Proposal for CHM Management Objectives for Software Preservation and a Software Repository

H.M. Gladney hgladney@pacbell.net
HMG Consulting
Saratoga, CA 95070

1 February 2005

Introduction

The Computer History Museum Software Collection Committee has started to discuss a long-term digital repository for its eventual collection of historical software.[1] To expedite the discussion of service requirements, it seemed efficacious to find and modify the requirements analysis work of some other institution with similar needs. For this I chose a British analysis which is described with:

> Our aim is to give the best possible advice and guidance to all who share our concern for the care and preservation of records and archives. To support this advice we promote accepted archival standards, and have published the National Archives Standard for record repositories to guide our own advisory and inspection services. The Standard is the recognised benchmark on caring for records and providing access to them. [PRO]

Part of the background is that long-range digital preservation is widely regarded to be an applied research topic. In contrast, digital repository technology is understood very well and represented by many high quality offerings.

Towards a Requirements Statement for a CHM Software Repository

Most published documents dealing with preservation repository requirements are written from the perspective of repository managers. I believe that we will obtain better results if we analyze requirements from the perspective of eventual users. This, together with what can be inferred from the following figures, suggests a taxonomy for user-oriented software technology requirements analysis, a topic which the current document takes up on page 36. The figures suggest aspects for discussion.

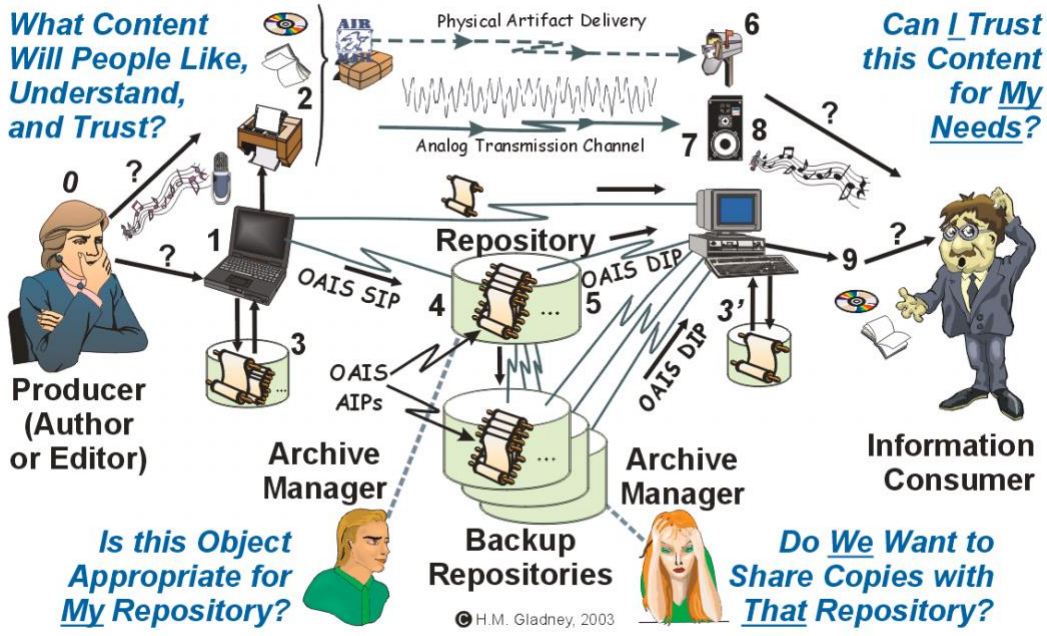

Figure 1: A model of digital communication,
suggesting human questions and alternative pathways by which a document might move from its producer to its eventual consumer. The numbers suggest copies that might be transformations of each other. The figure also suggests how easy it is to achieve [OAIS] conformance.

[1] In the current document, the word 'software' is extended to encompass all informational artifacts that are not computer hardware. I.e., it includes not only representations of computer programs, but also related documentation, manuscripts, pictures, and books.

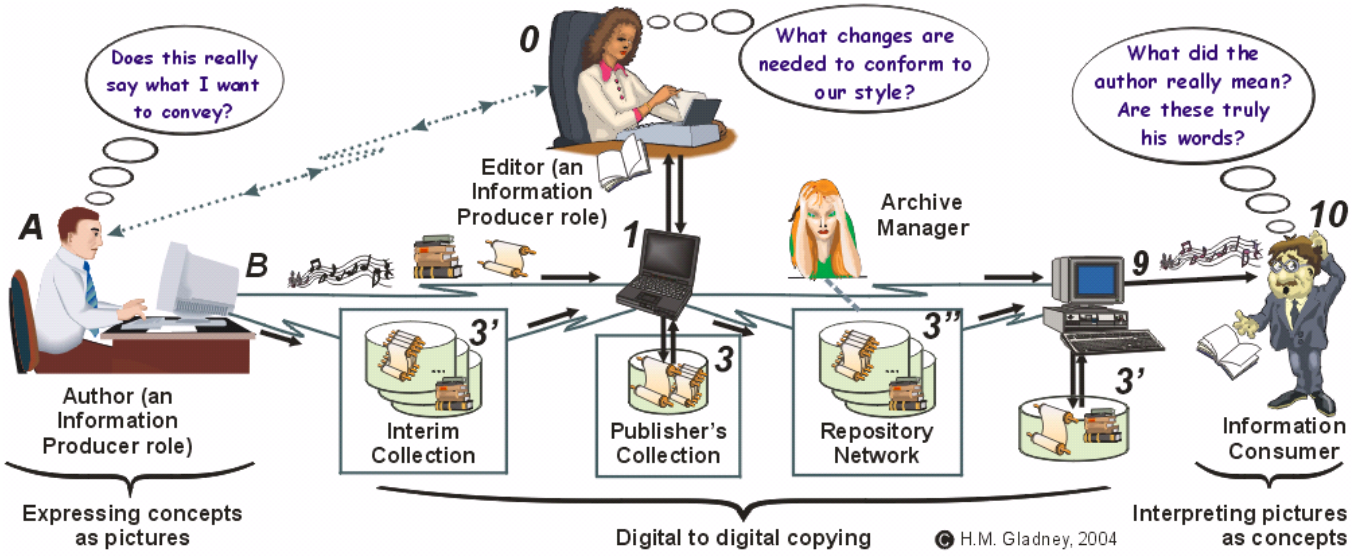

Figure 2: Human roles in cultural document archiving
typical for scientific, cultural, and scholarly documents. Cf. Figure 1.

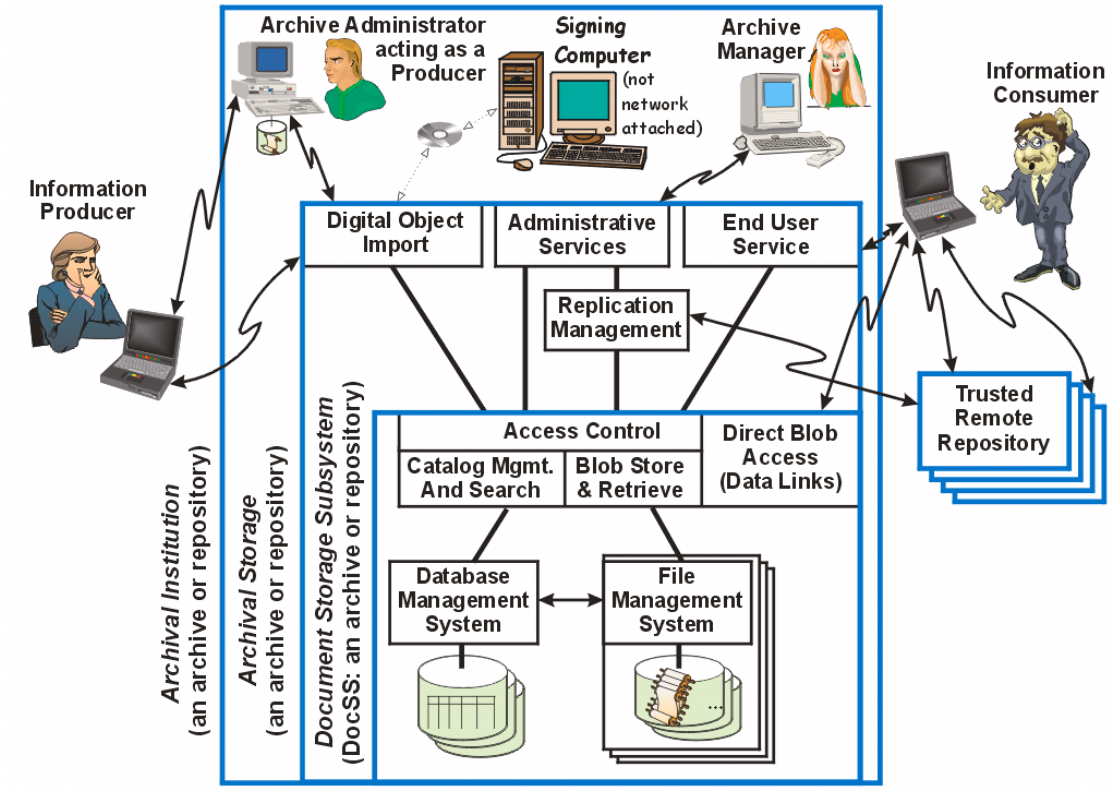

Figure 3: Illustrating nested repository structure
(and that the word 'repository' is ambiguous.)

I believe that the TNA document needs to be refined for use as a management objectives statement by CHM. This is partly because CHM is much smaller than TNA and partly because the TNA document seems to presume a solution that is not the only possibility.

Section 1: Attributes of Authenticity and Integrity

	TNA adaptation for CHM	Comments
1	***Summary***[2]	
1.1.1	Records for preservation are defined as those electronic objects and their concomitant metadata which defines them as records in the sense intended by archivists and which require continued retention until such time as they can be destroyed because they are no longer required for legal purposes or wanted by people authorized to access them.	
1.1.2	This document is a derivative of a set of four documents published in 2004 by The [U.K.] National Archives. It is intended to provide part of a requirements statement for durable software repository services providing public access to content thought to be of long-term interest access to Computer History Museum visitors, including its virtual visitors. The document from which it was derived is intended to define a standard of performance for electronic records that are to be authentic in accordance with *BS ISO 15489 Information and documentation—Records management standard.* A summation of the attributes, which would support an attestation of authenticity and integrity and which need to be maintained as part of the electronic record, is provided in this document.	If records are to be sustained there must be confidence that the maintained records possess authenticity, reliability, integrity and usability
1.1.3	N/A	
1.1.4	Section 2 describes the management controls required for such systems and section *3* addresses the technical requirements needed to maintain sustainable electronic records. Section *4* provides high-level guidance for information providers seeking to qualify their records to be maintained by the Computer History Museum and served to its visitors as authentic records.	
1.1.5	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.	
1.1.6	Other organizations might wish to use or adapt these requirements for their purposes.[3] Familiarity with the concepts of records as used in central government is assumed.	
1.1.7	Any enterprise wishing to use these requirements, as a baseline or benchmark, will 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 enterprise policy and practice, assessing whether any requirements listed in these volumes are highly desirable as opposed to mandatory for their own context	
1.1.8	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.	This needs to be extended to permanently durable records.
2	***Overview***	
2.1	Durably useful electronic records	
2.1.1	See 1.1.1. The records in question are intended to be useful and credible at least for many decades, and possibly indefinitely.	
2.1.2	*The Requirements for Electronic Records Management Systems: Functional Requirements* – Revision 2002 published by the National Archives (TNA) describe many of the controls that would also be required in ensuring that records captured into a managed environment are capable of being sustained as credible records over a defined period of time. The ability to capture the record so that modification, or editing, of the record is no longer possible is a key facet of authenticity and must be supported by credible metadata, audit trails and reports. However whilst it is possible to capture a	A concise formulation of the fundamental requirements for long-term document preservation is available in [Gladney 1].

2 The item numbering corresponds to that in the TNA document from which the current document is derived.

3 Any public service organization is free to do so. The courtesy of a citation is requested.

	TNA adaptation for CHM	Comments
	copy of the record in its original software format and store it unchanged in a repository linked to a corporate classification scheme it may not be feasible or desirable to maintain that record in its original format over the medium term.	
2.1.3	It needs to be recognized that no entirely safe procedure has been invented for so-called 'preservation migration'—repeated format migration when the current information representation is about to become obsolete. A safe alternative was invented at IBM Research by Raymond Lorie in 1995. [Lorie 04]	
2.1.4	In order to achieve sustainable records management each institution will require an appropriate level of functionality together with the requisite tools and business rules required to support sustainable solutions. It will be necessary to sustain electronic records over time as a valued historical asset, in a manner that retains their reliability and integrity for as long as they are required, preserving their value as a historical record. This will include prevention of changes to the content or context to retain authenticity, and continued maintenance in an appropriate format to retain accessibility.	
2.15	Records maintained in electronic form are continually at risk of inadvertent or intentional alteration. Absent suitable protective measures, alterations will be undetectable. The authenticity of electronic records is threatened also whenever unprotected records are transmitted (i.e., when sent between persons, systems or applications) or time (i.e., either when they are stored offline, or when the hardware or software used to process, communicate, or maintain them is upgraded or replaced). Authenticity can also be threatened during access if the environment permits unauthorised and undocumented modifications of the record. Requirements for assessing and maintaining the authenticity of electronic records that are preserved over the long term are necessary to support the presumption that an electronic record is, in fact, and continues to be, what it purports to be and has not been modified or corrupted in essential respects.	'Essential' implies that a statement of purpose for the record at issue has been communicatedand, for an archival record, that this purpose is documented and accessible.
2.1.6	*BS ISO 15489 Information and documentation – Records management standard* requires that a record *"should correctly reflect what was communicated or decided or what action was taken. It should be able to support the needs of the business to which it relates and be used for accountability purposes."*	
2.1.7	In this context this means that the content of a record should *"contain, or be persistently linked to, or associated with the metadata necessary to document a transaction".* The key elements can be summarised as: the structure of a record, that is, its format and the relationships between the elements comprising the record should remain intact; the business context in which the record was created, received and used should be apparent in the record (including the business process of which the transaction is part, the date and time of the transaction and the participants in the transaction); the links between documents held separately, but combining to make up a record, should be present.	
2.1.8	The important characteristics of a records are defined in section 7.2 of the *Information and documentation – Records management standard BS ISO 15489* as comprising authenticity, reliability, integrity, and usability.	A practical definition of authenticity is provided by [Gladney 9].
2.1.9	In order to maintain a sustained object as a credible record it is necessary to define the performance criteria which would provide credible evidence that authenticity, reliability, integrity and usability have been addressed and supported during the period the record has existed. These elements are examined in greater detail below. Ways of identifying the precise attributes of these characteristics are also explored in Section 4: *Guidance for categorising records to identify sustainable requirements*	
3	***Definitions***	
3.1	**Defining Authenticity**	
3.1.1	BS ISO 15489 *Information and documentation – Records management standard* states in section 7.2 that: *An authentic record is one that can be proven* *a) To be what it purports to be,* b) *To have been created or sent by the person purported to have created or sent it, and* c) *To have been created or sent at the time purported.* d) *To have the purported historical significance (relationship to events and circumstances of creation).*	
3.1.2	It goes on to state that "To *ensure the authenticity of records, organisations should*	

	TNA adaptation for CHM	Comments
	implement and document policies and procedures which control the creation, receipt, transmission, maintenance and disposition of records to ensure that record creators are authorized and identified and that records are protected against unauthorized addition, deletion, alteration, use and concealment".	
3.1.3	Reproduces the ideas of 2.1.8 above.	
3.1.4	In practice authenticity can only exist if sufficient elements of the other three characteristics are present. As such authenticity is an implicit value derived or presumed from the presence of the explicit elements that characterise the other three characteristics. A presumption of authenticity is an inference that is drawn from known facts about the manner in which a record has been created, handled, and maintained.	These authenticity requirements are more similar to than different from those for physical artifacts.
3.1.5	A presumption of authenticity will be based upon the number of requirements that have been met and the degree to which each has been met. Requirements are cumulative: the higher the number of satisfied requirements, and the greater the degree to which an individual requirement has been satisfied, the stronger the presumption of authenticity.	This is identical to any other case of evidence evaluation, e.g., as in a court of law.
3.1.6	To maintain a presumption of authenticity the records must be managed in accordance with procedures that ensure their continuing authenticity. The production of copies of the records must be done in accordance with procedures that ensure that their authenticity is not compromised by the reproduction process. The requirements are based on the notion of trust in record keeping and record preservation from the moment of a record's creation. Given some records will be subject to change or alteration if they are migrated to different software formats the standard of trust has to be considered in terms of circumstantial probability rather than certainty.	The TNA wording seems to assume more active curiatorial management than is in fact necessary. See [Gladney 2].
3.1.7	Assessing a record's authenticity involves establishing its *identity* and demonstrating its *integrity*. … The *integrity* of a record refers to its wholeness and soundness: a record has *integrity* if it remains complete and uncorrupted in all its essential respects throughout the course of its existence. This does not mean that a record must be precisely the same as it was when first created for its integrity to exist and be demonstrated. A record can be considered to be essentially complete and uncorrupted if the message that it is meant to communicate in order to achieve its purpose is unaltered.	Two records with the attributes mentioned can be different records! There is an untestable assumption in the TNA words, "if the message … is unaltered." See [Gladney 9].
3.2	**Defining Reliability**	
3.2.1	BS ISO 15489 regards a reliable record as one *whose contents can be trusted as a full and accurate representation of the transactions, activities or facts to which they attest and can be depended upon in the course of subsequent transactions or activities.* Records should be created at the time of the transaction or incident to which they relate, or soon afterwards, by individuals who have direct knowledge of the facts or by instruments routinely used within the business to conduct the transaction.	
3.2.2	BS ISO 15489 further states in clarification of the characteristic reliability that … The statements here are a weak subset of the requirements of any content management system—even if long-term preservation and access are not among the service objectives.	
3.2.3	Reliability therefore will be apparent if there is evidence that the records were created and captured as part of a legitimate business process and assigned to a logical and appropriate location within the businesses own classification schema or file-plan where the record will then be subject to enterprise management of its disposal. The identity and where possible the specific role of everyone involved in the creation and capture of the record should be clearly apparent and part of the historical record that is reliably preserved. The operational context or business process within which a record has been generated or managed should also be visible.	
3.2.4	Reliability should secure the identity of a record as described in paragraph 3.1.7 above. In order to implement policy and complete transactions every business needs reliable records placed within a logical context. If reliability is not built into the operational processes by the adoption of record management functionality at the time of a record's creation and capture it is unlikely that it can be asserted subsequently with any degree of confidence. The application of records management functionality should secure reliability—integrity however is a demonstration that the controls placed upon a record upon its capture into a "reliable environment" were secured and maintained for as long as the record is required.	
3.2.5	The need for reliability will differ from operational reliability needs and also according to the different types or categories of records created and held by a department	

	TNA adaptation for CHM	Comments
3.2.6	The characteristic of reliability itself can be broken down into three sub elements. These are: trust, relationship/context, and longevity.	
3.2.7	Trust is critical to reliability as without it there can be no meaningful faith in the accuracy of the retained records. Relationship and context refer to the comprehension of the meaning and value of records, which relies upon the ability of the reader to place the records in their operational context in a manner that their relationships with other affected records are clear and transparent.	
3.3	**Defining Integrity**	
3.3.1	BS ISO 15489 states that *"the integrity of a record refers to its being complete and unaltered".* It is necessary that a record be protected against unauthorized alteration. Records management policies and procedures should specify that no additions or annotations may be made to a record after it is committed to archival custody. Any desired annotation, addition or deletion to a record should be made in a new record version that is bound both to the base version and to a provenance statement for and a description of the alterations.	
3.3.2	BS ISO 15489 clarifies this with the following statement: *Control measures such as access monitoring, user verification, authorised destruction and security controls should be implemented to prevent unauthorised access, destruction, alteration or removal or records. These controls may reside within a records system or be external to the specific system. For electronic records the organisation may need to provide that any system malfunction, upgrade or regular maintenance does not affect the records*	
3.3.3	To sustain a presumption of authenticity it is necessary to identify the procedural controls over electronic records that provide a circumstantial probability of their integrity. The controls that define integrity include: establishing access privileges over the creation, modification, annotation, relocation, and destruction of records; instituting procedures to prevent, discover, and correct loss or corruption of records; implementing measures to guarantee the continuing identity and integrity of records against media deterioration and across technological change; where multiple copies of records exist, formally identifying the authoritative record; and clearly identifying and maintaining, along with the records, all the documentation necessary to understand their legal, administrative and technical content.	This can be accomplished without repository procedural controls.
3.3.4	If reliability emerges from the original operational purpose that caused the record to be created integrity should reflect the long-term business needs that are served by the continued existence of a record. BS ISO 15489 differentiates reliability from integrity which suggests there is a distinction to be drawn between the immediate operational need, which requires records to be reliable to ensure effective transactions, and the longer term business need, where those same records must display integrity through possessing a quality of auditability ensuring that they can be considered to be authentic over time. If integrity is absent authenticity is very difficult to adduce let alone assert.	
3.3.5	(Left blank deliberately, because the corresponding TNA item does not define integrity, but instead mentions methods for testing whether it is achieved.)	
3.3.6	In order to confirm the record is unchanged or that only authorised and appropriate changes have been made, the status of the records and the presence or absence of change has to be auditable or traceable.	
3.3.7	Integrity is bound to the need to demonstrate authenticity over time. It is presumed that all CHM software collection holdings are to be held forever.	
3.3.8	In certain instances it may be necessary or desirable to retain records related to a broad record category where the records were themselves generated in response to codes of instruction or standards in force at that time. In order to confirm if the record of a transaction was valid in these circumstances it may be necessary to reference the rules that applied at that juncture. The standards or rules need to be preserved as reliably as the records that depend on them	
3.3.9	The issue of record integrity is closely linked to effective business continuity planning in that in order to clarify the cost of maintaining record integrity it is necessary to evaluate the risk to the organisation if the records have been retained as incomplete or with limited auditable functionality.	

	TNA adaptation for CHM	Comments
3.4	**Defining Usability**	
3.4.1	BS ISO 15489 defines a usable record as *one that can be located, retrieved, presented and interpreted.* It should be capable of subsequent presentation and/or use as was the case when it was an active part of operations. The contextual linkages of records should carry the information needed for an understanding of the transactions that created and used them. It should be possible to identify a record within the context of historical usage. Any links between records that document a sequence of activities should be maintained.	
3.4.2	The aspects that will define usability will vary according to the business need. Initially the operational process, which created or captured a record will define how it is to be used and stored. However as the operational purpose that created that record evaporates other needs for this information may come to the fore. It may be necessary to provide access to the information in different forms linked or related to other material created subsequently. The issue for any business is that whatever drives the business need for the information and in whatever form this may take it should be immediately accessible to authorised users and the context in which the record was created and held should be apparent if that information is also required.	
3.4.3	Usability comprises at least four key elements: The form or forms which the organisation may wish to view or publish this information; The ability to produce new renditions in other formats as additional instances of the record whilst maintaining links to the original record; The access permissions which allow access to the record or to redacted instances of the record (e.g. where it is necessary to publish or release a limited subset of the information but where some details such as names or addresses are retained); The ability of the user to know where this information was obtained and where it can still be located and retrieved if a requirement for authentication is established; At no point should usability infringe upon the integrity of the record.	
3.4.4	The requirement for usability may appear superficially the easiest to scope and comprehend particularly where the records either consist of images or text. The issue can appear to revolve around the availability and presence of the appropriate rendering software or execution environment (for programs). However, the issue is more complex than the previous analysis might suggest as usability is also about ease of locating, quick retrieval and the quality of the presentation. The first question to ask is: What makes a record usable and how might this differ according to different types of records?	This needs to be extended to include executables (computer programs).
3.4.5	Four sub-elements then need to be considered in evaluating the requirement for the usability of records over time. These are locating, retrieving, presentation, interpretation. For programs, execution is also important.	
3.4.6	Locating refers to the means used to reliably identify without undue difficulty the record or records needed to satisfy the user's query. The location within the business classification schema or file-plan is one aspect but also the issue of accurate titling, meaningful nomenclature and the use of aliases or alternative titling fall into this area.	
3.4.7	Effective retrieval is dependent upon identification of the anticipated pattern of access demand and the application and continued management of appropriate access permissions across time.	
3.4.8	Effective presentation ensures the user can retrieve and view the records with the appropriate level of functionality required to undertake a meaningful interpretation. In some instance this may require the original program to be available so that the data can be manipulated or edited using the same functionality to create a new document or version, which can then be saved and added to the enterprise record without changing or deleting the original.	
3.4.9	Interpretation at its simplest can be addressed by an ability to view text or images using a simple browser without the enhancements offered by the original software, for example one can view document created in MS Word using a text file viewer such as WordPad although the formatting is lost in this view. In other circumstances seeing the content without the display and formatting built into the original document makes interpretation difficult if not impossible.	
3.4.10	In other instances interpretation also needs to be supported by linked contextual information, for example the ability to view the metadata of the record in both its original and existing context. This may require users having sight of both the current business classification system in which the records reside and the original classification system	

	TNA adaptation for CHM	Comments
	where that differs from the current version. This situation can arise where functions have been transferred between government bodies resulting in bulk exports and imports of metadata and data between EDRM platforms.	
4	***Attributes of authenticity and integrity***	
4.1.1	For a practical definition of 'authenticity' see [Gladney 9].	
4.1.2	A presumption of authenticity is an inference that is drawn from known facts about the manner in which a record has been created, handled, and maintained. Section 2 and 3 of these requirements describe the management and the technical requirements needed to support a presumption of authenticity. Section 4 of these requirements *Guidance for categorising records to identify sustainable requirements* also describes the elements that form the key characteristics of a record and provides a list of the key questions departments will need to address when formulating their strategies to sustain record categories over time.	
4.1.3	As stated previously in paragraph 3.1.7 assessing a record's authenticity involves establishing its *identity* and demonstrating its *integrity.* The *identity* of a record refers to the attributes, including external attributes such as context and provenance, that uniquely characterise it and distinguish it from other records (the name of the author, its date and place of origin, its subject); while the *integrity* of a record refers to its wholeness and soundness: a record has *integrity* if it remains complete and uncorrupted in all its essential respects throughout the course of its existence.	
4.1.4	The colloquial notion of 'the original' of any artifact is insufficiently precise to be used for digital documents. See [Gladney 4 §2.6] for a discussion of the issue and its resolution.	
4.1.5	When records are captured into a controlled domain required for long-term storage, it is necessary for the maintaining authority to establish whether, and to what extent, the records have been maintained using technologies and administrative procedures that either ensure their authenticity or at least minimise risks of change from the time the records were first set aside to the point at which they are subsequently accessed. The requirements described below deal with the maintenance of authenticity. After the attributes for supporting authenticity of the electronic records have been established, their authenticity needs to be maintained over the long term across different hardware and software platforms and in some cases changes of custodian. Authenticity has also to be maintained where records are selected for permanent preservation and transferred into the custody of a specialized archive.	The simplest way to achieve this for an authentic record is to avoid changing the record in any way.
4.1.6	To do so, that part of the organisation charged with the responsibility of maintaining and preserving reliable and authentic records must manage the electronic records in accordance with procedures that ensure their continuing authenticity. They must produce copies of those records in accordance with procedures that ensure that their authenticity is not compromised by the reproduction process.	Ditto
4.1.7	The organisation's own policies and procedures have to reinforce the characteristics of a trusted record management system. A trusted record management system includes the rules that control the creation, maintenance, and use of the creator's records, which support a presumption of the authenticity of the records within the system. The requirements have to identify the core information about an electronic record that must be persistently linked to it over time and across hardware and software platforms in order to establish and perpetuate its identity. Such information includes, among other things, the names of the creator, addressee, and custodian, the indication of the action or matter to which the record relates, the manifestation of the record's context within the classification system (what is referred to as the "archival bond"), and the indication of any annotations and attachments. These elements are clarified in the *Requirements for Electronic Records Management Systems Metadata Standard* published by the National Archives (TNA) and can be accessed at http://www.pro.gov.uk/recordsmanagement/eros/invest/2002metadatafinal.pdf	The notion of a 'trusted record management' system is flawed because no complete and testable procedure set has been written for it, or can (we believe) be written. (To some extent, this is discussed in [Gladney 3].)
4.1.8	The metadata standard indicates, for the first time, some metadata at the component level (i.e. a level below that of the individual record and consisting of the single physical object (i.e. the smallest level of granularity the operating system can handle—MS-DOS or UNIX file level). This is the first phase of extending PRO guidance on metadata into the areas of sustainability and preservation of business records within departments. This Standard is extensible to allow for these developments to follow. Element 16 in the standard – Preservation- is not yet fully defined at this stage to flag up an area that is to be developed within the next 12 months. It is expected that the definition of requirements and accompanying metadata for sustaining records in departments, as well as work on	

	TNA adaptation for CHM	Comments
	permanent preservation in the National Archives (TNA), will lead to additions to this area of the metadata framework.	
4.1.9	This metadata with additional information about changes the electronic records of the creator have undergone since they were first created will comprise the Preservation element in the metadata standard.	
4.1.10	To sustain a presumption of authenticity it is necessary to identify the controls over electronic records that provide a circumstantial probability of their integrity. Such controls include: establishing access privileges over the creation, modification, annotation, relocation, and destruction of records; instituting procedures to prevent, discover, and correct loss or corruption of records; implementing measures to guarantee the identity and integrity of records against media deterioration and across technological change; where multiple copies of records exist, formally identifying the authoritative record; and clearly identifying and maintaining, along with the records, all the documentation necessary to understand their legal, administrative and technical context.	This is insufficient without (1) a complete and objective prescription of business controls and (2) evidence that those business controls have been flawlessly exercised.
4.1.11	The requirements assume the existence of a role of a trusted custodian. The management requirements are published in Section 2 of these generic requirements: *Sustaining authentic and reliable records: management requirements.* These describe the criteria necessary to enable custodians to attest to the authenticity of electronic records after they have been transferred to their custody. Increasingly this will commence when a document is captured into an electronic record management system (ERMS). This role will require the custodian to actively intervene as part of the long-term maintenance process (e.g. software migration). Such interventions may require the application of approved and documented alterations to ensure the record remains usable whilst at the same time ensuring that the authenticity of the record is not affected. To be considered a custodian, an organisation must demonstrate that it provides no opportunity for unauthorised alterations to the records, or to allow others to alter them in such a manner that the alteration compromises the authenticity of the record; and that it is capable of implementing procedures that ensure that any loss or change to records over time is avoided or at least minimised.	Trusted by whom, and for what? The procedures alluded to in the TNA formulation depend on subjective judgments that have not been reduced to testable clerical steps.
4.1.12	These controls required by a custodian can be summarised as: maintaining unbroken custody of the records, implementing and monitoring security and control procedures; and ensuring that the content of records and any required elements of documentary form and annotations remain unchanged after any reproduction or transformation process.	How can the custodian demonstrate to a skeptical information consumer that transformations meet the criterion?
4.1.13	The maintaining organisation must also be able to demonstrate that the activity of reproduction has been thoroughly documented; and that the description of a given body of electronic records includes information about any substantial changes the records have undergone over time. Documentation and description are essential means of accounting for the integrity of the maintenance process in general and the reproduction process in particular and are necessary, therefore, to the proper fulfillment of the role of a custodian.	How can the organization demonstrate that essential information has not been distorted during transformations?
4.1.14	This means that the authority and legitimacy of the claims made for the authenticity of electronic records derive entirely from the integrity and internal coherence of the procedures adopted to manage them. It follows that an organisation needs, not only to design and implement procedures that provide a strong probability of record trustworthiness but also to provide an honest and adequate account of the choices and decisions taken, during the stewardship of the custodial organisation.	As far as I know, no-one has specified how this can be done to demonstrate that errors or malfeasance have not been introduced.
4.1.15	For further information on what controls and mechanisms a custodial organisation will need to adopt, refer to sections 2 and 3 of the Generic Requirements listed below.	

Section 2. Sustaining authentic and reliable records: management requirements

	TNA adaptation for CHM	Comments
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	N/A	
1.1.3	Section 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.* Section 3 addresses the technical requirements needed to maintain sustainable electronic records. Section 4 provides high-level guidance for departments seeking to categorise their records to scope the specific requirements needed to sustain these record categories as authentic records.	
1.1.4	This section defines the key elements that should be incorporated within any management strategic planning framework and the processes that will also have to be developed and supported in order to ensure that electronic records which are to be sustained over a defined period of time are able to satisfy the characteristics of a record as defined in *BS ISO 15489* that is authenticity, reliability, integrity and usability. If these characteristics are not maintained the sustained records will lose credibility and will lose evidential value. This section of the generic requirements will define performance indicators and non-functional requirements as opposed to the technical management requirements which are described in Section 3 of the *Generic requirements for sustaining electronic information over time - Sustaining authentic and reliable records: technical requirements.*	
1.1.5	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	
1.1.6	It is also recommended that any management process developed in response to the guidance given in this document should also be bench-marked against BS *ISI 7799* on information security.	
1.1.7	N/A	
1.1.8	Any enterprise 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 enterprise 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	***Operational frameworks***	
2.1.1	In order to develop and implement appropriate strategies to sustain their electronic records organisations will first need to establish a framework, which will define the business needs which are supported by each group of records in their custody, the style and content of the metadata, which will accompany those records and an appropriate technical solution for each software format included in the collection. It will also be necessary to ensure reproduction of copies of records upon demand together with any functionality that the business purpose may require when accessing these records. It will also be necessary to determine the performance criteria for evaluating the success of any action implemented within the framework. Criteria for assessing	N.B. Here we see examples of specifications that are not objective enough or technical enough to guide software engineers.

	TNA adaptation for CHM	Comments
	such frameworks will include: • definition of what an implementation plan will need to address • establishment of the business need serviced by the records • whether the business need has been met by the adopted strategy	
2.1.2	N/A	
2.1.3	The tools provided here will define how successful sustainability strategies will be evolved, the mechanisms to ensure their continued relevance together with performance criteria to provide measures to assess the effectiveness of the process.	
2.1.4	The management strategy will have to establish actions that must be executed either at specified times or under specified conditions. It will clarify and define the business purpose, and the technical requirements for maintaining the records in a form that enables that purpose. The document will have to provide performance criteria and quality indicators for the following sections.	
2.2	***Management Controls***	
2.1.1	Sustainability will be managed by producing a comprehensive framework consisting of sets of strategies and action plans. Each of these will be linked to a specific body of electronic records, categorised according to business requirements, which establish a need to maintain the records for a defined period. Each body of records will have a defined level of functionality with an identified technological infrastructure and methodology needed to implement the action plans.	See comment.
2.2.2	The elements listed below should be used to scope the scale of the work required to execute the sustainable strategy. The management process will also produce information about the maintenance function and about how the records are being sustained or preserved and this information should be used to refine and re-focus the strategy. It is also necessary to define the performance environment within which any preservation strategy will operate. The key elements to be addressed are: • Identifying the records that require to be sustained • External controls (e.g. applicable legal requirements and regulations, management requirements) • Defining the nature of the products (e.g. the standard and form of the records that are to be sustained • Resources required to execute the strategy (e.g. personnel, infrastructure etc.)	
2.2.3	These four components need to beb fully identified in every strategic framework with appropriate descriptions for each sub-element listed below. **Inputs** • Information about the content of electronic records required to be sustained • Information about the nature and capability of the software to be sustained • Management information and experience of sustainability (e.g. migration, configuration, emulation) **External environment and controls** • Institutional requirements (including any regulatory or legal constraints) • User access requirements • Current IT/IS infrastructure **Products** • Criteria for assessing if the sustained records meet the business need • Information about actively maintained records • Priorities for sustainability and action plans • Maintenance strategies • Assessment of continued authenticity of records • Proposed changes to technological infrastructure **Resources** • Dedicated facilities and infrastructure needed to deliver sustained records • Personnel needed to sustain the records	
2.2.4	The processes and mechanisms, which will need to be established to ensure that the strategic framework is credible are described in sections 2 through to 12.	
2.2.5	N/A	
3	**Determining the appropriate maintenance environment**	
3.1.1	Identifying the relevant regime to maintain records in an appropriate environment entails identifying the categories of objects that must be maintained. This includes specifying, for each category, the attributes and methods that must be preserved, as well as any requirements for certifying that any reproduced record is authentic.	

	TNA adaptation for CHM	Comments
3.1.2	Evolution of a requirement will be guided by evaluation of prior experience in applying such requirements to records that have been transferred to a sustained environment. The result of this process will be an informed and enhanced requirement, where the specification consists in identifying what operational, institutional and regulatory requirements apply to what records and how each can be implemented.	N.B.
3.1.3	A coherent set of requirements for maintaining electronic records in a manner which will enable reproduction of the records and where this is required certification of the authenticity of the reproduced records. Each set of requirements will apply to a specified collection of digital objects or records. The requirements encompass both the: • the digital objects themselves • record collections, categories or classes • storage media to be used the maintenance of digital files	
3.2	**Performance measures for maintained records**	
3.2.1	Requirements for media include: • standards and specifications of what media are to be used and for what purpose • how volumes are to be labeled • and how physical files are to be written.	
3.2.2	Requirements for digital objects include: • how both physical and logical files are to be identified • how logical files are mapped to physical files • how integrity of a file is ensured • specifications for appropriate software file formats • criteria for assessing and selecting current and future media and software file formats	These are less requirements statements than needed solution specifications.
3.2.3	Requirements for record collections include: • how records are to be composed from digital components • how records in an archival aggregate are to be arranged • how the business needs and concomitant maintenance need of records are to be expressed, captured and stored	These are topics to be addressed rather than requirements statements.
4	**Maintaining effective records – the role of a technology watch programme**	
4.1.1	N/A	Not needed. See [Gladney 2].
4.1.2	The storage media will need to be replaced with a different media type periodically as technology changes and a migration project to move to the new storage media should be undertaken before accessing the older storage media becomes problematic.	
4.1.3	The elements of a capture and preservation program, which will achieve safe and permanent preservation, which will define effective maintenance, are: • identifying the mechanisms to decide which formats of record components are to be captured into the archive; • identifying the criteria for a safe and durable preservation format; • determining what type of preservation format is required; • developing and maintaining appropriate translators; • identifying such auxiliary records as must accompany each record; • transformation to the chosen persistent format; • determining minimum information levels to be captured within the management audit trail for each process; • packaging with standard metadata and either packaging or linking to each required auxiliary record; • bringing each linked auxiliary record into the repository network. • identifying criteria to evaluate whether the conversion for preservation has been successful; • establishing criteria for assessing whether the entire process has been successful, and testing according to those criteria.	N.B.
4.2	***Capture into a secure record-keeping environment***	
4.2.1	Capturing the record within the electronic environment involves management of the interface between the record-keeping system and the applications, such as word processors or e-mail clients, which are used to create or receive records. Systematic capture requires both a technical interface and a set of rules or procedures, which govern its behaviour and successful application within the organisation. Maintenance of sustainable records requires that records should have been captured upon creation into	The security required is protection against unauthorized or inappropriate modification.

	TNA adaptation for CHM	Comments
	a managed environment and these controls should continue to apply for as long as the record is required by the organisation.	
4.2.2	However there will be cases where it is necessary to maintain records created in an unmanaged environment (that is held outside of an EDRMS environment). In such a case they should be imported into a managed environment where the mechanisms of sustainability can be applied and recorded. The record of capture will need to be annotated to record the circumstances in which the records were created and stored until their formal capture into a managed environment.	The CHM SW collection will hold records that have not yet been prepared for the long-term collection.
4.2.3	Depending on the business need for capture from such environments it may be useful to look at the example provided by forensic IT investigation, which is predicated on the capture of data from systems in a legally-admissible form. Within the United Kingdom the recommended basis for such procedures are the various codes of practice published by the British Standards Institution ref. *DISC PD 0008:1999* and *DISC PD 5000 - 1-6:2002.*	
4.2.4	The mechanisms for capture should ensure that: • appropriate records are captured. There should be a clear understanding of the information which should be captured as a record, and the operational means of identifying and capturing this within the working environment; • all types of record are captured. Workable mechanisms should exist for all record-creating applications in use to enable the capture of records from that application according to approved formats and standards; • complete records are captured. Capture mechanisms should be capable of acquiring all the elements which make up a record, and associating these together in a meaningful and useful manner; • metadata is captured and associated with records from the time of their creation, and that this descriptive metadata is closely bound with the record itself; • links to other records are established and maintained, within broader record assemblies, including mixed electronic and paper assemblies, and in other classification mechanisms if appropriate.	
4.2.5	The above requirements emerged in the definition of the *Requirements for Electronic Records Management Systems: Functional Requirements* – Revision 2002 published by the National Archives (TNA) Those requirements are relevant where a sustainable records policy is to be applied. Other typical requirements for electronic records management, which will also apply in a managed maintenance environment, are: • capturing, storing, indexing and retrieving all elements of the record as a complex unit, and for all types of record; • management of records within class categories or filing structures to maintain the narrative links between records; • record level metadata describing contextual information; • integration between electronic and paper records; • secure storage and management to ensure authenticity and accountability, including support for legal and regulatory requirements; • appraisal and selection of records for preservation and transfer to the keeping of the National Archives (TNA) or other permanent archive; • systematic retention and disposition of records; • migration and export of records for permanent preservation.	
4.2.6	All records upon capture into a managed environment should be accompanied by the following metadata as a minimum: • a unique identifier assigned from the system; • the data and time of registration; • a title or abbreviated description; • the author (person or corporate body), sender or recipient.	More complete sets are expressed in [OAIS] and [METS].
4.2.7	The *Requirements for Electronic Records Management Systems Functional Requirements 2002 revision: final version* provides much of the requisite functionality that is needed if records are to be pro-actively sustained. The Metadata Encoding and Transmission Standard [METS] describes the needed metadata elements.	
5	**Record attributes and linkages to records**	
5.1.1	The presumption of a record's authenticity is strengthened by knowledge of certain basic facts about it. The attributes identified in these requirements embody those facts. The requirement that the attributes be expressed explicitly and linked inextricably to the record during its life, and carried forward with it over time and space, reflects a need	Note especially "linked inextricably".

	TNA adaptation for CHM	Comments
	that such expression and linkage provide a strong foundation on which to establish a record's identity and demonstrate its integrity.	
5.1.2	The link between the record and the attributes is a conceptual rather than a physical one, and the requirement could be satisfied in different ways, depending on the nature of the electronic system in which the record resides. In an ERMS, this requirement is usually met through the creation of a record profile. When a record is exported from the live system, migrated in a system update, or transferred to an external specialist archive, the attributes should be linked to the record and available to the user. When pulling together the data prior to export, the creator should also ensure that the data captured is the right data.	[Gladney 3] describes and justifies a method for accomplishing this.
5.1.3	To support a presumption of authenticity the custodian must possess, obtain and maintain evidence that the following metadata attributes defined in the *Requirements for Electronic Records Management Systems Metadata Standard* are supported: • Identifier System ID; • Title; • Creator; • Date Created; • Date Acquired (mandatory for e-mail); • Date Declared; • Addressee (mandatory for e-mail); • Type Record type (mandatory where applicable); • Relation Copy (pointer) (mandatory where applicable); • Relation Parent object; • Relation Redaction/Extract (mandatory where applicable); • Relation Reason for redaction/extract (mandatory where applicable); • Relation Rendition (mandatory where applicable); • Aggregation; • Rights Protective marking.	Compare [METS]. (See 5.1.3b.)
5.1.3b	CHM records should be marked according to METS (the Metadata Encoding and Transmission Standard).	
5.1.4	To support a presumption of integrity the custodian must possess or obtain evidence that the following attributes are supported: • name of the creating organisation that regards the record as part of its official corporate record; • indication of types of annotations added to the record; • indication of technical modifications.	A more careful definition of a 'complete' provenance assertion is needed.
6	**Access Control**	
6.1.1	Defining access privileges means assigning responsibility for the creation, modification, annotation, relocation, and destruction of records on the basis of competence, which is the authority and capacity to carry out an administrative action. Implementing access privileges means conferring exclusive capability to exercise such responsibility. In electronic systems, access privileges are usually articulated in tables of user profiles. Effective implementation of access privileges involves the monitoring of access through an audit trail that records every interaction that an officer has with each record (with the possible exception of viewing the record).	
6.1.2	An information producer or an archive manager should define and implement access privileges concerning the creation, modification, annotation, relocation, and destruction of records.	
6.1.3	The custodian has to maintain and update and where necessary extend existing access privileges to implement all changes relating to modification, annotation, relocation, and destruction of records.	
7	***Technical modifications***	
7.1.1	Technical modifications are any changes in the digital components of the record. Such modifications would include any changes in the way any elements of the record are digitally encoded and changes in the methods applied (e.g. software) to reproduce the record from the stored digital components; that is, any changes that might raise questions as to whether the reproduced record is the same as it would have been before the technical modification. The indication of modifications might refer to additional documentation external to the record that explains in more detail the nature of those modifications.	The feasibility of meeting this objective completely, correctly, and reliably is far from clear!
7.1.2	The reason for any modification has to be fully documented as must the nature and date of application of a specific process together with references of the records, objects	

	TNA adaptation for CHM	Comments
	or components that have been subject to modification. It is essential that this information be incorporated within the database charged with the management oversight of the records being sustained or preserved by the custodian.	
7.1.3	No archival record should be destroyed without appropriate authorization that this be done. (Archival records are valuable institutional resources, and as such should be subject to conventional asset controls.)	
7.14	No archival record should be altered. Instead, if a revised version is needed, a new version should be created and inextricably linked to the version from it was constructed. Metadata to satisfy 7.1.2 should be inextricably linked to the new version.	Note "inextricably linked."
7.2	**Protective Procedures: documenting management of loss and corruption**	
7.2.1	Procedures to protect records against loss or corruption include: prescribing regular back-up copies of records and their attributes; maintaining a system back-up that includes system programs, operating system files, etc.; maintaining an audit trail of additions and changes to records since the last periodic back-up; following any system failure ensuring that the back-up and recovery procedures will automatically guarantee that all complete updates (records and any control information such as indexes required to access the records) contained in the audit trail are reflected in the rebuilt files and also guarantee that any incomplete operation is backed up. The capability should be provided to rebuild forward from any back-up copy, using the back-up copy and all subsequent audit trails.	This is conventional backup and recovery management required for any valuable data collection.
7.2.2	The creator or custodian has to establish and effectively implement procedures to prevent, discover, and correct loss or corruption of records.	N.B.
7.3	**Protective Procedures: Media and Technology**	
7.3.1	Procedures to counteract media fragility and technological obsolescence include: planning upgrades to the organisation's technology base; ensuring the ability to retrieve, access, and use stored records when components of the electronic system are changed; refreshing the records by regularly moving them from one storage medium to another; and migrating records from an obsolescent technology to a new technology.	This is part of conventional data center procedures.
7.3.2	The creator or custodian has to establish and effectively implement procedures to guarantee the continuing identity and integrity of records against media deterioration and across technological change.	N.B.
7.4	**Media Refreshment and Migration**	
7.4.1	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 time scales 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.	This is conventional backup and recovery management required for any valuable data collection.
7.4.2	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 exercise should be undertaken. 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 technology watch strategy. Following a successful media migration a new media refreshment scheme must be established and maintained as described above.	7.4.1 through 7.4.3 should be applied to data considered as a set of bit-strings. I.e., without regard to the meaning or format of any record.
7.4.3	After selection and prior to refreshment or migration taking place, a new media handling process guide must be approved if an appropriate one does not already exist.	
7.4.4	The process 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.	This does not seem to be important for CHM data.
7.4.5	N/A.	No CHM SW holdings are expected to require rigorous security against unauthorized disclosure.

	TNA adaptation for CHM	Comments
8	**Establishment of Document Types**	
8.1.1	The document type of a record as defined in the *Requirements for Electronic Records Management Systems Metadata Standard* may be determined in connection to a specific administrative procedure, or in connection to a specific phase(s) within a procedure. The document type or form may be prescribed by a business process or workflow, where each step in an administrative procedure is identified by a specific record or document type. If a creator customizes a specific application, such as an electronic mail application or template within MS Word for a pro-forma, to carry certain fields, the customized form becomes, by default, the required document type. It is assumed that the creating organisation, acting on the basis of its own needs or because of pre-existing legal requirements, will establish the required document types for their records.	
8.1.2	When the creator establishes the document type in connection to a procedure, or to specific phases of a procedure, that will allow for the maintenance of the authenticity of the record. As that determination will vary from one form of a record to another, and from one creating organisation to another, it is not possible to predetermine or generalise the relevance of specific elements of documentary form in relation to authenticity.	
8.1.3	The creator or custodian has to establish the document types or forms of records associated with each procedure (e.g. templates, forms etc) either according to any legal or regulatory requirements system or those of the creating organisation and ensure these are documented.	
9	***Authentication of Records***	
9.1.1	In common usage, to authenticate means, or serves, to prove the authenticity of something. More specifically, the term implies establishing genuineness by adducing legal or official documents or expert opinion. For the purposes of these requirements, authentication is understood to be a declaration of a record's authenticity at a specific point in time by a person entrusted with the authority to make such declaration. It may take the form of an authoritative statement (which may be in the form of words or symbols) where there is a specific legal provision for such a statement or it may take the form of a digital signature whose authenticity can be verified using the public key infrastructure (PKI). The effect of either method when added to or inserted into a copy of the record is to attest that the record is authentic. The requirement may also be met by linking the authentication of specific types of records to business procedures and assigning responsibility to a specific office or officer for authentication or by the adoption of enabling technology such as watermarking, digital or biometric signatures.	What is commonly understood under the label 'PKI' has flaws that make it inappropriate for the purpose called for here.
9.1.2	It should be emphasised that the adoption of such authentication methods described above should not normally be inserted into the preserved record whereas it may be appropriate, or required, for the custodian to use one of these methods when providing an authenticated copy to a third party. If an authentication method is inserted into the preserved record this will change its attributes and could compromise the integrity of the record thereby calling into question its authenticity. Intrusive techniques such as watermarking may themselves compromise the authenticity of digital data, and certainly may not be sustainable over time. In general any technique, which alters the bit stream of the actual record held and maintained by the custodian should, be avoided.	[Gladney 3] describes a safe and reliable way for combining changes with base document versions—a way that propagates authenticity.
9.1.3	The authentication of copies differs from the validation of the process of reproduction of the digital components of the records. The latter process occurs every time the records of the creator are moved from one medium to another or migrated from one technology to another.	
9.1.4	Where authentication is required by statute or the needs of the organisation, the creator or custodian has to establish specific rules regarding which records must be authenticated, by whom, and the means of authentication.	
9.2	**Identification of Authoritative Record**	
9.2.1	An authoritative record is a record that is considered by the creator to be its official record and is usually subject to procedural controls that are not required for other copies. Normally any record subject to management by an ERMS has the capability of being an authoritative record but the organisation may regard certain document types (e.g. affidavits) as being especially authoritative as they are required to use them in discharge of legal obligations. In such cases such a designation should be apparent in the record retention schedule applied to those records and the metadata designating an authoritative record should form part of the Rights and Mandate elements as defined in the Record Management Metadata Standard.	

	TNA adaptation for CHM	Comments
9.2.2	If multiple copies of the same record exist, the creator or custodian has to establish procedures that identify which record is considered authoritative.	
9.2.3	It is understood that in certain circumstances there may be multiple authoritative copies of records, depending on the purpose for which the record is created.	
10	**Removal and Transfer of Relevant Documentation**	
10.1.1	Where there is a transition of records from active status to semi-active and inactive status, which involves the removal of records from one platform to another system or an archive external to the organisation, the creating organisation has to establish and effectively implement procedures determining what documentation has to be removed and transferred to the new system which is to be maintained and preserved along with the records.	
10.1.2	This requirement implies that the custodian needs to carry forward with the removed records all the information that is necessary to establish the identity and demonstrate the integrity of those records, as well as the information necessary to place the records in their relevant contexts.	
11	**Controls over records, export, maintenance, and reproduction**	
11.1.1	The controls over the transfer (export) of electronic records across platforms include establishing, implementing, and monitoring procedures for registering the records' export; verifying the authority for export; examining the records to determine whether they correspond to the records that are designated in the terms and conditions governing their export; and formally importing the records onto the new platform.	
11.1.2	As part of the export process, the assessment of the authenticity of the creator's records should be verified. This includes verifying that the attributes relating to the records' identity and integrity have been carried forward with them along with any relevant documentation.	The focal event is the action of transfer of a record from one responsible party to another.
11.1.3	The custodian should establish access privileges concerning the access, use, and reproduction of records; establish procedures to prevent, discover, and correct loss or corruption of records, as well as procedures to guarantee the continuing identity and integrity of records against media deterioration and across technological change. Once established, the privileges and procedures should be effectively implemented and regularly monitored. If authentication of the records is required, the custodian should establish specific rules regarding who is authorized to authenticate them and the means of authentication that will be used.	
11.1.4	The controls over the reproduction of records include establishing, implementing, and monitoring reproduction procedures that are capable of ensuring that the content of the record is not changed in the course of reproduction.	
11.1.5	The procedures and system(s) used to transfer or export the records to another platform; maintain them in-situ; and reproduce them must provide adequate and effective controls to guarantee the records' identity and integrity, and specifically that: • unbroken custody of the records is maintained; • security and control procedures are implemented and monitored; • the content of the record and any required annotations and elements of documentary form remain unchanged after reproduction.	
12	**Documentation of reproduction processes and outputs**	
12.1.1	Records in order to be access will require to be produced the existing software format or rendered or re-produced in another format such as XML or PDF. Documenting the reproduction process and its effects is an essential means of demonstrating that the reproduction process is transparent (i.e., free from pretence or deceit). Such transparency is necessary to the effective fulfillment of the preserver's role as a custodian of the records. Documenting the reproduction process and its effects is also important for the users of records since the history of reproduction is an essential part of the history of the record itself. Documentation of the process and its effects provides users of the records with a critical tool for assessing and interpreting the records.	
12.1.2	In those cases where a copy of a record is known not to reproduce the elements expressing its identity and integrity fully and faithfully, such information should have been documented by the custodian and this documentation has to be readily accessible to the authorised user.	
12.1.3	Reproduction includes viewing both rendition in the original software format and viewing of renditions in alternate formats determined as needful by the custodial organisation The activity of reproduction has to document: • the date of the records' reproduction and the name of the responsible person;	

	TNA adaptation for CHM	Comments
	• the relationship between the records acquired from the creator and the copies produced by the custodian; • the impact of the reproduction process on their form, content, accessibility and use.	

Section 3: Sustaining authentic and reliable records: technical requirements

	TNA adaptation for CHM	Comments
1	***Introduction***	
1.1.1	Records to be sustained 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 section 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 Section 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	N/A	
1.1.5	N/A	
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	N/A	
1.1.8	Any enterprise 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 Sections 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.	N.B.
2	**Ingest—Importing across platforms**	
2.1.1	Electronic documents and records will require to be ported 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 create a visible warning and generate an exception report detailing the specific exceptions. The points listed below break the required information into specific categories 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	

	TNA adaptation for CHM	Comments
	▪ 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 conforms to applicable standards, 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 archive managers and 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 its administrator that any error has been detected and that a report has been generated.	
2.2.7	Every holding in the collection should be accompanied by its provenance, which can be construed to be an audit trail documenting every known significant event in the creation of this holding in its preserved form. In cases in which it is impractical or too expensive to document some significant creation or update events, this fact should be noted.	

	TNA adaptation for CHM	Comments
2.2.8	Every audit trail element should identify who created it, when this was done, and any other creation circumstances that eventual information consumers are likely to value as historical evidence.	
2.2.9	The audit trail for any holding should be reliably bound to the holding.	
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 all sources of degradation, at least 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	The system must maintain at least two copies of each object and its metadata, doing so locally and in remote repositories, so as to make the probability of the loss of any set of holdings extremely unlikely.	
3.2.2	See 3.2.1.	
3.2.3	The system must have the ability to reconstruct its indices should they be corrupted by any event.	

	TNA adaptation for CHM	Comments
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	Prudent backup procedures are widely known, and should be adopted.	
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.	
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.	If the new copy is identical to what it replaces, no metadata change should be needed.
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	See 3.2.1.	
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 the effects of media degradation**	
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.	

	TNA adaptation for CHM	Comments
	• 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	N/A	
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 long-long term preservation are not mature. The system should provide for such new methods as might appear and promise to improve significantly the safety and usability of the repository holdings.	
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	If format migration is used, 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	If format migration is used, each migrated record should be saved 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.*	[OAIS] defines distinct ingest, archival, and dissemination formats.
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*	This will be discussed in [Gladney 5].
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*	

	TNA adaptation for CHM	Comments
	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:	

	TNA adaptation for CHM	Comments
	in their native format, or a current format to which they been migrated and in order of preference; • an XML format which conforms to applicable standards, 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:	

	TNA adaptation for CHM	Comments
	• 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 conforms to applicable standards, 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.	

Section 4. Guidance for categorizing records to identify sustainable requirements

	TNA adaptation for CHM	Comments
1	***Summary***	
1.1.1	See Section 3, 1.1.1.	
1.1.2	This section provides high-level guidance for conservators seeking to categorise institutional records to scope the specific nature of the requirements needed to sustain these record categories as authentic long-term records.	
1.1.3	N/A	
1.1.4	N/A	
1.1.5	N/A	
1.1.6	See Section 3, 1.1.6.	
1.1.7	N/A	
1.1.8	See Section 3, 1.1.8.	
1.1.9	See Section 3, 1.1.9.	
2	***Introduction***	
2.1	**Purpose**	
2.1.1	Redundant.	
2.1.2	Records produced during the course of operational business will be used in different ways for varying lengths of time. The inherent differences between various records means that multiple strategies will need to be developed for sustaining records throughout their retention period. The differences in the records also needs to take account of how their use might change as this will affect their sustainable requirements.	
2.1.3	The sustainable requirements for particular records are associated with the need to retain them for operational and other business uses. This means that there is a need to be able to categorise records in a way that reflects how the records are used and will be able to be used as a broad measure of their different requirements. This guidance takes a high level view about the development of record categories although in practice it might be necessary to develop more discrete groupings. Through developing these categories it will be possible to ensure that the records of most value to the department are not compromised and will be fit for purpose.	
2.1.4	Redundant. Furthermore, see [Gladney 9].	
2.1.5	The costs of sustaining or preserving records for long periods are potentially high even if the overall storage costs appear to be low. Once the profile of a category is established which clarifies the elements that are needed to preserve the records as reliable, authentic and usable assets it will also be possible to identify the overall costs and resource implications of applying a particular maintenance strategy to a given category of records.	
2.2	**Benefits**	
2.2.1	The benefits of adopting and implementing sustainable strategies to targeted record categories can be summarised as follows: • identification of the known or potential use of the records and how this may change over time. • identification of the level of reliability required if the records are to be fit to meet the known business and operational use. • identification of the requisite qualities that need to be maintained if the records are to demonstrate a meaningful degree of integrity. • identification of the changing usability need to present and interpret the record in an intelligible manner. • the ability to justify need and allocation of resources into sustaining particular record categories. • the ability to determine when the sustainability requirements of a set of records might change and work out the implications this might have in terms of risk and resources.	

	TNA adaptation for CHM	Comments
	• the ability to predict where resources will need to be allocated according to changes either in software or in terms of machinery of government changes to ensure records are sustained to the appropriate level of authenticity. • identification of the risks and consequences involved if the records are not sustained. – e.g. are the benefits of sustaining greater than the benefits of not sustaining? • identification of the resource requirements and concomitant costs needed to sustain particular record categories to a defined level of quality.	
2.3	**Audience**	
2.3.1	This methodology is designed to help conservators and others charged with record management responsibilities to scope the profile and volume of the high level categories of electronic records held by their organisations, which are generating a need for sustainable requirements.	
2.3.2	There will be other stakeholders in the organisation who will participate in the assessment of electronic record collections, from an operational, business or IT perspective. All concerned should ensure consistency with the organisation's corporate policy and procedures, and general working practices.	
2.3.3	N/A	
3	***Developing a strategy***	
3.1	**General**	
3.1.1	This document is intended to assist departments to clarify the record attributes that need to be sustained over time. These in turn will help identify broad record categories and the resource requirement needed to sustain the records to a standard appropriate for the duration of the continuing business need.	
3.1.2	The rigour with which sustainable requirements need to be applied will not be the same for all records as the length and type of business and operational use will not be the same. The differences in business and operational use will affect records in a way that will affect their need for authenticity, for example records used in court proceedings need to have a higher level of authenticity than those used for research purposes.	
3.1.3	In developing a sustainable records strategy account needs to be taken of the differences of the length and type of use of different records. It would be impractical to do this at the level of individual records and unnecessary as many records have similar features. This means that it is necessary to determine how records can be categorised at a high level in a manner where their characteristics reflect similar sustainable requirements	
3.1.4	The nature of these categories will vary according to the nature of a department's activities and roles but it is to be emphasised that a broad-brush approach needs to be adopted to scope the sustainable requirements. Looking at every record type created within an organisation and trying to establish a preservation need at that lower level is unnecessary as it is felt that this would be an expensive and problematical approach. Effective risk evaluation ultimately is critical to the success of a corporate strategy to maintain sustainable records.	
3.1.5	Section 1 will help first assess the value of a record category to the organisation and their relationship with other records across the organisation. Having established the broad need it then becomes necessary to identify the precise elements that need to be preserved to maintain these records as authentic records as defined by *BS ISO 15489 Information and documentation – Records management standard.*	
3.1.6	Redundant.	
3.1.7	N/A	
3.1.8	Having identified the record characteristics that need to be maintained it then becomes necessary to identify the required resource requirements and overall costs of maintaining records to that degree. Section 8 posits some of the questions that need to be consider at this stage.	
3.1.9	Having undertaken this analysis it is then recommended that the enterprise apply a risk evaluation methodology to ascertain the risk of not maintaining records to the recommended degree. The final outcome should provide a robust basis for decisions for the development and implementation of sustainable strategies.	
3.1.10	It should be emphasised that the questions proposed in this document do not represent	

	TNA adaptation for CHM	Comments
	a comprehensive or an exhaustive list nor would it be necessary to define responses to all the questions referenced in this document.	
4	**Assessing the Value of Records**	
4.1.1	N/A	
4.1.2	A record category's relationship with other records is also a factor which can increase or decrease the value of the records. When used in combination with other records the value of a category may increase. The category might also duplicate or overlap with another record category in such a way that the repository needs to keep one of category.	
4.1.3	To assist the enterprise in identifying the value of the records to the business two areas need to be considered. These are: • content and business use identifying the value of the material based on the record category alone and, • relationship with other records assessing the material in the context of other, related record categories	
4.1.4	Uninteresting because it is obvious in the light of other objectives.	
4.2	**Content and business use**	
4.2.1	The questions below are proposed to help identify the business need • To what extent are these records needed to document the history of decisions taken/actions carried out for future operational use? • How important is the user's/creator's continuing need for this information in the future? • To what extent are these records needed to fulfill legal requirements? • What implications for accountability arise from a decision to dispose of these records?	
4.3	**Relationship to other records**	
4.3.1	The questions below are proposed to help identify the business need • To what extent do the records in this category support the interpretation and use of other records? • What value do these records add to a wider set of information? • To what extent do these records have a logical relationship to other record categories that are being kept? • If these records are derived from a wider body of information, how much value do they add to the original information? • If these records contain personal data, to what extent does their retention create a risk, and to whom?	
5	***Identifying the requirement for reliability***	
5.1.1	N/A	
5.1.2	The characteristic of reliability itself can be broken down into three sub elements. These are: trust, relationship/context, longevity.	
5.2	**Trust**	
5.2.1	Trust is critical to reliability as without it there can be no meaningful faith in the accuracy of the retained records. The issue here is not so much the precise characteristics of an individual document as the characteristics of the records of an activity or transaction, which have to be maintained if the records are to continue to be serviceable. The questions that need to be addressed in order to substantiate trust are: • What makes up or constitutes the record? (i.e. what is it that has been captured that is critical to the business) • Who was the creator and how are they identified? (what are the critical elements just the individual's identity or name or the profile allocated within the ERMS at the time of the creation of the record? – the profile may be required to confirm if the officer named possessed the appropriate authority to undertake or authorise the transaction) • Which dates have been captured in relation to the creation and modification of the records and which are significant? (i.e. what if any are the critical stages of the work	Precision is needed. Who is to be trusted for what? This consideration affects what should be provided in metadata for provenance recording.

	TNA adaptation for CHM	Comments
	process or transaction which have been captured)	
5.3	**Relationship/Context**	
5.3.1	Comprehension of the meaning and value of records relies upon the ability of the reader to place the records in their operational context in a manner that their relationships with other affected records are clear and transparent. Again it is not so much the precise characteristics of an individual document that should be considered as the characteristics of the records of an activity or transaction. Here it is the links and relationships with other records and the location of these within the business classification schema, which need to be considered. It should be noted that the adoption of the records management metadata standard provides for the creation and management of relationships under the element Relation. The existence of pointers within an ERMS giving multiple locations is also relevant here. The questions provided below will help determine what contexts or relationships must be maintained for the records to be considered reliable. • What is the scope of the records and what do they cover? (e.g. in the case of case records or transactions an understanding of the business process and possibly the legal or regulatory context in which the records were created is essential to understanding them over time) • Which records would be maintained in the same vicinity of the classification schema or file-plan, which are critical to the understanding of the activity? • Which other significant records were produced in conjunction with the records of the activity under consideration? • How long do these relationships continue to be meaningful? • What cross-references or pointers exist and what is the relevance of the link between the two sets of records?	N/B. This consideration affects what should be provided in metadata for provenance recording.
5.4	**Longevity**	
5.4.1	Longevity refers to the duration of the period for which the business still depends on the records to fulfil a residual business need. The requirement for reliability may differ according to the different types or categories of records created and held by a department. Establishment of this sub-element will assist in clarifying the requirement for maintaining the characteristic of integrity, which follows in the next section of this document. • How long are the records used by the business centre that creates and manages the records? • How often are the records updated while they are open? • When are the records considered to be closed? (i.e. no longer updated)? • How frequently, by whom and for what purposes are the records referred to once they are closed? (this helps identify the scale and nature of the continuing access requirement) • What makes up the record and which parts are considered to be dispensable if any? • Which dates or other information would be captured subsequently in relation to the modification of these records? (e.g. requirement to amend following a data protection subject access application)	
6	**Identifying the requirement for integrity**	
6.1.1	Redundant.	
6.1.2	What has to be determined is what gives a record category its required level of integrity and how might this differ across the various categories of records identified.	
6.1.3	The characteristic of integrity itself can be broken down into four sub elements. These are: traceability, retention periods, applicable rules, standards and regulations, risk.	
6.2	**Traceability**	
6.2.1	In order to confirm the record is unchanged or that only authorised and appropriate changes have been made, the status of the records and the presence or absence of change has to be auditable or traceable. The questions that follow can be used to scope the both need and the degree for auditable information • For audit purposes, what are the minimum requirements of events to be recorded? For example,	

	TNA adaptation for CHM	Comments
	▪ changes in access provisions ▪ additions to records (e.g. annotations and modifications) ▪ movement history (e.g. exports or imports due to transfers of function or re-classifications within the business classification schema) ▪ who has accessed the record and when ▪ formats into which the record has been rendered, how this was achieved and ▪ why changes in retention periods, why and when this was done • How long would the events captured in the audit trail be needed for business purposes? • Is there a need to maintain a record of the decisions relating to the access permissions applied to the records? • When will a review of access provisions and permissions be required and what type of notification will be required? • Is there a need to maintain an ongoing record of who has been permitted to have access to the records and the dates relating to the period of permitted access? (note: this is separate from a record of changes in access permissions) • Is there a need to maintain a record of who has been permitted to modify the records? (note this is separate from a record of authorised changes or modifications)	
6.3	Retention periods	
6.3.1	As integrity is bound to the need to demonstrate authenticity over time it is necessary to clarify the specific business retention requirements. In doing this it will be possible to establish the overall duration of the retention period to be applied to a category of records and clarify the profile of retention taking into account that the cost of the maintenance period is related to the length of the retention period. Where it is possible to reduce the number and complexity of the records required to be sustained without compromising business effectiveness this will help justify the business case for expenditure on sustainable strategies. • What are the retention requirements for these types of records? • How is the retention period determined? (i.e. is it specified by a regulatory or legal requirement) • Are there some records in this area where parts of the records have longer retention periods than the rest of the record? (e.g. certain key documents relating divorce decrees are retained for 75 years whereas the bulk of the material relating to a case are deleted after 25 years) • Are there examples where it is more appropriate for a subset or abstract of the record of a transaction, rather than the whole record, to be retained for a longer period? (e.g. summary of employment service retained for superannuation purposes)	
6.4	Applicable rules, standards and regulations	
6.4.1	In certain instances it may be necessary or desirable to retain records related to a broad record category where the records were themselves generated in response to codes of instruction or standards in force at that time. In order to confirm if the record of a transaction was valid in these circumstances it may be necessary to reference the rules that applied at that juncture. For example a query or claim for an entitlement to a benefit may only be validated by crosschecking the standards that were extant at the time the adjudication was made. • Is there a need to maintain the requirements and standards needed when considering for how long the records should be maintained? • Is there a need to maintain the requirements or standards relating to the maintenance conditions of the records? (i.e. certain key document types may be explicitly referenced. For example, the server on which the records were stored, maintenance operations conducted on server, architecture of ERMS)	
6.5	Risk	
6.5.1	The issue of record integrity is closely linked to effective business continuity planning in that in order to clarify the cost of maintaining record integrity it is necessary to evaluate the risk to the organisation if the records have been retained as incomplete or with limited auditable functionality. The following questions are intended to help identify the scale of the risk to the organisation if integrity is compromised. This in turn will help cost justify the selection and application of specific sustainable strategies.	

	TNA adaptation for CHM	Comments
	• What are the potential problems if the records are not available over x, y, z number of years? • What are the potential risks of not effectively disposing of the records at the correct time? • What are the potential problems if access controls and permissions are not properly maintained? • What are the potential consequences of inaccurate information? • What are the potential risks of not knowing where related records are located? • What areas of the business would be of particular concern in relation to risk and contingency management and which records are considered to be the vital and/or the emergency records?	
6.5.2	N/A	
6.5.3	N/A	
6.5.4	N/A	
7	Identifying the requirement for usability	
7.1.1	The requirement for usability may appear superficially the easiest to scope and comprehend particularly where the records either consist of images or text. Providing the appropriate viewer or browser is available the end users should have no difficulty accessing the record. The issue can then appear to revolve around the availability and presence of the appropriate viewing software. However, the issue is more complex than the previous analysis might suggest as usability is also about ease of locating, quick retrieval and the quality of the presentation. The first question to posit is: What makes a record usable and how might this differ according to different types of records? Four sub-elements then need to be considered in evaluating the requirement for the usability of records over time. These are locating, retrieval, presentation, interpretation.	
7.2	Locating	
7.2.1	Locating refers to the means used to reliably identify without undue difficulty the record or records needed to satisfy the user's query. The location within the business classification schema or file-plan is one aspect but also the issue of accurate titling, meaningful nomenclature and the use of aliases or alternative titling fall into this area. • How are the records titled? • What cross references/pointers are also required to be maintained and how are they made visible? • What is used to show the location of records within the business classification schema or file-plan? • What thesaurus terms are used and are these industry standard or user defined (the latter need to be identified if they are to be maintained)?	
7.3	Retrieval	
7.3.1	Effective retrieval is dependent upon the anticipated pattern of access demand and the application and continued management of appropriate access permissions.	
7.3.2	N/A	
7.3.3	In practice the business requirement will mean the first two options are preferred for their greater convenience. However, the economics of far-line and off-line storage may be very attractive if the use of the records is estimated as being residual and very infrequent.	
7.3.4	Access requirements can be characterised by the following estimates: • Total number of retrieval requests in a given period • Average number of documents requested • Average total size of request in megabytes • For databases, cost of database query (rows retrieved or examined) • Anticipated methods of retrieval (e.g. use of keywords, full text indexes and thesauri)	
7.3.5	Access must be capable of being defined in response to an organisation's business needs and is likely to vary according to the organisation's information requirements.	
7.3.6	The critical elements for an effective access strategy can be summarised as follows: • Identify who can make requests, and who can execute them • Understand management parallels with paper records	

	TNA adaptation for CHM	Comments
	• Beware of using past access to predict future access patterns • Take disaster recovery into account in planning	
7.3.7	The questions to be asked to ascertain the retrieval and access requirements are: • Are the access permissions likely to change over time? • What type of access permissions would be set? • What prompts would be set for access permissions to be changed? • What level of frequency of access is required? • What is your policy on encryption and password-protected documents/objects—are these routinely removed upon capture into the EDRMS?	
7.4	Presentation	
7.4.1	Effective presentation ensures the user can retrieve and view the records with the appropriate level of functionality required to undertake a meaningful interpretation. In some instance this may require the original program to be available so that the data can be manipulated or edited using the same functionality to create a new document or version, which can then be saved and added to the corporate record without changing or deleting the original. In other cases it may be sufficient to view the image in a more static environment either by using viewer technology or be generating a rendition, which is a faithful image of the original.	
7.4.2	Different groups of users may have different presentation requirements. In some cases a small group may need the original functionality when viewing the record. The cost of supporting such a service may not be too onerous for a small group of specialised users but excessive for the whole organisation where the opportunity of viewing a rendition would normally be a satisfactory alternative. The technology used to interface and view the record must therefore reflect the ongoing business need. The questions that need to be addressed are: • What form do the records currently take, what format are they associated with, word, excel, spreadsheets, word processing, slides, html? • What level of presentation is essential to enable the users to undertake the work anticipated and required by the business?	
7.5	Interpretation	
7.5.1	Interpretation at its simplest can be addressed by an ability to view text or images using a simple browser without the enhancements offered by the original software, for example one can view documents created in MS Word using a text file viewer such as WordPad although the formatting is lost in this view. In other circumstances seeing the content without the display and formatting built into the original document makes interpretation difficult if not impossible. If, for example, a respondent has cited a specific paragraph or entry of a code of instruction by its original number as the authority for undertaking an action or receiving an entitlement and this data is not visible to the user in the business, it will not be possible to either confirm or deny the validity of the claim. This type of information is often built into the format display properties of the software in which the document was originally conceived and can only be viewed either if the original program is available or an appropriate rendition, which has captured this detail, has been created and maintained.	See [Gladney 2].
7.5.2	In other instances interpretation also needs to be supported by linked contextual information, for example the ability to view the metadata of the record in both its original and existing context. This may require users having sight of both the current business classification system in which the records reside and the original classification system where that differs from the current version. This situation can arise where functions have been transferred between government bodies resulting in bulk exports and imports of metadata and data between EDRM platforms. In those cases a portion or subset of the earlier or original classification system will have been transferred before the records are relocated in the new business's classification system. Maintaining a copy of the original classification system can assist understanding of the full context in which the records were created and used as well in addition to how they are seen in the current classification plan.	Linked information is likely to be needed for almost every archival record.
7.5.3	The questions that need to be addressed are: • What is it about the document that will require interpreting, the content, the presentation or both? • What level of contextual information is essential to a full understanding of the	

	TNA adaptation for CHM	Comments
	records?	
8	Assessing resource implications	
8.1.1	Departments need to identify and quantify the resource implications required to maintain an existing record collection. This section is provided to help departments to undertake this process by supplying the key questions that need to be asked in order to clarify the overall resource requirement of continuing to apply a sustainable strategy to a record category.	
8.1.2	Departments need to compare the value of the records to the business suggested at section 1 with their concomitant requirements for reliability, integrity and usability proposed at sections 5, 6 and 7 against the cost and resource implications of applying a strategy which will secure the value and the authentic properties of the records. Extended or indefinite storage of electronic records does incur significant overheads and recommendations to either dispose of, or retain, a category of records will be informed by this knowledge.	
8.1.3	N/A	
8.1.4	N/A	
8.1.5	The questions that need to be asked at this juncture are: • Should the records be reviewed for sensitivity? • Are these records accessible via the current hardware/software platform? • If accessed via their current platform, will the records continue to be accessible on this platform for the short term? (1 to 2 years) • If accessed via their current platform, will the records continue to be accessible on this platform for the medium term? (3 to 5 years) • What percentage of the records require migration in the short terms (1 to 2 years) to a different software format to retain access? • What percentage of the records require migration in the medium terms (3 to 5 years) to a different physical format to retain access? • Are there specific difficulties in migration due to e.g. proprietary formats, non-standard design structures? • Should the records be sampled to verify technical decisions?	
9	Compliance assessment – evaluating the implementation	
9.1.1	All sustainable strategies should be subject to regular review to assess their relevance and effectiveness. Department's need to assess whether the assumed pattern of use and concomitant retention requirements are still appropriate taking into account the perceived business benefit, the contingency requirements and the overall cost of continuing to apply the selected strategy. It is recommended that the interval for evaluating these strategies should not exceed 5 years.	

Discussion

Why Change the TNA Requirements Statement?

I believe that the TNA document needs to be refined and extended if it is to be used as a requirements statement for a future CHM software collection repository and service to remote CHM visitors. This is partly because CHM is a much smaller institution than TNA, partly because much of the TNA document reads more as management objectives than as technical requirements, and partly because the TNA document seems to presume a solution-class that is neither the only possibility nor (in my opinion) the best possibility.

The institutional size difficulty is that the TNA document implicitly calls for quite a bit of hands-on repository management with rule and procedure definitions. Compliance might be unaffordable by CHM and therefore unsustainable with demonstrable reliability.

The management objectives difficulty is that many of the TNA line item statements seem to be subjective (requiring human translation into executable steps) rather than objective (specifications for which compliance can be objectively demonstrated or tested). Further work would be required to extend them to be a guide for software engineering.

The solution-class difficulty is that the TNA document seems to presume that the only way to create and manage a reliable long-term software collection is by repository management rules (as suggested by *Trusted Digital Repositories: Attributes and Responsibilities* [RLG] and closely related proposals). [Gladney 1] and related papers show an alternative and argue its superiority.

Details about the Changes to TNA documents for CHM Use

In the top level sections above, each objective was derived from the corresponding TNA objective. The pattern of changes made is suggested by the following considerations.

- Four TNA documents were collapsed into one document, with consequent minor changes. Some TNA statements have become redundant.
- The TNA document is written as if the so-called "Trusted Digital Repository" approach were the only way, at its level of description, of preserving digital information reliably for the long term. This assumption has caused numerous requirements to specify methods for which there are, arguably, better alternatives.
- The documents cited in the bibiliography below themselves contain numerous citations that provide entry points into the scholarly literature on digital preservation. The worthwhile part of that literature is expressed in 100 to 200 papers and unpublished reports. (The digital preservation literature is between 3 and 5 times that large. However, it is also full of redundancy.)
- TNA documents allude to 'procedures' to be followed.[5] Many of these are not procedures in the sense usually understood by software engineers (in which each specification has been, or could be, reduced to a sequence of mechanical steps that can be executed in finite time by some machine or clerk), but are instead subjective criteria for the behavior of repository employees.
- In some places, the TNA document depends on the notion of 'essential' information without defining an effective procedure for deciding which aspects of a deposited document are essential and which are accidental. (This distinction is, to some extent, discussed in [Gladney 4]. It will be carefully analyzed in a paper that colleagues and I hope to have ready later in 2005. [Gladney 5] The challenge is to capture, in objective form, the intentions of each information producer.)
- Content management technology is well understood, and many satisfactory software offerings support what is required. In contrast, digital preservation is widely considered to be a research topic. The TNA document does not distinguish these topics as clearly as would be helpful to managers and implementors.
- 'Statutory' has been replaced by 'legal' throughout, as 'statutory requirements' is a subset of 'legal requirements'.
- The phrase 'trusted custodian' has been replaced by 'custodian', since the presumption that any custodian is fully trusted is questionable.
- I feel that the TNA audit trail statements are not strong enough for a repository that might hold records tempting for malfeasance. Some addition has been made, but this should not yet be regarded as complete. (See, for example, Section 3, 2.2.7 ff.)

Next Steps for the CHM Software Collection Committee

What might someone a century from now want of information stored today? The figures on pages 1 to 2 help us discuss preservation reliability. In addition to what content management offerings and published metadata schema already provide, a complete solution would:[6]

- Ensure that a copy of every preserved document survives as long as wanted;
- Ensure that authorized consumers can find and use any preserved document as its producers intended, avoiding errors introduced by third parties;

---

5 Section 2, item 2.1.1 provides an example of this problem.

6 The first part of this subsection is copied almost verbatim from [Gladney 1].

- Ensure that any consumer can decide whether information received is sufficiently trustworthy for his application; and
- Hide technical complexity from end users (information producers, information consumers, and also archive managers).

Viable solutions will allow repositories and their clients to use deployed content management software without disruption.

My colleagues and I believe that the design of a digital preservation system should protect against all accidents or misfeasances that might jeapordize the authenticity or reliablity of any archived record, to the extent that doing so is feasible. If this is accomplished, then the repository infrastructure will be "strong" enough for any kind of holding, including records whose adulteration or destruction is a tempting target for fraud. Of course, such an attitude would be impractical if it resulted in a solution that was significantly more costly than some practical alternative. Happily, the objective is not only feasible and likely to be cost-effective, but is likely to be less expensive than any alternative for which a design has been proposed!

Structure for a SW Repository Statement of Requirements

Each statement in such a future document should be such that purported compliance can be objectively tested with little risk of controversy. The statement of requirements might best be structured into sections that deal with:

(1) What is needed to please the eventual information consumer. (See the figures to understand 'information consumer' and other human roles.)
(2) What is needed to please any information producer (either an author or and editor), over and above requirements identified in (1).
(3) What is needed to make archive (repository) managers productive, over and above requirements identified in (1) and (2).
(4) What should be provided by the document storage subsystem (the inner core of a digital repository), which is hidden from any user and should be useful for any repository institution.
(5) What should be provided by the next repository level ('archival storage' in Figure 3), which should contain most of the institution-specific repository software needed.
(6) What needs to be provided by a combination of the next level of repository software and by repository managers (people), and is not already specified by responses to (1) through (5).

To the extent possible, the individual statements of requirements should avoid solution design assumptions, except for statements of conformance to international standards (e.g. [METS] and assumptions about the use of widely accepted software (e.g., XML).

Digital repository (a.k.a. content management or 'digital library') technology is well understood and represented by highly refined implementations. It is therefore prudent to distinguish repository requirements that apply even for content that need not be durable for the long term from the additional requirements needed for long-term digital preservation.[7]

Suggested Action by CHM Software Collection Committee Members

Because the current document reads more as a statement of management objectives than as a statement of technical requirements, we should both refined it and also prepare a software requirements document.

Readers are urged to consider carefully each numbered item in Section 1 through Section 4 above with the following questions in mind, always with respect to what CHM realistically will need.

- Is the objective a sufficient treatment of the topic it addresses? Alternatively, is the objective too rigorous?

[7] Throughout this document, 'long-term' for a record should be construed to mean 'beyond the time when the authors and editors of the record might be available to clarify confusions or authenticity questions'.

- Is the objective an objective assertion? If not, can it be reformulated so that compliance can be objectively judged?
- What major management objectives have been overlooked?
- Is the objective stated in a form that the CHM would be pleased to submit for review by external commentators?
- What specific software requirements are needed to satisfy the management objective?
- What changes or extensions are needed for my Introduction and Discussion sections?
- What important citations have been overlooked?
- How might the requirements taxonomy in *Structure for a SW Repository Statement of Requirements* be improved?

Please consider also what the next steps should be and what target completion date we should set for ourselves for each such step.

Bibliography

[Gladney 1] Gladney, H.M. *Principles for Digital Preservation*, Comm. ACM, to appear in 2005. Longer version abstract at http://arxiv.org/abs/cs/0411091 and paper at http://arxiv.org/ftp/cs/papers/0411/0411091.pdf

Published difficulties with long-term digital preservation prove to be largely confusions with language. Similar difficulties were addressed in early twentieth-century philosophy. We describe prominent confusions, show how to clarify the issues, and summarize a 'Trustworthy Digital Object (TDO)' method that solves all the technical challenges described in the literature. Other TDO reports provide detailed design and analysis of the TDO method.

A purpose of this article is to invite searching public criticism before anyone invests significant resources in creating preservation data objects.

[Gladney 2] Gladney H.M. Lorie, R.A. *Trustworthy 100-Year Digital Objects: Durable Encoding for When It's Too Late to Ask*, expected to appear in ACM Trans. Info. Sys., 2005. First draft, June 2003. Preprint available at http://eprints.erpanet.org/archive/00000007/. Abstract at http://arXiv.org/abs/cs/0411092 http://arxiv.org/ftp/cs/papers/0411/0411092.pdf

How can an author store digital information so that readers can surely understand and use it as he intends?

We present a solution in a way that readers not steeped in computer programming can readily understand. The core idea is the feasibility of specifying a "universal virtual computer" (UVC) that is extremely simple, but that still can handle any computation whatsoever. Since we can specify such a machine correctly in every detail, our descendants will be able to execute its programs on the computers of their time. It is easy to provide a UVC program today to interpret whatever information we want to be useful in the future.

Such a general solution might be more elaborate than needed to preserve data such as e-mail, image, audio, and video files. Sufficiently simple files can be preserved by encoding them in conformance with well-known standards. We sketch convenient methods for files ranging from very simple structures to those containing computer programs.

Methods that depend on something that might work in the future are not good enough. We demand a method that is sure for any data saved today. Prior proposals—called "migration" and "emulation"—fail because what they save depends on knowledge about today's information technology. In contrast, our proposal depends on avoiding such irrelevancies in preservation bitstreams."

[Gladney 3] Gladney, H.M. *Trustworthy 100-Year Digital Objects: Evidence After Every Witness is Dead,* ACM Trans. Info. Sys. 22(3), 406-436, July 2004. May 2003 version available at http://eprints.erpanet.org/archive/00000008/.

How can a publisher store digital information so that any reader can reliably test its authenticity, even years later when no witness can vouch for its validity? What is the simplest security infrastructure sufficient to protect and later test evidence of authenticity?

In ancient times, wax seals impressed with signet rings were affixed to documents as evidence of their authenticity. A digital counterpart is a message authentication code fixed firmly to each important document. If a digital object is sealed together with its own audit trail, each user can examine this evidence to decide whether to trust the content—no matter how distant this user is in time, space, and social affiliation from the document's source.

We suggest technical means for accomplishing this: encapsulation of the document content with metadata describing its origins, cryptographic sealing, webs of trust for public keys rooted in a forest of respected institutions, and a certain way of managing document identifiers. These means will satisfy emerging needs in civilian and military record management, including medical patient records, rEgulatory records for aircraft and pharmaceuticals, business records for financial audit, legislative and legal briefs, and scholarly works.

This is true for any kind of document, independently of its purposes and of most data type and representation details, and provides each user with autonomy for most of what he does. Producers can prepare works for preservation without permission from or synchronization with any authority or service agent. Librarians can add metadata without communicating with document originators or repository managers. Consumers can test authenticity without Internet delays, apart from those for fetching cryptographic certificates.

Our method accomplishes much of what is sought under labels such as "trusted digital repositories", and does so more flexibly and economically than any method yet proposed. It requires at most easy extensions of available content management software, and is therefore compatible with what most digital repositories have installed and are using today.

[Gladney 4] Gladney, H.M. *Trustworthy 100-Year Digital Objects: Syntax and Semantics—Tension between Facts and Values,* 2003 preprint at http://eprints.erpanet.org/archive/00000051/

Prior *Trustworthy 100-Year Digital Object* articles describe a method for preserving digitally represented information. Trustworthy Digital Object (TDO) representation and packaging makes any digital content reliably meaningful to consumers, no matter how distant these are in time, in space, and in social affiliation from their information sources. The current article focuses on digital document authenticity and on evidence a consumer can use to decide whether to trust the content.

Such considerations are necessarily epistemological. Arguing the issues must start by conveying as unambiguously as possible what we mean by words like 'authenticity' and 'evidence' and by distinguishing between statements that are 'objective' and those that are 'subjective'. Our analysis applies Wittgenstein's teaching to pictorial models of digital and conventional communication.

This analysis leads us to identify an ethical imperative for digital preservation, and to suggest that the TDO method defines a quality standard against which any method of digital preservation can be judged.

[Gladney 5] Gladney, H.M. Bennett, J.L. Lucas, P. *Trustworthy 100-Year Digital Objects: What's Meant? Intentional and Accidental in Documents,* work in progress, 2005.

[Gladney 9] Gladney, H.M. Bennett, J.L. *What Do We Mean by "Authentic"? What's the Real McCoy?* D-Lib Magazine 9(7), July 2003.

[Gladney 11] Gladney, H.M. *Preserving Digital Records: A Method Guided by Scientific Philosophy*, preprint, Nov. 2004.

Preserving digital information has received steadily increasing attention since 1995. However there has been little substantial progress towards resolving the key technical challenges: first, ensuring the future ability to use producer's information with computers whose design cannot today be known; and second, creating durable evidence so that each future user can prudently decide whether to trust saved information.

We outline a solution to these challenges—a solution that we call the Trustworthy Digital Object (TDO) method. TDOs provide reliable packaging for any type of digital object, no matter how distant its eventual recipients are in time, space, and organizational affiliation from the information sources. Each preserved object carries its own provenance audit trail. Information producers can prepare documents for archiving without help or permission from anyone. Archivists can add metadata without communicating with producers. Consumers can test the authenticity of preserved documents without human assistance.

Deep analyses and searching peer critiques are important as preludes to implementation and deployment of any proposed digital preservation solution. Early twentieth-century philosophy and pictorial models help clarify dilemmas expressed in Archivaria articles and elsewhere. Our analysis leads us to suggest that the TDO method achieves technical quality against which any method of digital preservation should be judged.

[IFLA] International Federation of Library Associations and Institutions, *Information Policy: Copyright and Intellectual Property*, 1999.

[IFLA 2] International Federation of Library Associations and Institutions, *Study group on the functional requirements of bibliographic records*, http://www.ifla.org/VII/s13/frbr/frbr.pdf

[Lorie 04] Lorie, Raymond Armand. *Long term archiving of digital information*, U.S. Patent 6,691,309, February 10, 2004.

Digital data is preserved by ***archiving*** on a removable medium. In the long term, the save data bit stream must be correctly interpreted. For a computer program or system to be archived, the bit stream constituting the program must be archived and the code must be executable at restore time. The program that restores the data does not "see" the contents of the data itself, but accesses it by issuing a function call to an executor. A description of which methods are available to restore the information hidden in the data is always available. A text tells the client which functions are available and what their purposes are. The ***archiving*** method is based on using a virtual computer instruction set and saving the algorithm as a program written int hat virtual machine language. For machine instructions to be executed many years later, for example 100 years, an emulator of the original machine would be written on the future hardware. Any machine manufactured in the originating year would develop for each architecture a Universal Virtual Computer (UVC) description of the machine. Each originating instruction would be mapped into a small program of UVC instructions. All manufacturers of new architectures would then have to write a UVC executor which would be able to execute UVC instructions on the machine running 100 years in the future.

[METS] *Metadata Encoding and Transmission Standard* (METS), 2001-2.

The METS schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the XML schema language of the World Wide Web Consortium. The standard is maintained in the Network Development and MARC Standards Office of the Library of Congress, and is being developed as an initiative of the Digital Library Federation

[OAIS] CCSDS 650.0-R-2, *Reference Model for an Open Archival Information System (OAIS),* Red Book, Issue 2, July 2001. The ISO draft is at http://www.ccsds.org/RP9905/RP9905.html. An overview of the development of OAIS is at http://ssdoo.gsfc.nasa.gov/nost/isoas/us/overview.html.

See, for example, the description of the use of the reference model in recent D-Lib Magazine articles on the NEDLIB project <http://www.dlib.org/dlib/september99/vanderwerf/09vanderwerf.html>, in the British Library <http://www.dlib.org/dlib/november00/brindley/11brindley.html>, and at the National Archives and Records Administration <http://www.dlib.org/dlib/february01/thibodeau/02thibodeau.html>.

The *Open Archival Information System (OAIS)* is a high-level reference model developed by the Consultative Committee for Space Data Systems with representatives of the leading space science agencies in North America, Europe and Japan.

The OAIS reference model provides a unifying set of concepts for an archive. It consists of an organization of people and systems that has accepted responsibility for preserving information and making it available to a designated community. The OAIS model provides terminology and concepts for describing and comparing the architectures and operations of archives, defines the responsibilities of an open archival information system, and offers detailed models for the functions, components, and processes necessary to support long-term preservation and access to digital information. The model was developed to assist with the preservation of large databases of space science information.

[OAIS 2] CCSDS –651.0-W-2, *Producer-Archive Interface Methodology Abstract Standard*, June 15, 2002. This draft recommendation for a Space Data System Standard can be seen at http://bill.cacr.caltech.edu/cfdocs/usvo-pubs/files/CCSDS_651_W2.pdf.

The object of this document is to regulate the relationships and interactions between an information Producer and an Archive. It defines the methodology to allow all the actions to be structured that are required to apply from the time of contact being made between the Producer and the Archive until the objects of information are received by the Archive. These actions imply the first stage of the Ingest Process as defined in the Reference Model OAIS [1]. It describes part of the functional entities: Administration ("Negotiate Submission Agreement") and Ingest ("Receive Submission" and "Quality Assurance").

This methodology document:

Identifies the different phases in the process of transferring information between a Producer and an Archive,

Defines the objective of each of these phases, the actions that must be carried out during these phases and the expected results (administrative, technical, contractual, etc.) at the end of a phase,

Forms a general methodological framework, which should be able to be applied and reused in those processes that relate to the Producer-OAIS Archive interface. This general framework should also provide sufficient flexibility for each particular case,

Forms a basis for the identification and/or development of standards and implementation guides, in the domain in question,

Forms a basis for identification and/or development of a set of software tools that will assist the development, operation and checking of the different stages in the process of information transfer between the Producer and the Archive.

[PRO] The National Archives, *Standard for Record Repositories*, 2004. accessible via http://www.nationalarchives.gov.uk/archives/framework/. See also *Generic Requirements for Sustaining Electronic Information over Time*, 2003. accessible via http://www.nationalarchives.gov.uk/electronicrecords/reqs2002/

[RLG] Beagrie, Neil. Bellinger, Meg. Dale, Robin. Doerr, Marianne. Hedstrom, Margaret. Jones, Maggie. Kenney, Anne. Lupovici, Catherine. Russell, Kelly. Webb, Colin. Woodyard, Deborah. *Trusted Digital Repositories: Attributes and Responsibilities,* RLG-OCLC Report, May 2002. http://www.rlg.org/longterm/repositories.pdf

Other Bibliography of Interest to CHM

[Buchanan 04] Buchanan, George. David Bainbridge. Katherine Don. Ian H. Witten. *A New Framework for Building Digital Library Collections*, 2004.

This paper introduces a new framework for building digital library collections and contrasts it with existing systems. It describes a radical new step in the development of a widely-used open-source digital library system, Greenstone, which has evolved over many years. It is supported by a fresh implementation, which forced us to rethink the entire design rather than making incremental improvements. The redesign capitalizes on the best ideas from the existing system, which have been refined and developed to open new avenues through which users can tailor their collections. We demonstrate its exibility by showing how digital library collections can be extended and altered to satisfy new requirements.

[Reich 01] Reich, Vicky. Rosenthal, David S.H. *LOCKSS: A Permanent Web Publishing and Access System*, D-Lib Magazine 7(6), June 2001. See also http://lockss.stanford.edu.

LOCKSS (Lots Of Copies Keep Stuff Safe) is a tool designed for libraries to use to ensure their community's continued access to web-published scientific journals. LOCKSS allows libraries to take custody of the material to which they subscribe, in the same way they do for paper, and to preserve it. By preserving it they ensure that, for their community, links and searches continue to resolve to the published material even if it is no longer available from the publisher. Think of it as the digital equivalent of stacks where an authoritative copy of material is always available rather than the digital equivalent of an archive.

LOCKSS allows libraries to run web caches for specific journals. These caches collect content as it is published and are never flushed. They cooperate in a peer-to-peer network to detect and repair damaged or missing pages. The caches run on generic PC hardware using open-source software and require almost no skilled administration, making the cost of preserving a journal manageable.

LOCKSS is currently being tested at 40+ libraries worldwide with the support of 30+ publishers.

Archiving for permanent retention is facing some major challenges as we move into the next millennium. These include issues relating to selection from a burgeoning mass of information being produced in a wide range of formats; issues relating to media longevity and equipment obsolescence; migrating information across formats; the commercialisation of activities; the growing impact of IT requirements and the complexity of copyright and other rights in digital materials.

[Witten 00] Witten, I.H. McNab, R.J. Jones, S. Apperley, M. Bainbridge, D. Cunningham, S.J. Managing complexity in a distributed digital library, Computer - IEEE Computer Magazine, 32(2), Feb, 1999, pp 74-79

[Witten 00b] Ian H. Witten, Stefan J. Boddie, David Bainbridge and Rodger J. McNab, Greenstone: a comprehensive open-source digital library software system, in Digital Libraries 2000, San Antonio, Texas, USA, pp175-184, 113-121, June, 2000. Also in Comm. ACM 44(5), xxx-xxx, May 2000.

[Witten 01] Witten, Ian H. *How to Build a Digital Library Using Open-Source Software,* JCDL 2001. See also The New Zealand Digital Library and pubs & downloads.

This tutorial describes how to build a digital library using the Greenstone digital library software, a comprehensive, open-source system for constructing, presenting, and maintaining information collections. Collections built automatically include effective full-text searching and metadata-based browsing facilities that are attractive and easy to use. They are easily maintainable and can be rebuilt entirely automatically. Searching is full-text, and different indexes can be constructed (including metadata indexes). Browsing utilizes hierarchical structures that are created automatically from metadata associated with the source documents. Collections can include text, pictures, audio, and video, formed using an easy to use tool called the Collector. Documents can be in any language: Chinese and Arabic interfaces exist. Although primarily designed for Web access, collections can be made available, in precisely the same form, on CD-ROM or DVD. The system is extensible: software "plugins" accommodate different document and metadata types. The Greenstone software runs under both Unix and Windows, and is issued as source code under the GNU public license. Attendees will receive an extensive user manual and should learn enough to download the software and set up a digital library system. Those with programming skills should be able to extend and tailor the system extensively.

[Witten 01b] Witten, Ian H. Bainbridge, David. Boddie, Stefan J. *Greenstone: Open-Source Digital Library Software*, D-Lib Magazine 7(10), October 2001.

The Greenstone digital library software is an open-source system for the construction and presentation of information collections. It builds collections with effective full-text searching and metadata-based browsing facilities that are attractive and easy to use. Moreover, they are easily maintained and can be augmented and rebuilt entirely automatically. The system is extensible: software "plugins" accommodate different document and metadata types.

Greenstone incorporates an interface that makes it easy for people to create their own library collections. Collections may be built and served locally from the user's own web server, or remotely on a shared digital library host. End users can easily build new collections styled after existing ones from material on the Web or from their local files (or both), and collections can be updated and new ones brought on-line at any time.