



Citable References for Electronic Records in TNA¹

Seamless Flow Programme

Release

Date: 30th March 2007

Author:

Project Manager

Owner:

Senior Responsible Owner (SRO)

Client:

The National Archives (TNA)

**Document
Number:**

0.3

¹ The first edition of this report was presented to, and endorsed by, the Seamless Flow Programme Board in July 2005.

Document History

Document Location

This document can be accessed here: *[insert Objective link]*
Hard copy versions of this document are only valid on the day that they are printed

The Objective file path for this document is: *[quote Objective file path]*

Revision History

Date of next revision:

Revision date	Version	Summary of Changes	Changes marked
12/02/2007	0.1	Original draft	No
13/02/2007	0.2	Amended appendices	No
02/04/2007	0.3	Amended appendices	No

Role	Form of Approval	Version	Location
Programme Manager			<i>Insert Objective link here</i>
Programme Director			<i>Insert Objective link here</i>
SRO, Transfer to TNA			<i>Insert Objective link here</i>
			<i>Insert Objective link here</i>

Distribution This document is stored in Objective with the following see, open and edit privileges: *[insert Objective privileges]*

An Objective link to this document has been sent to the following individuals or groups.

Name	Title	Role	Date of Issue	Version
Seamless Flow Project Managers				
Programme Board				

Contents

Topic	See Page
Role of this document	5
Scope	5
General definitions	5
Requirements	6
Citable References : format structure and allocation	7
Redactions	8
Hybrid series	9
Persistence	10
Usage	10
Appendix 1:	11
Meeting on Hybrid Records 23 rd Aug 2006	
Meeting on Metadata Migration and Referencing 7 th Feb 2007	14
Appendix 2: Terminology	17
Appendix 3: Sources used	19

1. Role of this Document

1.1. This is a high-level document recommending developments relevant to the workflow of Seamless Flow as a whole.

1.2 This document makes no recommendations as to the implementation of its recommendations.

2. Scope

2.1 The scope of this document incorporates the automatic referencing of all electronic records during the processes of capture and transfer to TNA, storage and maintenance in TNA, redaction by Other Government Departments (OGD) (or by TNA at the request of OGD), and preservation and presentation by TNA.

3. General Definitions

3.1 Definition of a citable reference for an electronic record

3.1.1 A citable reference for an electronic record is reference used by members of the public either onsite or online, to identify a born-digital record, and by members of TNA staff to locate and identify that same record.

3.1.2 This definition is consistent with the current version of *CATALOGUING GUIDELINES. PART A, (DATA ELEMENTS), (2nd rev) 2005*: 'Reference: The unique identifier which links the Catalogue to the records it represents and allows the user to order and refer to them'.

3.1.3 The referencing system of which citable references are a part is alphanumeric and similar to the referencing system used in PROCAT.

3.1.4 The referencing system of which citable references are a part is separate and distinct from any referencing system used internally to identify accessions or accumulations, or to locate objects in the Digital Object Store (DOS) or to locate metadata in the Generic Metadata Management System (GMMS)².

3.1.5 The citable reference does not indicate the open or closed status of a record. This information is held elsewhere: publicly available in the access status field of metadata relevant to an electronic record; and for internal recording purposes in the System for Access Regulation (SAR).

3.2 Definition of an electronic record

3.2.1 For the purposes of this document and the processes to which it relates, the pragmatic working definition of an electronic record which has been developed elsewhere in Seamless Flow: 'an electronic record is an electronic

² Such systems may, of course, be the subject of additional documentation.

object, or a collection of objects, with its accompanying metadata', has been adopted here³. Electronic records in a Seamless Flow context may also be referred to as 'born-digital records' to distinguish them from their digitised counterparts which are digital copies of paper records.

4. Requirements of the citable reference

4.1 In identifying the nature of the citable reference for electronic records and the system of allocation of those **references**, the following requirements have been agreed:

No.	Description:
1	Any citable reference should be easily printable and quotable
2	Any citable reference should be relatively easy to recall
3	Any selected referencing system should be compatible with our current paper referencing system
4	Any selected referencing system should be easily extensible to cater for capacity requirements
5	All citable references should be persistent
6	Citable references should be restricted to a mixture of numbers and letters; the use of symbols should be kept to a minimum
7	Allocation of citable references to record sets should be amenable to automation
8	Allocation of citable references should be by TNA
9	Allocation of as much of the citable references as possible should occur at preparation for transfer
10	Departments need to know as much of the citable references possible at transfer or before to manage records that are closed, retained or redacted
11	Where possible/appropriate the citable references should be resolvable to a url
12	Any selected referencing system should allow for a common searching mechanism across both paper and electronic records
13	Where a citable reference is presented to the user, context information must also be available for supply
14	All appropriate metadata fields need to be captured for a given citable reference
15	It should be possible to provide the full range of metadata at any level of the referencing system
16	However it should not be mandatory to supply metadata at all levels of record definition
17	It is essential we maintain the current mechanism for identifying departments
18	We need to maintain the concept of series number but with respect to

³ See Seamless Flow Glossary, <http://www.nationalarchives.gov.uk/documents/glossary.pdf>

No.	Description:
	electronic records this may reflect functional area rather than record series as with paper, or a departmental file plan
19	We need to catalogue a record in same sense as for paper records
20	We need to ensure conformance to standards – ANSI/ISO
21	Users should be able to browse electronic records
22	Users should not need to understand the basis of the referencing system
23	Users should not need to differentiate between PROCAT and electronic records when searching
24	Users should only need to visit a single site to search and locate both paper and electronic records
25	The citable reference given to the User doesn't necessarily need to cite down to the lowest object
26	To insure investment for the future we should minimise any changes to PROCAT

5. Citable reference: format structure and allocation

5.1 Format

5.1.1 The full format of the **citable reference** will be composed of an Archival reference

eg. DCA 3
PREM 47

and an object reference

eg. DCA 3/46
PREM 47/16

5.1.2 In cases where the record or its metadata have been redacted in TNA, the citable reference will also include a numerical reference to the redaction see below, 6.4.

5.2 Structure

5.1 The **Archival reference** relates to the Department and Series

5.2. The **Object reference** which follows refers to the **Deliverable Unit** level (**DU**). A **Deliverable Unit (DU)** is, as its name suggests, the unit which is delivered to the user. A **DU** is defined as : 'logically deliverable' ... possessing 'a unique reference.'⁴

⁴ Seamless Flow Glossary <http://www.nationalarchives.gov.uk/documents/glossary.pdf>

5.3 Reference Allocation

5.3.1 The allocation of references shall take place in the Transfer process except in the case of Redactions; see 6.4.2 below

6. Redactions

6.1 A redaction can be defined as 'a record, part or parts of the text of which have been removed, withheld or hidden due to either the application of an FOI Exemption or a decision by TNA to restrict access where sensitivity, copyright or Data Protection issues arise'⁵.

6.2. A redaction may be created by one of two processes, either by an OGD, or by TNA on behalf of an OGD; TNA shall make provision for both.

6.3 Redaction in OGD

6.3.1 Where a redaction of a closed or partially closed record is transferred to TNA, the citable reference will be as at 5 above.

6.3.2 it may be however, that the citable reference of the redacted record will differ significantly from its unredacted relation: its object reference may not be the same, and as some time may have elapsed between the arrival of the closed and the redacted record, it may be at quite a distance, in numerical terms, from the **citable reference** of the unredacted record.

6.3.3 Where this is the case, provision should be made in the pre-accessions and cataloguing processes to ensure accurate metadata and appropriate cross referencing so that these relationships can be made clear to OGD, TNA and to all users.

6.3.4 As an alternative to cross referencing, the provision of a 'marker' or place holder for closed records, providing a place for the redaction(s) to go which is numerically near the closed record, might also be investigated.

6.4 Redaction in TNA

6.4.1 It has been agreed in the *Digital Records Redaction Referencing Document*⁶, that where a redaction is carried out by TNA on behalf of OGD, the redacted version of a digital record shall indicate redaction status in its citable reference. In the case of

DCA 3/46

then, the first redaction of this Deliverable Unit (DU) will be

⁵ See Seamless Flow Glossary, see <http://www.nationalarchives.gov.uk/documents/glossary.pdf>

⁶ *Digital Records Redaction Referencing Document*, June 2005; revised June 2006

DCA 3/46/1

In the case of

PREM 47/16

the first redaction of this Deliverable Unit (DU) will be

PREM 47/16/1

and if there are further redactions

DCA 3/46/2

DCA 3/46/3

PREM 47/16/2 ...

and so on.

6.4.2 The allocation of references shall take place in the Record Preparation System (RPS)⁷.

6.5. In the unlikely circumstance of a redaction of one record being created in OGD, and a further redaction of the same record being created in TNA, provision will be made through use of the Transfer system to allocate TNA redaction references to both versions.

7. Hybrid records and hybrid series

7.1 Definition

7.1.1 Hybrid records may be loosely defined as records, part of which are paper, and part of which are digital. In TNA terms, these form hybrid series, where some pieces and items are paper and some pieces are items are digital. Three types of hybrid records have been identified as being in current use in OGDs⁸:

7.1.2 **Type 1:** These arise from the high security classification of certain records which cannot be kept on an Electronic Records Management System (ERMS) and remain in paper. These records do however relate closely to electronic records of the same provenance but because of security considerations no link is maintained between the paper and digital records.

⁷ Technology Watch (TW) will have overall responsibility for the process: *Business Change Team Leaders Workshop*, 25 January 2007.

⁸ See *Notes of a Meeting on Hybrid Records* 23rd August 2007, in Appx 2 below

7.1.3 **Type 2:** These are paper records which are registered within an ERMS by a marker or place holder, ie. a folder with a metadata 'stub' which describes the paper file⁹. An example of these are case files.

7.1.4 **Type 3:** Unstructured, mixed media records which are predominantly paper. Examples of these records are hard-copy reports that contain information stored on a CD-ROM.

7.2 The continued maintenance of hybrid series in TNA would require the GMMS and PROCAT to allocate and present discontinuous references. As the GMMS is unable to do this, the solution for hybrid records is as follows:

7.2.1 Accession and reference the digital records and the paper records of a hybrid series as separate series and cross reference as necessary¹⁰.

7.2.2 In cases where the hybrid series is of **Type 2** and where the paper records are eventually transferred to TNA as digital copies and where those digital records slot into the place holders appropriately, then accession and reference as digital records. Where the paper records of the hybrid series are eventually transferred to TNA as paper, accession and reference as at 7.2.1 above.

8. Persistence

8.1 All **citable references are to be persistent** and to resolve to a url¹¹.

9. Usage

9.1 The usage of citable references as defined above shall pertain throughout Seamless Flow, shall be used in TNA's online resources where material catalogued from OGDs is presented, and shall be used by members of the public either online, or onsite at TNA, to refer to authoritative citable sources.

⁹ As described in PRO, *Requirements for Electronic Records Management Systems, 2002: Part 1. Functional Requirements, B.3.9-B.3.13; Part 2. Metadata Standard, Section 12*

¹⁰ see PRO, *Requirements for Electronic Records Management Systems 2. Metadata Standard, 2002 Section 9*

¹¹ For the use of persistent identifiers in the archival context, see *A practical approach to ensuring the persistence of digital collections at the National Library of Australia, May 2002*

Appendix 1

Title: **Meeting on Hybrid Records & Seamless Flow**
Date of Meeting: **23rd August 2006**
Location: Conference Room B
Subject or Purpose: Ad Hoc meeting/workshop to find solutions.
Attendees: Project Team members

Apologies:

Action

- 1 Current practice**

The current practice of processing/accessioning hybrid series was discussed.

The current practice for the processing/accessioning of electronic records into Electronic Records Online (ERO), some of which may be part of a hybrid series, is detailed in the linked paper.

- 2 Types of hybrid records**
 1. Confidential records kept outside of the Electronic Records Management Systems (ERMS) and with no link between – records which due to their high security classification could not be kept on the ERMS.
 2. Records such as case files which are registered within the ERMS by having a folder and perhaps stubs for the paper files. Metadata describing the paper records would appear in the ERMS. There is a close linkage between the ERMS and the paper files in question.
 3. Unstructured, mixed media records which are predominantly paper – records such as hard-copy reports that contain information stored on a CD-ROM.

3 **Issues arising from supporting hybrid records in Seamless Flow**

- Related paper records are being identified in the notes field of the electronic record
- Timing of accession may not be the same for paper as electronic
- Allocation of reference needs to be synchronised
- Release of records may not be simultaneous
- Marker locations exist in ERMS systems for paper records
- Marker locations in ERMS systems may relate to more than one paper record
- Marker locations could relate to media other than paper.
- Paper records will need to be entered into the catalogue with metadata and allocated a reference
- Cross referencing between paper and electronic records needs to be implemented

4 **Possible Solutions**

4.1 Type 1 hybrid records

These records are generally kept separately from each other and may be accessioned at different times. They are likely to be accessioned as separate series. Therefore it was agreed these should always be accessioned as separate series. This will enable resolution of all issues above.

4.2 Type 2 hybrid records

Since there is a strong formal link between the marker in the ERMS and the actual paper records, it was felt inappropriate to split these into two separate series. The information was such that it should be archivally registered as a single series. This meant that all the issues listed above would need to be addressed. Detailed proposals for handling this type of hybrid records are detailed below.

4.3 Type 3 hybrid records

Since these are predominantly paper records, they should be accessioned into the Catalogue in the usual way and the digital records entered into GMMS. This would of necessity have to be a handcrafted process and thus should enable issues to be overcome.

5 **Proposals for handling Type 2 hybrid records**

5.1 It is assumed that **all** associated paper records will have place holders within the ERMS.

5.2 The GMMS will process the full metadata including place holders for hybrid records.

5.3 It will allocate a reference to the place holder.

5.4 It will also allocate a reference to the paper object(s)

5.5 This reference will be placed in the related material location of the place holder to provide the cross reference to the paper object(s).

5.6 An export of the metadata for the place holder but with the reference interchanged with the cross reference will be produced.

5.7 This export will then form an import to the Catalogue to create the catalogue entry for the paper record(s).

5.8 This mechanism will ensure:

- a) integrity of reference allocation
- b) cross-reference between the electronic and catalogue
- c) maintenance of the integrity of the original ERMS structure
- d) provision of some automation of metadata for the Catalogue
- e) reduction of onus on OGD to prepare list for paper accession

6 Next steps

6.1 Some unproved areas exist with the proposed solution so it was agreed that further investigation was required.

6.2 LC agreed that someone (probably PS) would look into the feasibility of importing discontinuous references within a series into the Catalogue

6.3 IH agreed to look into the feasibility of exporting the relevant metadata from the GMMS for import into the Catalogue.

6.4 It was agreed that the above proposals be explained to a wider client manager forum in about a months time.

6.5 The timetable for action was discussed. It was not anticipated that hybrid records would be part of the pilot and that an implemented working process and software tools would not be required until next FY. It was decided to confirm viability of process; discuss with client managers and gain feedback before committing to identified solution.

Meeting on Referencing and Metadata 7 February 2007

Title: **Metadata Migration and Referencing Meeting**
Date of Meeting: **7th February 2007**
Location: **Con D**
Subject or Purpose: **Review Actions Required**
Attendees: **Project Team members**

1 Apologies

2 Welcome and Background

2.1 LC welcomed everyone and explained the background to the meeting – what do we need to do to ensure a successful go-live date for Transfer ?

3 Citable References

3.1 LC briefed team with main changes to Reference paper

- Format and structure
- Definitions into appendices

3.2 **Do we need to cite presentation and preservation manifestations ?**

3.2.1 We will not be keeping past presentation copies – metadata will keep audit of past versions – if we cite must be able to produce – but we will not be keeping past copies

3.2.2 Information content the same no need to cite a manifestation

3.2.3 Team agreed we do not need to cite manifestation copies

3.3 **Redactions**

3.3.1 If done by OGDs could send in closed and redacted records via Transfer system – would receive 2 sequential DU reference numbers – different to those done at TNA

3.3.2 IH to look at system / process to load via Transfer system as redacted

3.3.3 Team agreed it is likely that OGDs will not ‘do’ redactions

3.3.4 Team agreed preference to use redacted reference as if done at TNA – some might slip through – deal as ‘one off’s’ and monitor volumes / numbers

3.3.5 Team accepted Citable Reference Document for sign-off at next Wednesday review

4 What we need to do for migration and related issues

4.1 LC presented key points from discussion paper :

- Options on referencing items in Digital Archive (DA)
- One off process or something for the future

4.1.2 Some suggested options discussed :

Action

IH

- Straight conversion – see app 2 CAB 172
 - Concatenate into 1 piece – see app 2 JS3/1
 - Use Procat numbers – see app 2 JS5 and AN192
 - Hybrid – use high number to make visible / 10000 – see app 2 IR122 and JC8
- 4.1.3 Some issues discussed :
- Can we use number already used
 - Can we concatenate
 - High numbers for hybrid
 - Technical possibilities
 - OGD assurances that digital series are not hybrids
- 4.1.4 Team agreed :
- Close old numbers and cross reference
 - Could concatenate – but what about individual items
 - Can't use discontinuous numbers in GMMS
- 4.1.5 *GMMS would recognise existing Procat numbers as Redactions therefore cannot use existing Procat references***
- 4.1.6 Agreed not to concatenate as would not be able to show records as before
- 4.1.7 *Team agreed to create as new individual DU references and cross reference***
- 4.2 The National Archives Catalogue (PROCAT)
- 4.2.1 Series level and above metadata to be held in Procat for the foreseeable future
- 4.2.2 *The use of our referencing system would mean we could not enhance a digital record and keep the same reference number***
- 4.3 Surrogates
- 4.3.1 Issues still to be identified and agreed
- 4.4 Summary of agreements to date :
1. ***Team agreed we do not need to cite manifestation copies***
 2. ***Team accepted Citable Reference Document for sign-off at next Wednesday review***
 3. ***GMMS would recognise existing Procat numbers as Redactions therefore cannot use existing Procat references***
 4. ***Team agreed to create as new individual DU references and cross reference***
 5. ***Team agreed that hybrids would have to be created as a new series and cross referenced in Procat***
 6. ***The use of this referencing system would mean we could not enhance a digital record and use the item reference for that enhancement as is currently done for some paper records in TNA***

5 Agreed Actions

- 5.1 Team agreed the following actions :
- 1) Agree new series for DA items
 - 2) Need mapping table – old → new
 - 3) Need process for ERO to tell GMMS what presentation copies exist
 - 4) Metadata from DA into GMMS
 - 5) Assign new references in GMMS
 - 6) ERO database query for file names and put into GMMS and create presentation copies
 - 7) Communication plan :

- (1) Internal
- (2) External
- 8) Trial run of process - March
- 9) Target live date – early June

- 5.1.1 Team to consider any missing issues
- 5.1.2 Compile project plan
- 5.1.3 Confirm project leader

**ALL
DB / DH
DB**

In preparation for this meeting, the following Summary was drawn up by Programme Manager on 31 January 2007:

Summary of proposed referencing scheme for use within the Seamless Flow environment.

Archival component		Object	Redaction	Manifestation	Comment
Department	Series	Deliverable Unit	Redacted unit	Archival or presentation	
DCA	23	10001		A1	Archival manifestation
DCA	23	10001		P1	Presentation manifestation
DCA	23	10001	1	A1	Redacted archival manifestation
DCA	23	10001	1	P1	Redacted presentation manifestation
DCA	23	10002		A1	Archival manifestation
Etc.					

The current citable reference is:

DCA 23/10001 for the actual deliverable record

DCA 23/10001/1 for the redacted record

Need to confirm that the manifestation reference is NOT citable

Also need to confirm format for the manifestation part of the reference.

A further area for discussion is referencing of redacted versions if a) done at OGD or b) done at TNA. If latter, the DCA 23/10001/1 reference would be used but if done by OGD before accession then would have deliverable unit reference allocated ie DCA 23/10002.

Appendix 2

Terminology used throughout SF

1.1 It has been found useful, throughout SF, to adopt the terminology developed by the *Open Archival Information Systems (OAIS)*, an ISO standard gaining considerable international acceptance as a common language to discuss digital archiving¹².

1.2 Electronic records will be transferred from OGD to TNA in the form of **Information Packages** consisting of digital objects and preservation metadata packaged together, along with some descriptive metadata to enable the user to locate it.

1.3 An Information Package ready for transfer may be called a **Submission Information package (SIP)**. A **SIP** contains the digital objects making up the record and surrounding metadata describing them, packaged for transfer from the record producer to the archive.

1.4 A SIP ready for loading into the archive may be called an **Archive Information Package (AIP)**. An **AIP** contains the digital objects making up the record and surrounding metadata describing them in a form suitable for loading into the archive for permanent **preservation**.

1.5 A *manifestation* of the digital objects making up the record and the surrounding metadata describing them migrated from an AIP and containing information suitable for **presentation** to consumers, may be called a **Dissemination Information Package (DIP)**.

1.6 **Manifestation** is the name given to any AIP or DIP.

1.7 **Derivation**: a *derivation* is the name given to a manifestation in which the information content is derived from, and different to, an original manifestation, regardless of the data format. The most common example of a derivation is a *redaction*.

1.8 It is possible using the terminology above to visualise the various manifestations of a record in 3-D. The vertical axis here shows **presentation manifestations** (DIPs) over time. The horizontal axis represents **preservation manifestations** over time (AIPs). The final axis, going into the page, represents **derivations** (changed information content – possibly **redactions**) of a record over time. The original transfer, an SIP or Submission Information Package, contains the original digital objects making up the

¹² ISO 14721 at <http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf>

For usage in Seamless Flow, see *Digital Records Terminology in Seamless Flow*, March 2005, 3.2.1.1, 3.2.1.2 and 3.2.1.3; 3.2.2 and 3.2.3.

record and surrounding metadata describing them, packaged for transfer from the record producer to the archive.¹³

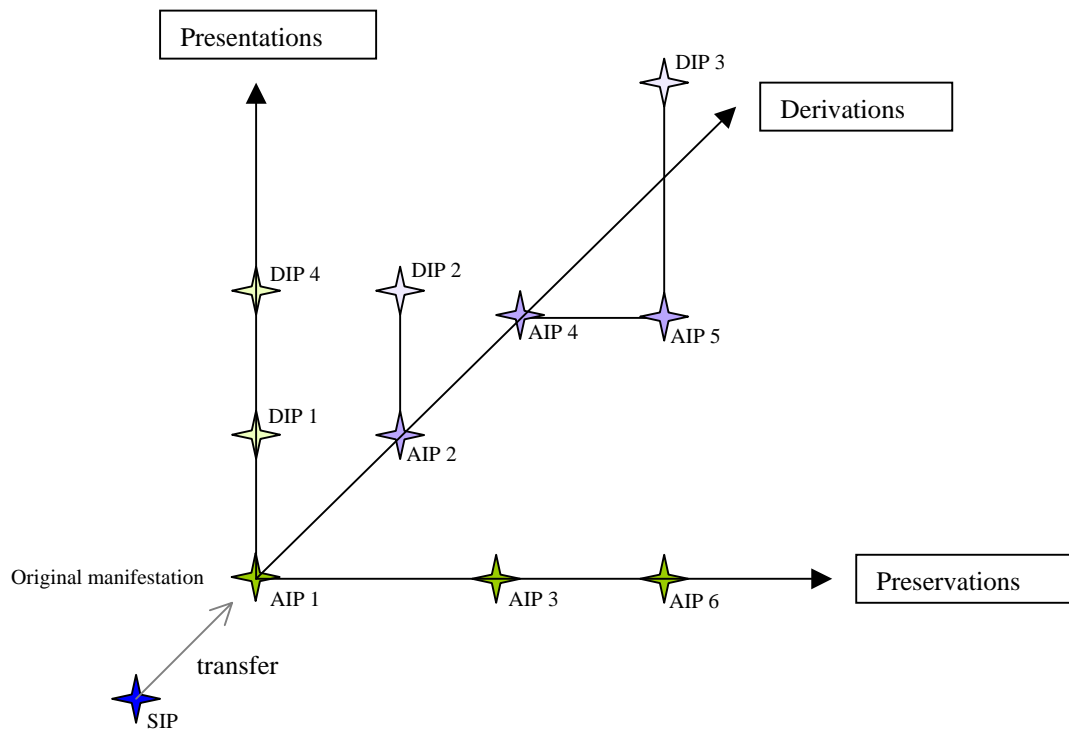


Figure 1 – Manifestations of a record.

1.9 The citable reference will not indicate the preservation or presentation manifestation status of any born-digital record; derivations in the form of redactions will however be indicated in the citable reference; see above, 6.4.

¹³ *Digital Records Terminology in Seamless Flow*, March 2005 3.3.

Appendix 3: Sources Used

Digital Object Identifier System (DOI), at <http://www.doi.org/>

National Library of Australia, *A practical approach to ensuring the persistence of digital collections at the National Library of Australia, May 2002* at: <http://www.nla.gov.au/nla/staffpaper/2002/boston2.html>

Open Archival Information System (OAIS), ISO 14721 at <http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf>

PRO, *Requirements for ERMS, 1: Functional Requirements; 2 Metadata Standard; 3 Reference Document, 2002; 4 Implementation Guidance, TNA. 2004*

Seamless Flow Documentation, including:

Referencing electronic records, July 2005

Appraisal Policy (Background paper), 2004

Digital Records Redaction Referencing Document, June 2005 revised June 2006

Statement of Requirements for Public Web Search, 2004

Digital Records Terminology in Seamless Flow, March 2005,

Seamless Flow Glossary, see <http://www.nationalarchives.gov.uk/documents/glossary.pdf>

TNA, *CATALOGUING GUIDELINES, PART A, (DATA ELEMENTS) (2nd rev)* 2005,

Unique Identifiers, A brief Introduction at <http://www.bic.org.uk/uniquid.html#DOI>