS Word and XML*

4.2.3 The system must be able to import objects, which are related to objects already imported; and must, in this case, update the metadata of all relevant objects to reflect the correct relationships.

For example:

The system may have to import a redacted instance (e.g. where a decision has been taken to release a record under Freedom of Information (FOI) but that certain specific elements are to be withheld due to their personal sensitivity). In this case the system must update the instance's metadata with information about the original object; and must update the original object's metadata with information about the instance.

4.2.4 The system may have to import a part of a record imported previously. In this case the system must update the part's metadata with information about the record; and must update the record's metadata with information about the part.

4.2.5 At time of import, the system must deduce and store sustainable metadata from the objects being imported.

4.3 Management of relationships between copies of the same object in different formats

4.3.1 The system must not discard records after migration to a new format without the authorisation of the record manager.

In some circumstances it will be necessary to maintain copies of the records in both the newly-migrated format and the older format

4.3.2 Whenever any action is taken which changes an object in any way (such as a migration), the system must record this change in the appropriate metadata element(s).

The metadata elements include history elements and elements which relate different manifestations of objects.

5 Reproduction of electronic records

5.1.1 Any solution will have to deliver or reproduce copies of records upon demand to authorised users in a form that meets the business requirement. In order to do this the following elements need to be clarified:

- clarify how to present copies whilst safeguarding the “original” components
- determine the information flows that need to be captured in the management audit trail when copies are rendered for viewing
- define when to apply appropriate certification or authentication mechanisms if required e.g. watermark or digital signature attesting to the authenticity of the content

5.1.2 The system must generate a report of why a request for a record and/or information about a record could not be satisfied in whole or in part.

5.1.3 Functional requirements will need to articulate the services that would define a compliant reproduction and presentation system and the criteria against, which the outputs could be evaluated.

5.2 Authentication mechanisms

5.2.1 In certain circumstances it will necessary for departments to provide copies of sustained records together with a certificate or attestation of authenticity that one or more records are authentic. Logically this would be undertaken by the person or persons responsible for the active maintenance of the sustained records and could take the form of a document, an attachment, or an annotation, which attests to the authenticity of one or more records.

5.2.2 In order to determine the basis of authenticity it will be necessary to identify the information that indicates whether records can be considered as authentic. This will have to be founded on the basis of how the records creator addressed the requirements for authenticity up through the time when the records were imported into the sustained environment. Alternatively authenticity could be verified through corroborating evidence. Where records were stored or created within an EDRMS environment this can be in addressed by the record management metadata held within the XML schema.

5.3 Export Requirements

5.3.1 Each specified collection of digital objects or records will require their own subset of the generic requirements that appear here. The requirements encompass both how the records will be written in physical and logical files both for transfer and for storage to produce requirements for physical and logical files.

5.3.2 The system must be able to retrieve and export on agreed removable media or by network one copy of all objects stored for any specified collection or series

in response to a single request, exporting them in digital form, together with all their metadata and (at the administrator's option) audit trail data.

5.3.3 The choice of media or network, and the formats are to be agreed at the time of the request for an export.

This requirement implicitly includes export of all the records and metadata in the system although the choice of media and format at time of export would obviously be limited by the chosen system design.

5.3.4 The system must also be able to export records:

- in their native format, or a current format to which they been migrated and in order of preference;
    - an XML format which falls within the UK e-GIF framework, where possible
    - a rendition which is consistent with the range of formats currently specified in the e-GIF set, where an XML format is not available.

Such renditions may be achieved by:

- capturing an appropriate rendition as part of the record capture process
- rendering the record as part of the export process
- exporting directly to another package which is capable of rendering the record within a controlled environment (e.g. to PDF).

5.3.5 The system must also support the storage and management of schemas and style sheets required for rendering.

5.3.6 Where an appropriate XML format is not available the system must be able to export electronic records in the native format, or the migration format currently stored in the host system

5.3.7 Where an appropriate XML metadata schema exists the system must have the capability of supporting the schema to permit the export of metadata in accordance with the schema.

5.3.8 The system must be able to export all types of records, which it is able to capture, regardless of the presence of the generating application software.

5.3.9 In addition to the export of record management metadata in XML the system should support the mapping and configuration of metadata from the existing scheme into the scheme used by the target system. This should be done by creating and exporting formatted XML files to which an appropriate XSL style-sheet has been applied, thus enabling the transferred metadata to be viewed externally from the exporting platform in a manner which either maintains the display provided on the browser of the exporting system, or in a form which can be interpreted by users who have little, or no familiarity, with the exporting system.

6 Security

6.1.1 The system must as a minimum provide the same capability to specify and allocate access permissions as required by the *Functional Requirements for Electronic Records Management Systems 2002 revision.*

6.1.2 The system must have an overall security capability to meet the information security requirements required by *BS ISO 7766.*

6.1.3 Any object within the database should have the capacity to have an individual access control protocol assigned to it.

6.1.4 The systems should allow the system administrator to create of any number of roles to which specific access permissions can be allocated along with any subset of administrative privileges for any one object or group of objects.

6.1.5 The system must be able to store objects classified up to highest security classification applied to the records held within the system. The design and operation of the system is to follow normal UK government guidelines for this classification.

6.1.6 The system must protect objects containing personal data consistent with data protection legislation.

6.1.7 The system must have the capability to interface with applications which have lower security levels, maintaining its security level at all times.

6.2 Audit controls

6.2.1 The system must maintain automatically an audit trail of all actions carried out on all objects. Actions are to include, but need not be limited to

- import processes;
- migrations;
- replacement of corrupt copies;
- changes to metadata

6.2.2 The system must maintain an audit trail of all changes of system configuration or metadata configuration.

6.2.3 The system must store its audit trail securely in a manner which ensures it cannot be changed or deleted.

6.2.4 The Service Provider must store the audit trail information including the audit trail of migration for at least as long as the objects to which it refers. It may be necessary is some instances for the audit trail to be preserved in perpetuity.

6.2.5 The system must provide, on request, audit trail listings to show, for a specified time interval:

- all actions affecting a specified object;

- all actions affecting the system;

in a form which can be interpreted by management and external legal advisors or auditors who have little or no familiarity with the system.

6.2.6 The system must store audit trail data in an XML format which falls within the UK e-GIF framework, where possible and designed so that the audit trail data can be exported and preserved in future.

6.2.7 In addition to the export of audit trail data in XML the system should support the mapping and configuration of audit trail metadata from the existing scheme into the scheme used by the target system. This should be accomplished in same manner as the export of record management metadata as described in paragraph 5.3.9 above by creating and exporting formatted XML files to which an appropriate XSL style-sheet has been applied, thus enabling the transferred metadata to be viewed externally from the exporting platform in a manner which either, maintains the display provided on the browser of the exporting system or in a form which can be interpreted by users who have little, or no familiarity, with the exporting system.