Catalogue Model Proposal

[TNA-CMP20/1]

Prepared for: Jone Garmendia, The National Archives

Prepared by: Adam Retter, Evolved Binary
adam@evolvedbinary.com

17 June 2020

Proposal version: 1.0.4

EVOLVED BINARY

ABSTRACT

The National Archives has initiated a new project: Omega, to act as a PoC (Proof-of-Concept) in evaluating technologies to replace its ageing catalogue management system for paper records: PROCat. This proposal looks not only at PROCat but also the larger organisation. Through analysis it is discovered that there are several catalogue like/dependent systems that are found to be in use as well as PROCat, which are then described herein.

This document first proposes widening the scope of Project Omega, to develop a Catalogue Model which is suitable not only for PROCat's paper records, but also for digitised, born-digital, web archival, and retained records. To achieve that end, this document examines the strengths and weaknesses of existing Catalogue Models and applicable standards, and makes recommendations for a new model.

This document also recognises that implementing a Pan-Archival Catalogue is out-of-scope for Project Omega itself, but that selecting a suitable model as the basis for the PoC replacement of PROCat is fundamental to achieving a future unified system within TNA.

Cover photo by Beatriz Pérez Moya on Unsplash.

TABLE OF CONTENTS

EVOLVED BINARY

EVOLVED BINARY

EVOLVED BINARY

EVOLVED BINARY

1. BACKGROUND

The original scope of Project Omega was to identify and prove a new data model and technology stack for a PROCat and ILDB replacement.

PROCat Editorial and Viewer are two web-based GUI applications, which allow staff to catalogue records held by TNA (The National Archives). The PROCat applications allow staff to perform CRUD (Create, Replace Update, and Delete) operations against a relational database called ILDB (Inventory Lists Data Base). ILDB is comprised of two distinct sets of data:

1. A physical implementation of the TNA-CS13 (The National Archives, Cataloguing Standards - June 2013) conceptual model in a relational database system.
2. Application state required by the PROCat applications to provide the business processes of operating the catalogue (predominantly workflow around record editorials and auditing).

TNA-CS13 which in the main is itself derived from ISAD(G) 2nd Edition (General International Standard Archival Description), was designed before archives were expected to accession and make available large collections of born-digital records. TNA-CS13 dictates a very rigid arrangement of records into a seven-level hierarchical structure, which for the most par has served TNA well for paper records, but more recently is often not flexible or extensible enough to describe digital records which may have more complex arrangement and locational histories.

As such, the ILDB TNA-CS13 implementation predominantly contains entries for Paper Records, which is the main concern of PROCat. Outside the domain of Paper Records, descriptions of Digitised, Born Digital, and Records held by other Archives, are held in other distinct systems at TNA, however those records may also have high-level catalogue descriptions in ILDB (typically only at the first three TNA-CS13 levels: *Department*, *Division*, and *Series*).

2. DOCUMENT SCOPE

This document concerns itself solely with a new Data Model for a TNA Catalogue. It does not consider the application UI or stateful business process data needed for replacing PROCat Editorial or Viewer, those will be dealt with separately. The Data Model will predominantly be considered at the Conceptual Level, i.e. its concepts and terminology are of more importance for this document than its technological implementation.

EVOLVED BINARY

3. IMPACT

PROCat and ILDB are absolutely core to TNA's business of both accessioning and presenting records. During our discovery phase of Project Omega, it has been recognised that as well as PROCat there are several other systems that interact with the TNA-CS13 data in the ILDB database. We briefly detail the dependency of these systems below.

3.1 System for Access Regulation

The SAR (System for Access Regulation) system contains additional data about the catalogue records in ILDB. SAR is the authoritative source for information on the *Closure Status* of records held by TNA. The Closure Status of a record determines if/when its description and/or manifestation are made available to the general public. SAR consists of a relational database and an Excel Spreadsheet. The database contains entries about both the Closure Status for records held by TNA, and information about the approvals of closure and retention applications to the Advisory Council. The Spreadsheet, which is stored securely, contains any Closed Descriptions of records which are considered sensitive. Closure Status data from the SAR database is replicated automatically every morning into the ILDB database. Closed Descriptions are manually copied from the Spreadsheet into the SAR database when they become *open* and are no longer considered sensitive, from there they then flow into ILDB via the afore mentioned database replication. The SAR database and ILDB share a flag named *AR* which indicates to PROCat that a record has *access restrictions* described in the SAR database. When a record becomes open, the Closure Status replicated from SAR to ILDB causes the *AR* flag to be cleared.

3.2 Document Ordering and Reader Information Service

The DORIS (Document Ordering and Reader Information Service) system holds information about the location and properties of the physical paper records at TNA, this is used by staff to retrieve records for the general public. In addition DORIS also shows whether a record is *orderable* by the general public. Only those records which are not restricted by *Closure Status* and are flagged as *orderable* can be requested (ordered) by the General Public. The PROCat ILDB database holds the canonical *Orderable* flag for records, this flag may be set at the TNA-CS13 Piece or Item levels. The orderable status of new or updated catalogue entries are replicated from ILDB to DORIS every 24 hours.

3.3 Transfers

The Transfer Team handle the transfer of records from external organisations to TNA. As part of the transfer process the Transfer Team receive either, a Transfer Form (also called AA2 or Blue Form) as a Microsoft Word Document, or a Transfer Template as a Microsoft Excel Spreadsheet. This is then used to prepare a new Microsoft Excel Spreadsheet which contains catalogue details of the incoming transfer of records. The final spreadsheet forms the starting basis of TNA's catalogue holdings. It is converted to a CSV (Comma Separated

Value) file, transformed via a tool called PET (Perl-based Extraction & Translation) into EADv1 XML, and then loaded into ILDB. A copy of the CSV file is also passed to The Accessions Team who manually enter it into their Accessions Database (Microsoft Access).

3.4 Historical Manuscripts Commission

TNA-CS13 makes use of Authority Files, each of which provide a single canonical name for an organisation, department, or person. Due to legacy technological restrictions, the HMC (Historical Manuscripts Commission) database is the authoritative source for Authority Files, this data is then replicated back into ILDB every 24 hours.

3.5 Discovery

The Discovery website allows users to survey and locate records held and/or described by TNA. New or updated catalogue entries in ILDB are replicated to Discovery every evening, which allows the records held by TNA to be searched, viewed, downloaded, or ordered online.

4. PROPOSAL OF A PAN-ARCHIVAL CATALOGUE MODEL

In addition to those systems that are directly dependent on the TNA-CS13 data in the ILDB database, during our discovery phase of Project Omega we have identified at least thirteen other distinct systems holding catalogue like information. These other systems, with prime examples being DocsOnline's ePRO and DRI Catalogue, were previously created specifically because the ISAD(G) based TNA-CS13 data model behind PROCat (as implemented in ILDB) was not extensible for digital records. Many of these additional catalogue systems are authoritative for a particular domain where ILDB is not.

With the ever forward march of digital technology replacing pen/print and paper, TNA is set to receive an increasing number of born-digital records and a decreasing number of paper records. Focusing solely on paper records when developing a new system would surely be short-sighted.

4.1 Perceived Business Advantages

A single pan-archival catalogue data model that allows records to be described consistently across the entirety of TNA could bring many benefits including:

1. by making the organisations assets visible from a single system, costs could be reduced through:

    1. reducing the time needed to locate and deliver assets
    2. avoid duplication of work and assets; previously several series have been digitised multiple times by mistake
    3. reducing inconsistencies across systems; previously when two systems described the same record in different ways, staff time is needed to investigate and resolve
2. by cataloguing both records and their digital surrogates, existing investments in digitisation could be reused to increase revenue
3. by improving the description of the organisations assets, communication both internally and externally could be improved; making records more easily searchable could lead to increased access and revenue opportunities for delivery
4. by reducing multiple catalogue systems to a single catalogue system, costs could be reduced through:
    1. purchasing and operating a single hardware and software system, as opposed to many
    2. staff would only need to be trained for a single system
    3. staff can assist and/or stand-in for each other, as they all understand how the same system operates

Thus, we propose that the scope of Project Omega be widened to develop a Catalogue Data Model which will be suitable for describing any type of record regardless of medium, whether it is deposited at TNA or held by another archive. Such a model must be able to operate on the abstract concept of a record, or expressed more concretely - a variety of different types of record. In addition, we cannot know what form or description will be required for records in the future, so such a model must also be easily extensible.

4.2 Additional Catalogue Systems at TNA

To both add weight to our proposal of a pan-archival catalogue model and for the sake of completeness, we describe the additional catalogue-like systems that we discovered.

4.2.1 DocsOnline

DocsOnline was the first system at TNA to hold and deliver digital surrogates of paper records via the Web. Due to limitations of TNA-CS13 with regards to digital records, it was found that PROCat and ILDB would not be suitable for storing those catalogue descriptions. Part of the DocsOnline system is the ePRO (Electronic Public Record Office) relational database which contains the catalogue of TNA's earlier digitised record projects. DocsOnline has been partially retired, it still serves the information that it holds, but any new transfers of digitised records are now accessioned via DRI. The long-term plan is to move the records currently held in DocsOnline

into DRI, and then to decommission DocsOnline and ePRO. DocsOnline has a substantial holding of records, and should not be underestimated, currently it holds approximately 9 million records.

4.2.2 DRI Catalogue

DRI (Digital Records Infrastructure) is TNA's Digital Archive, and as such holds digitised, born-digital, and some digital surrogate records. Whilst DRI was being built, it was recognised that there was a need for holding rich catalogue information and metadata about the digital records being accessioned. At that time, for likely similar concerns as those expressed in the DocsOnline project, it was recognised that PROCat and ILDB would not be flexible enough to hold the required catalogue data. DRI opted for a split catalogue approach, whereby PROCat would remain authoritative for TNA-CS13 Series level catalogue information, but DR'Is own custom-designed catalogue would be authoritative for records below Series level. The DRI Catalogue is backed by a triple based graph store, which was designed to be flexible enough to accommodate new metadata fields as and when required. The DRI Catalogue currently holds details of approximately 3.1 million records, composed of some 34 million triples.

4.2.3 Historical Manuscripts Commission

The HMC relational database holds catalogue information about records and papers that are predominantly in private hands, i.e. not transferred to TNA, this includes the MDR (Manorial Documents Register). It is worth noting that the MDR also contains records which are held by TNA. HMC only describes records at a very high-level but none-the-less it is a catalogue of records. As previously mentioned, instead of ILDB, the HMC database is also the canonical source for Authority Files used by PROCat, and so it also forms a small but important part of TNA's paper catalogue. HMC currently holds over 600 thousand records about private collections, and an additional 300 thousand records for the MDR.

4.2.4 Manage Your Collections project

The MYC (Manage Your Collections) project provides holds information about records held by other archives. This dataset originally came from the A2A (Access to Archives) catalogue which it replaced; A2A itself was a pure ISAD(G) model which was populated from external archives uploading EAD XML files. Each external archive uses the MYC website to describe a mapping between fields in a CSV file and ISAD(G), the archive then uploads CSV files which contain details about their catalogue, this is transformed into an ISAD(G) structure and stored in a Mongo database. MYC is a catalogue of catalogues. The entries uploaded by external archives are often exported from their own institution's Catalogue Management system and so MYC is usually non-authoritative, although there are some smaller archives that use MYC as their authoritative catalogue. It is also worth considering here that MYC's reliance on an ISAD(G) model, similarly to TNA-CS13 used by PROCat, likely restricts its ability to provide external archives with an effective catalogue system for digital records. MYC holds root entries for more than 2,500 external archives, 400 of which have uploaded catalogue data, with 150 of those archives still providing updates. MYC holds approximately 10 million records spread across the 400 external archive catalogues.

4.2.5 UK Government Web Archive

The UKGWA (UK Government Web Archive) captures snapshots of UK Government Websites, preserves them, and subsequently makes them available for access online. Two types of captures are performed, recurring scheduled captures of a specific Government Website or part(s) thereof, and event based targeted captures whereby one-or-more websites are captured as part of an event related collection. An example of a possible event based capture would be the "UK Climate Change Protests in London, October 2019".

For each initial snapshot to be captured, a *Collection* is created in the UKGWA relational database, this consists of metadata fields: Title, Description, and Note, and a configuration profile for the Web Crawler. Each collection in the UKGWA database may be associated with one-or-more TNA-CS13 CCRs at the Series Level; these are later entered into ILDB via PROCat Editorial. Currently there is approximately a two year backlog of UKGWA Collections, which have not yet had Series created for them in ILDB. It is important to note that a single Series is only created and assigned once for a particular Government website capture; in effect these Series refer to the abstract concept of a governmental website and are not anchored to any specific time. Multiple snapshots of the same Government website may be captured over time, and associated with that Series.

Each snapshot of a website, is produced as a WARC (Web ARChive) file by the Web Crawler. The WARC file contains both the content of the original website, and additional harvested and computed metadata to help describe and identify the original resources. The holdings of the UKGWA are substantial, it currently holds records for over 5,000 domains, consisting of some 6 billion resources. It should be noted that resources from recurrent snapshots of a website are not de-duplicated, and therefore there can be many copies of the same resources.

4.2.6 Arabian Gulf Digital Archive

The TNA joint project with National Archives of United Arab Emirates, digitised a number of records and collected descriptive metadata for those records as well as expanding the descriptive metadata for other records that were considered in scope. These records used TNA-CS13 and some may already be present in PROCat's ILDB, however not all of the records and/or extended descriptive metadata has been bought into ILDB. The records not in ILDB and the expanded descriptive metadata resides in a separate system, which is believed to be comprised of Excel Spreadsheets.

4.2.7 Supplementary Finding Aids

It has been reported that there are a number of paper based SFAs (Supplementary Finding Aids) which were partially digitised, but have not yet been transferred into PROCat and ILDB. These SFA contain additional catalogue type information about various records. These are believed to be works in progress, some of which

may have stalled for a prolonged period of time, likely due to limited resources and the varied challenges of translating them into meaningful and usable data, further work would be required to get these into a TNA-CS13 compliant form. It is believed that these are predominantly maintained as Excel Spreadsheets.

4.2.8 System for Access Regulation

SAR whilst discussed earlier, is repeated here as it contains the access information for each catalogue entry in PROCat's ILDB. Therefore, the catalogue data which is needed to understand a specific catalogue entry and act with it, is split between the two ILDB and SAR databases.

4.2.9 Record Copying

The Record Copying department, delivers a service to provide copies of paper records for the readers, not just those readers physically present in the reading rooms at TNA, but also those readers externally which are accessing records via the website. The readers accessing records via the website make up the vast majority of the orders placed with the Record Copying department.

A record is ordered via its PROCat Catalogue Reference. Each order creates a Job with a generated Job Number which identifies the readers request and the catalogue records involved. The Job information is stored and maintained in a SQL Server database as part of the RC2015 (Record Copying 2015) system. It is worth explicitly noting, that readers may also order a copy of just part of a record. Copies of records or parts thereof, whether ordered via the website or in-person, need to first be appraised by the Record Copying department, this is known as a "*page check*". The page check is needed because the information about the records available to the record copying department from the catalogue is insufficient. The page check, involves retrieving the physical original record, and verifying that the original is copyable and determining the extend of the copying required to fulfil the readers demands; This can be a somewhat complex process as the record (or record part) requested may be for example, a large non-regular sized piece of parchment, or even a box of thousands of letters. Once the page check is complete, options for copying and pricing are presented to the user.

When a record or part(s) thereof are copied, a "*Research Quality*" digital copy is made using a two dimensional document scanner with an acquisition quality of 300 DPI. The resulting digital files are stored in a lossy JPEG format on a networked file-system within the corporate IT infrastructure. They are organised into a hierarchy of folders, first by date, and then by Job Number. The files themselves are named after the catalogue reference of the paper record. Where there is more than one digital file for the same paper record, the catalogue reference in the each file's name is appended with a 3 digit sequential postfix. Finally, the reader according to their request is then supplied with either, a digital copy of the JPEG file, or a physical reproduction of the JPEG in an appropriate monochrome or colour printed format.

At present Record Copying rely on the catalogue references provided by ILDB via Discovery, from paper records they are effectively creating digital surrogates which are then unknown to either ILDB or Discovery. The existence of these digital surrogates is known only to the RC2015 system. These digital surrogates could be a useful asset to TNA if they were catalogued, for example, they could be used for web presentation purposes. In addition the RC2015 system is very limited, when receiving an order, it is very hard for the Record Copying team to determine if they have previously copied the requested paper records, improved cataloguing of digital surrogates could answer such questions.

The Record Copying service is available for over 15 million records, which potentially means that there is a large volume of uncatalogued digital surrogates. For the period of the 1st of February 2016 to the 9th of January 2020, the total number of record copying orders placed was 35,363.

4.2.10 Image Library

The Image Library is a web based application which provides users access to a library of digital images which were digitised from paper records held by TNA. These digitised images are in fact digital surrogates of paper records. The original paper records were catalogued in PROCat with their data stored in ILDB, however the digital surrogates are not.

The digitised image files making up the digital surrogates, are stored on a network file-system at TNA in TIFF format. The metadata about each digital surrogate file, which includes the catalogue reference of its original paper record, are stored alongside the digital file in an XMP XML *side-car* file. The digital file and its XMP file are related through a file naming convention.

The web based Image Library application is a branded version of the Asset Bank software from Bright Interactive; who also operate the online service on behalf of TNA. The digitised TIFF images, are converted to lower resolution JPEG images which are injected with the XMP metadata before being uploaded to Asset Bank. Upon upload, Asset Bank also assigns its own unique identifier to each digital image, which is used within the URLs of the Image Library website. There is a perceived risk in having these identifiers allocated and maintained externally to TNA.

The Image Library contains approximately 80,000 digital images. It is worth also noting that some digital surrogates are only part of a paper record, and not a representation of the entire record.

4.2.11 Digitiser

The Digitiser Tool (or Digitiser) is a relatively recent project, although some form of it has been operated by DRI since before 2011. TNA has in its possession many digital surrogates of paper records which have been created in the past, but never inventoried. Many of these surrogates were creating through licensing agreements with 3rd-parties such as Ancestry. Such agreements usually grant a period of exclusive access to Series of newly opened paper records by a 3rd-party in exchange for a fee and/or the digitisation of those paper records by the 3rd-party. After the period of exclusivity, these digital surrogates are then delivered to TNA by the 3rd-party in some agreed upon file format upon some agreed physical computer storage unit, typically - Tape, Optical Disk, or Hard Disk. Historically TNA has then stored these physical computer storage units in a similar manner to its paper records, most often within the Archival Stacks and/or placed in fire-proof safes.

The Digitiser project (and its predecessors) have been collecting these physical computer storage units from across the organisation, inventorying the units and their contents (digital surrogates), and uploading a copy of those digital surrogates into TNA's Dark Archive for redundancy. The inventory itself is a SQL Server database which contains details about the unit, file formats, file sizes, and identifiers (most often TNA-CS13 CCR) where known. These inventoried items may also have some licensing information added to them if it is available from TNA's Licensing Team.

Some interesting observations have been made by the Digitiser project:

1. A single digital surrogate file may represent many records. This is most-often the case where microfilm has been digitised. An example of this is the TNA-CS13 Series with the CCR: RG 10.
2. A single digital surrogate file may only represent part of a record. This is most often the case when a record is an entire book or folio or similar. An example of this is the TNA-CS13 Series with the CCR: PROB 11.
3. Many records have been digitised more than once, so there are duplicates or similes. The presumption is that this was done by mistake at expense due to a lack of inventory. An example of this is the TNA-CS13 Piece with the CCR: CAB 129/204, which was digitised at least three separate times.
4. Some of the digital surrogates that have been inventoried appear to be available via Discovery. It is not clear how to easily determine which were previously added to Discovery, or the reason for their inclusion (or indeed exclusion).
5. Some digital surrogates that have been supplied by 3rd-parties use the internal identifiers of the 3rd-party rather than TNA-CS13 CCRs; only in some cases is a mapping available. Examples of this include the TNA-CS13 Series with CCRs: RAIL 264, CUST 102, and BT 26.
6. Shockingly the majority of these digital surrogates are not registered in the organisations "Digital Pipeline", which means that not only are they not being monetised through reuse, but that they may even be re-digitised by mistake.

The most staggering thing about the Digitiser project, is its sheer scale and rate of growth. At the time of writing the Digitiser project holds an inventory of some 89 Million digital surrogates, making it the single largest catalogue of digital surrogates with in TNA by a large margin. It has grown by 1 Million digital surrogates within the last two weeks, and is predicted to reach 100 Million digital surrogates before the end of Q1 2020.

4.2.12 Super Integrated System

The physical activities of preservation and palliative care of TNA's physical records is undertaken and managed by the staff of the Collection Care department.

Collection Care record provenance information about the work that they perform upon TNA's records. This information is recorded into a system known within Collection Care as SID (Super Integrated System). SID is actually an external product called Collection Care which is provided by a company called Naqoda.

SID holds details of the work undertaken organised by TNA-CS13 CCR, whilst SID has only been in use for 2 years, it holds data going back 20 years. At present SID contains details of work undertaken upon approximately 285,000 records. Many records have had work undertaken on them more than once, so a conservative estimate is that SID holds the details of 500,000 remedial actions.

4.2.13 Multi-Spectral Images

More recently TNA's Collection Care department have gained the facility to perform multi-spectral imaging of physical records. This process can provide new insights into existing records.

Examples of TNA-CS13 records by CCR where multi-spectral imaging has offered new insight include:

- E 26: Exchequer: Treasury of the Receipt: Barons' Letter. The iron gall ink was faded and the ultraviolet reflectance image enhanced the text.
- FO 371/23231: General situation: Arab and Jewish claims. Code 31 file 6 (papers 2106 - 2230). A sentence in page 319 had been crossed out and could not be read. The infrared reflected image hid the fountain pen ink and revealed the text written in carbon-based typewriter's ink.
- HCA 32/1832 73 (Bundle A). Prize papers: letter written in Armenian, from a Captured French merchant ship Santa Catherina, leased to Armenian merchants. The bottom part of the letter could not be unfolded as the ink had adhered it to the page. The ultraviolet luminescence image and the phase-shift images (combination of visible transmitted and reflected images) enhanced the visibility of the text in the folded area.

The multi-spectral imaging process generates a series of digital image files (typically in TIFF or JPEG format), and a text file that describes the configuration of the multi-spectral imaging process. At present these are stored on a Network Filer organised into folders by TNA-CS13 catalogue reference.

Each multi-spectral image could be considered a new digital-surrogate manifestation of the record. As the process is relatively new, the volume of multi-spectral digital surrogates is assumed to be low at present.

5. CATALOGUE MODELS IN OPERATION AT TNA

There are several predominant catalogue data models in use at TNA. This section details each in turn and examines their strengths and weaknesses, with a focus on any constraints they impose which would make then unsuitable for describing a variety of record types.

5.1 TNA-CS13 Model

TNA-CS13 as used in PROCat and ILDB organises records into a hierarchy that is closely related to the physical arrangement of paper records within the archive stacks. TNA-CS13 uses a maximum of 7 levels, comprised of the 6 levels of ISAD(G) Second Edition and an additional level (as permitted by ISAD(G)) called Sub-sub-series which was introduced between Sub-series and File. TNA-CS13 does not entirely use the same naming for its levels as ISAD(G), and even before TNA-CS13 was standardised TNA in the past used various other naming conventions.

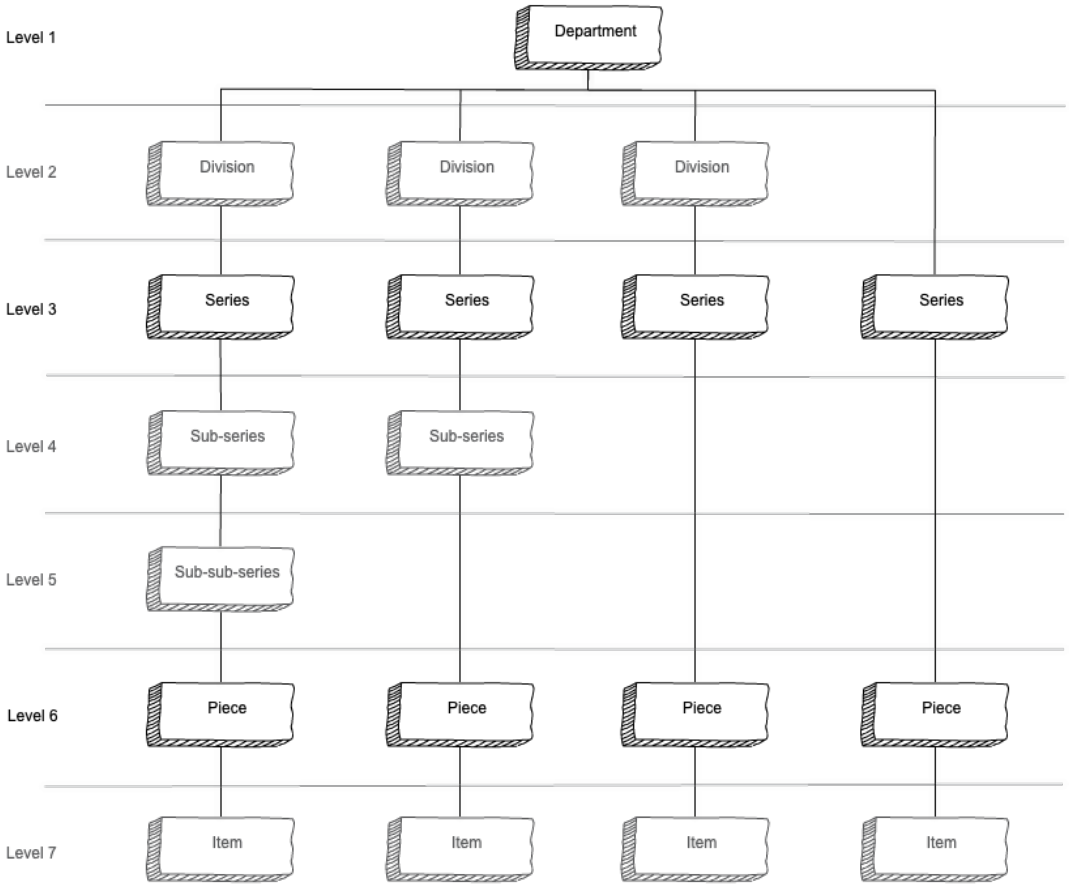
Catalogue Level Naming

Level#	Required	TNA-CS13 Name	Pre-2000 Name	ISAD(G) Name
1	Yes	Department	Letter Code	Fonds
2	No	Division	Division	Sub-fonds
3	Yes	Series	Class	Series
4	No	Sub-series	Header	Sub-series
5	No	Sub-sub-series	Sub-Header	*N/A*
6	Yes	Piece	Piece	File (*Level 5 in ISAD(G)*)
7	No	Item	Item	Item (*Level 6 in ISAD(G)*)

Records are catalogued under TNA-CS13 at either Level 6 - Piece, or Level 7 - Item. A Piece may (but not always) represent some form of physical container, e.g. a folder, folio, or box, containing one or more records.

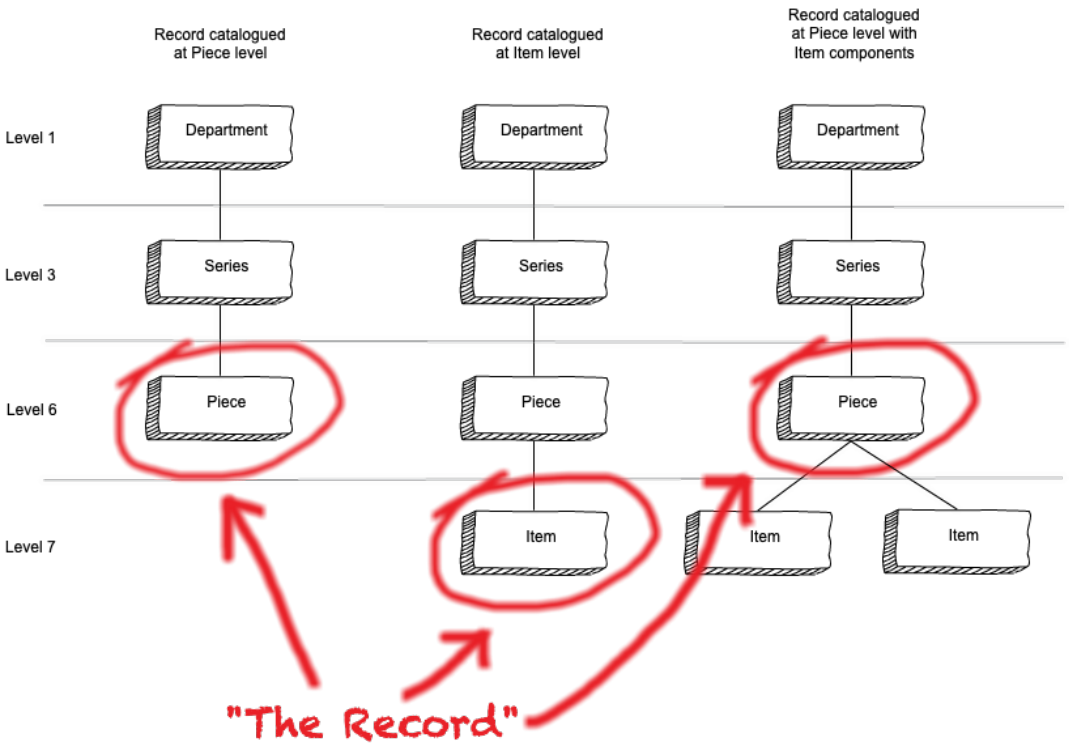
Whereas an Item most often represents an individual record, which, for example, itself may be composed of multiple paper pages. In the extreme case an Item may be a line on a page.

Fig 1 - TNA-CS13 Catalogue Levels



Often records are only catalogued down to the Piece level as cataloguing to the Item level may require an amount of effort which is not economically, politically, or socially viable. More generally the concept of The Record should not be conflated with the level at which it is catalogued; TNA has records catalogued with TNA-CS13 where *the record* is catalogued at Piece level, but the catalogue also contains Item level data representing components of the record.

Fig 2 - TNA-CS13 Catalogue Level Record Placement



The levels above Piece and Item, then serve to organise the records within the context of transfers from various creating bodies. Levels 2 - Division, 4 - Sub-series, and 5 - Sub-sub-series, are designated as Optional or Not-Required, these are generally only employed for large tranches of records where further dividing their arrangement makes sense either, physically for logistical purposes, or logically so as to better represent their original or archival arrangement.

5.1.1 Classic Catalogue References

TNA-CS13 also describes a catalogue reference scheme, whereby each Piece or Item has a unique reference. Herein, we refer to references within this scheme as CCR (Classic Catalogue Reference). CCRs literally encode the hierarchical arrangement of the records under TNA-CS13, however only the required levels are exposed in the reference.

For example, we may have a record which is arranged in the following manner:

- **Department**: *ABC*
- **Division**: *12*
- **Series**: *3*

- **Sub-series**: *4*
- **Sub-sub-series**: *5*
- **Piece**: *6*
- **Item**: *1*

From that catalogue arrangement only the Department, Series, Piece, and Item will be exposed in the CCR. CCRs are often written using the syntax:

```
<Department> <Series>/<Piece>/<Item>
```

For example:

```
ABC 3/6/1
```

It is worth noting that the `/` character which is often used as a visual separator when communicating a CCR can also appear within the Piece and/or Item part of the reference, this can make it difficult (or even impossible) to determine visually from a CCR whether a record is a Piece or Item.

5.1.2 TNA-CS13 Record Metdata

TNA-CS13 defines 40 *Data Elements* (or properties) which hold metadata about each level. Some of these properties are mandatory at various levels, or under various conditions.

TNA-CS13 Data Elements

Data Element	Used at Level(s)	Mandatory	Type
Summary Data Elements			
Legal Status	1 to 7	Highest/Orderable levels	Enumeration
Language	1 to 7	Highest level	Enumeration
Reference	1 to 7	Yes	Text
Former Reference (Department)	3 to 7	When applicable	Text
Former Reference (PRO)	3 to 7	When applicable	Text
Title	1 to 7	Levels 1 to 5	EADv1 XML
Map Designation	6 and 7	When applicable	EADv1 XML
Creator Name(s)	1 to 7	Highest level	Foreign key(s) to Authority
Covering Dates	1 to 7	Yes	Text
Physical Description, Extent	1 to 7	Levels 1 to 3	Text

Data Element	Used at Level(s)	Mandatory	Type
Physical Description, Form	1 to 7	Levels 1 to 3	Enumeration
Dimensions	6 and 7	No	Text
Map Scale Number	3 to 7	When applicable	Text
Physical Condition	3 to 7	No	Text
Place of Deposit	1 to 7	Levels 1, 3, 6, and 7	Foreign key(s) to Authority
Note	1 to 7	No	Text
Access Data Elements			
Access Conditions	1 to 5	Levels 1 and 3	Enumeration
Closure Type	6 and 7	Yes	Enumeration
Closure Code	6 and 7	Yes	Text
Record Opening Date	6 and 7	Yes, if level records closed	Date
Closure Status	6 and 7	Yes	Enumeration
Restrictions on Use	1 to 7	Highest level	Text
Immediate Source of Acquisition	1 to 6	Highest level	Foreign key(s) to Authority / Text
Custodial History	1 to 6	When applicable	Text
Accumulation Dates	3 to 5	When applicable	Text
Appraisal/Destruction Information	3 to 5	When applicable	Text
Accruals	3	When applicable	Text
Location of Originals	1 to 7	When applicable	Foreign key(s) to Authority / Text
Copies Information	1 to 7	No	Foreign key(s) to Authority / Text
Content Data Elements			
Scope and Content	1 to 7	Levels 1 to 3, 6, and 7	EADv1 XML
Arrangement	1 to 6	Yes, if diff. level arrangement	EADv1 XML
Publication Note	1 to 7	No	Text
Unpublished Finding Aids	1 to 7	No	Text
Related Material	1 to 7	No	Foreign Key to Reference
Separated Material	1 to 7	No	Foreign Key to Reference
Admin History Data Elements			

Data Element	Used at Level(s)	Mandatory	Type
Administrative/Biographical Background	1 to 7	Level 1 and 2, when applicable	EADv1 XML
Index Terms Data Elements			
Index Terms: Corporate Names	1 to 7	No	Foreign key(s) to Authority
Index Terms: Personal Names	1 to 7	No	Foreign key(s) to Authority
Index Terms: Places	1 to 7	No	Foreign key(s) to Authority
Index Terms: Subjects (*DEPRECATED*)	1 to 7	No	Foreign key(s) to Authority

- Highest level - Mandatory at the highest level to which it applies.
- Highest/Orderable levels - Mandatory at the highest level to which it applies, and for every Orderable unit (whether Piece or Item).
- When applicable - if the information is known and/or accurate for that level.

5.1.3 Strengths of TNA-CS13 Model

1. Through virtue of being based on the ISAD(G) standard, it provides a consistent international approach.
2. Closely represents the physical arrangement of paper records. This provides consistent language when communicating, expectations of service, and experiences across various departments, both within TNA for staff and without for the wider general public.
3. CCRs are concise, easily memorable and communicable by humans, and also infer context about the record.

5.1.4 Weaknesses of TNA-CS13 Model

1. Closely represents the physical arrangement of paper records. Digital records, and in-particular born-digital records are rarely organised in the same manner.
2. Weak data typing which could lead to inconsistencies and non-machine readable data; Most data elements allow just text, with editorial standards proposed around the values of those elements.
3. Apart from those few data elements that allow EADv1 XML, or the text-and-link data element, none of the other data elements allow structured content or hyper-links.
4. None of the data elements are extensible. Those that allow EADv1 are limited to that specific grammar.
5. Some data elements are specific to the record's medium. e.g. Map Designation, and Map Scale Number, which can be used for Maps. However, there are no additional data elements specialised for

the many other mediums of physical (paper) records, including - Books, Microfilm, VHS/Audio Cassettes, etc.

6. A CCR does not confer the complete arrangement of the record (it omits levels 2, 4, and 5). This is also a strength, as it keeps the CCRs concise.
7. CCRs have in the past used a `/' character within the names of some of their levels, when displayed this can be confused with the `/' character which is used to separate the Series, Piece, and Item.
8. There has historically been some confusion between the use of Title and/or Scope and Content data elements. At the highest levels of description the Title data element is mandatory (levels 1-5). At the Piece and Item levels (6-7), Scope and Content is the mandatory data element.
9. TNA-CS13 has no mechanism for describing record redactions, or surrogates, both of which could be thought of as manifestations of the record. At present redaction of a record is managed through modifying an existing catalogue entry, and adding a new entry for each redaction of that record.
10. There is almost no facility for describing and maintaining the provenance of records.

5.2 DRI Catalogue Model

The DRI Catalogue Model was originally developed because it was felt that the seven levels imposed by TNA-CS13 were not enough to describe the arrangement of digital records, and at the same time that the data elements of TNA-CS13 were lacking when it came to storing the rich technical and descriptive metadata available for digital records.

The DRI Catalogue Model was designed for the needs of both digitised and born-digital records. Its design was influenced by both the TNA-CS13 and XIP (XML for Information Package) version 4 standards. It also reused aspects of several other standards including Central Government Ontology, Dublin Core, FOAF, W3C's The Organization Ontology, OPM (Open Provenance Model), and W3C's PROV. Further information about the design of the DRI Catalogue Model can be found in the paper: "The National Archives Digital Records Infrastructure Catalogue; First Steps to Creating a Semantic Digital Archive. 2013. Rob Walpole".

The DRI Catalogue Model is constructed around two main activities:

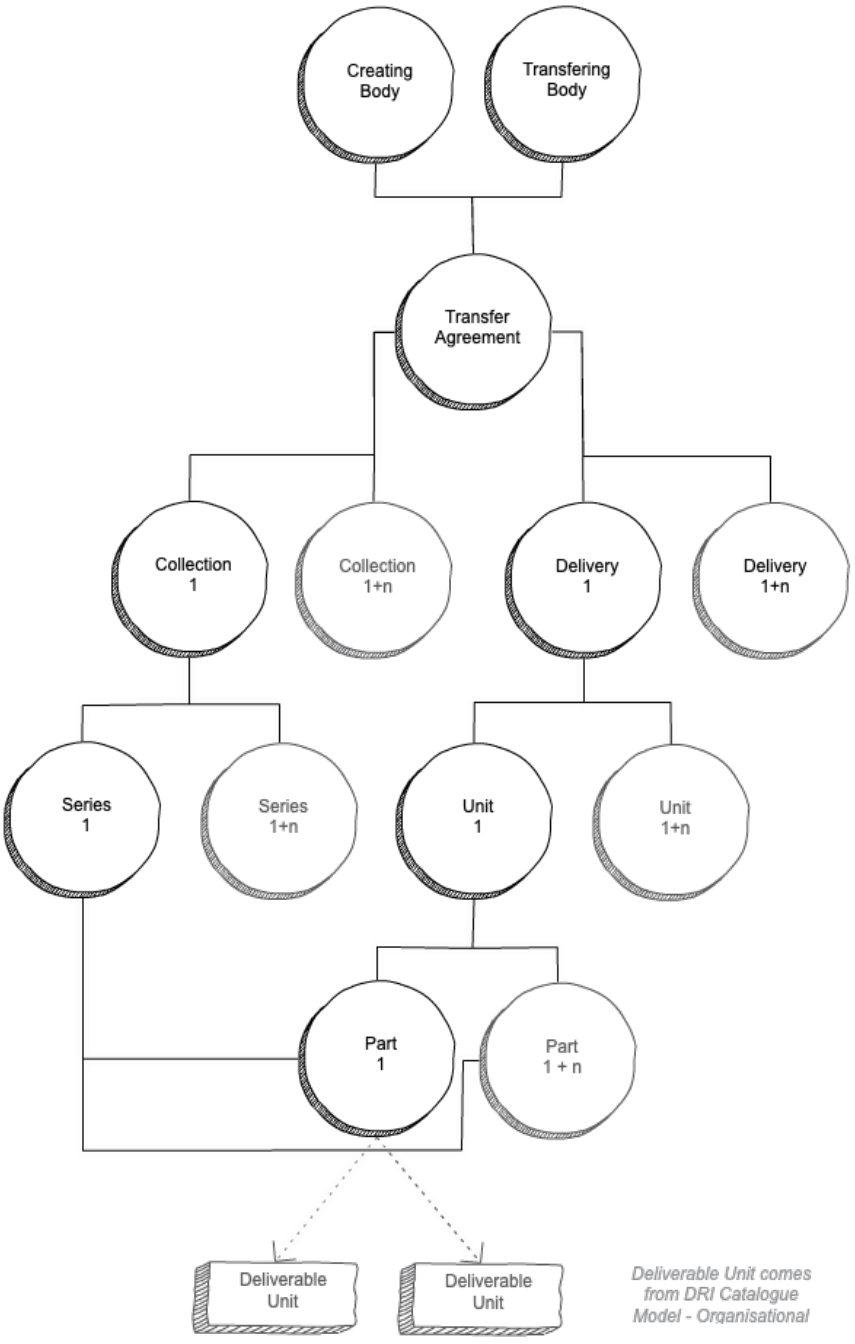
1. The transfer of digital records to TNA
2. The arrangement of those digital records

Where the DRI Catalogue Model expresses the transfer of digital records this goes beyond the scope of TNA-CS13, and allows DRI to understand and describe records within the larger process of accessioning. The

specifics of the transfer

Fig 3 - DRI Catalogue - Transfer

expression are not of

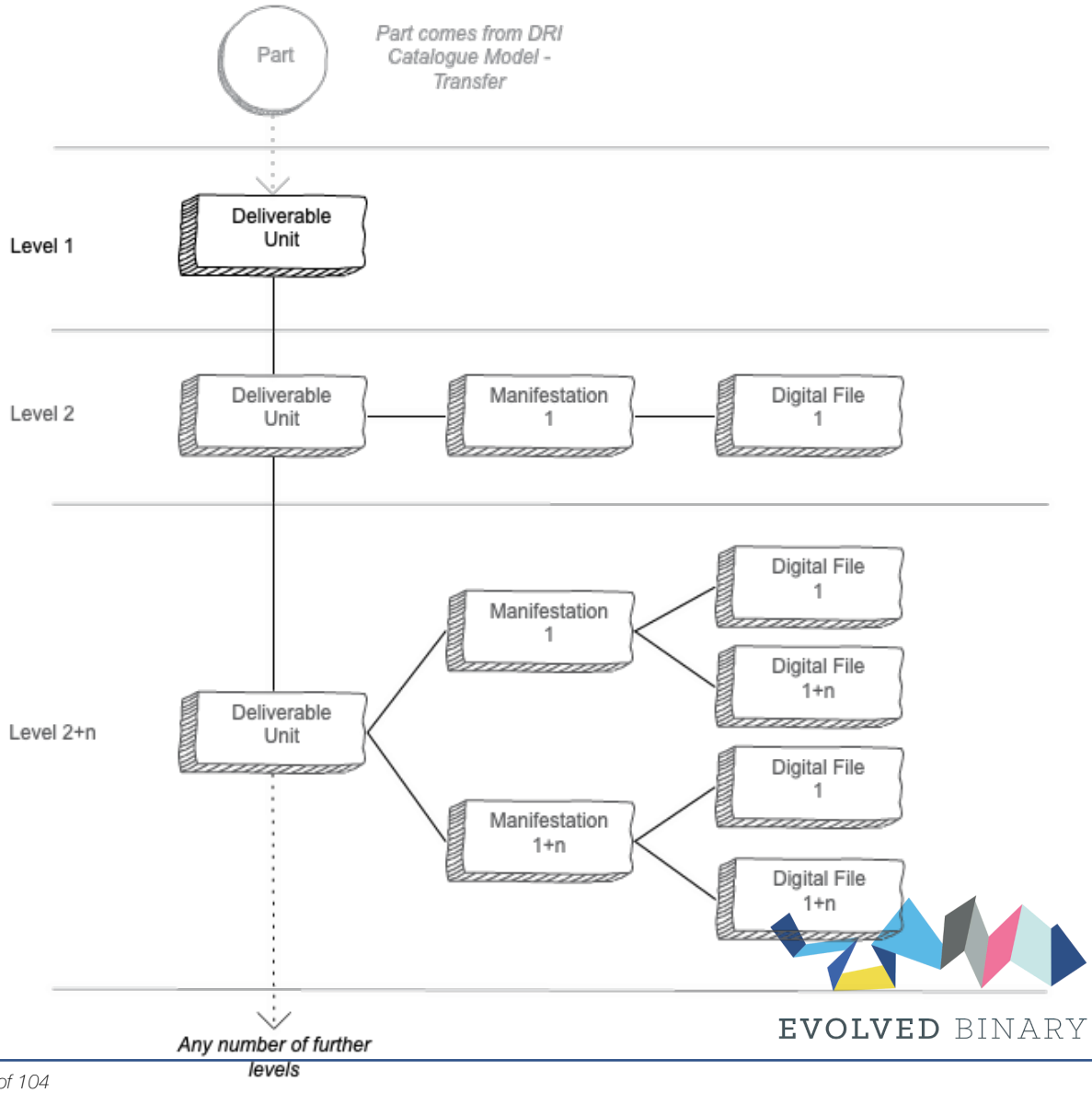


particular interest at this point apart from to point out that they allow records to be associated with a TNA-CS13 Series.

The concept of a Collection in DRI is not the same as a Series in TNA-CS13. DRI references TNA-CS13 Series, but also has *Collection* as a construct for grouping together related series that follow a common theme, e.g. the same provenance and/or creator.

The DRI Catalogue Model organises records into a hierarchy of unlimited levels. Each level is simply called a DU (Deliverable Unit). This unlimited hierarchy allows the DRI Catalogue to reflect deep arrangements of digital files. Each DU may have zero-or-more *Manifestation*. A Manifestation describes a representation of the digital record, for example there may be a Preservation Manifestation of the digital record for archival purposes, Presentation Manifestations for display/access purposes, and Format Migration Manifestations for preservation purposes. In turn these Manifestations may have zero-or-more Digital File, these Digital File describe the technical and digital location metadata for retrievable files. Digital File may also be shared between manifestations.

Fig 4 - DRI Catalogue - Record Levels



Each topmost DU for a record may be associated with a TNA-CS13 Series which is maintained separately in PROCat's ILDB. PROCat is authoritative for Series level descriptions. When needed, further compatibility with the first seven levels of TNA-CS13 can be maintained by assigning the topmost DU a TNA-CS13 level type of Piece and associated CCR, and if present, its child DU a TNA-CS13 level type of Item and associated CCR. The DRI Catalogue Model deviates from ISAD(G) Second Edition in three major areas:

1. By allowing any number of levels below the ISAD(G) level of Item.

   ISAD(G) technically allows for the division of intermediate levels, but prohibits adding levels below Item. If some degree of compatibility with TNA-CS13 could be ignored, then it would be possible to create a mapping from the DRI DUs to ISAD(G) levels. This could be done for each record by mapping the topmost DU to the ISAD(G) File level, and the bottommost DU to ISAD(G) Item level, any intermediate DUs would then be dynamically mapped to unnamed sub-divisions of the ISAD(G) File level.

2. Through the modelling of Manifestations

   ISAD(G) has no abstraction for describing different manifestations of a public record. For digital records, it is often desirable to hold multiple digital formats of the content of the record and by association varying technical metadata. The closest equivalent in ISAD(G) is the *Related Material* data element, but it would link to another record itself instead of a manifestation of the same record.

3. Through the modelling of Digital Files

   ISAD(G) has no data elements for describing the physical location and mechanisms of retrieval for a record. Rather it implies that the record's reference, may be used to locate the physical record in another system. For digital records, it is both obvious and desirable to link the description of the record with the content of the record.

5.2.1 Generated Catalogue References

The TNA-CS13 CCR scheme reflects the hierarchical arrangement of records in a catalogue. If it were to be extended to reference the records in the DRI Catalogue, as that model allows *n-levels* for each record, we would end up with potentially very long catalogue references!

Therefore, whereas a full length CCR for a TNA-CS13 record might look like: `AB 123/20/30`, there would be no bound on the length of a CCR describing a record in the DRI Catalogue Model, a theoretical CCR for a record at level 8 in DRI might look like: `AB 123/20/30/12/43/29/1001/1/26487`. It was quickly recognised that such long catalogue references were not very humane; they would be difficult to commit to short-term memory, or communicate in verbal and written forms. An additional concern for DRI, was that the CCR scheme has no mechanism for explicitly referencing different manifestations of a record.

As an extension of the CCR scheme would not be suitable for human use, DRI defined a new catalogue reference scheme designed for, but not limited to, digital records: the GCR (Generated Catalogue Reference) scheme.

A GCR just like a CCR consists of 4 components. The first two components of a GCR are the same as those in a CCR: Department, and Series; these were kept as even in the digital world, records are transferred from Departments and are likely arranged by those departments into Series, also as human concepts these allow staff and the general public to infer some context for a record.

Unlike CCRs, GCRs then do not further describe the arrangement of a record, in the digital world the arrangement of a record may be highly complex, deeply nested, or there may be multiple arrangements. With a GCR, the reference of the record, and the arrangement(s) of that record (below Series level), are completely isolated.

The third component of a GCR is the Record Number, this is simply a sequential number maintained for each Series that is incremented for each new record. For presentation purposes, the record number is encoded into a restricted Base25 alphabet. The restricted alphabet removes both, vowels to prevent sensitive (i.e. foul) words from being generated, and visually confusing characters such as `0` (Digit Zero) and `O` (Latin Capital Letter O). The Base25 alphabet also compresses the number, making it easily communicable by humans in verbal and written forms. The fourth and final component of a GCR is a Revision Number which indicates the manifestation of the record, the first (or only manifestation) always has the value `Z`, and the next `1+n` manifestations have the value `Z`*n* (e.g. `Z1`, `Z2`, etc); The `Z` identifier was chosen as it never appears at the corresponding Item level in CCRs, thus allowing CCRs to be visually distinct from GCRs.

GCRs are often written using the syntax:

```
<Department> <Series>/<base25(Record Number)>/Z[Revision Number]
```

for example, the first manifestation of a record:

```
ABC 12/7BF/Z
```

or, the second manifestation of the same record:

```
ABC 12/7BF/Z1
```

5.2.2 DRI Catalogue Record Metadata

DRI has a series of Metadata Guidelines and associated schemas which define the record metadata which is stored at each level of the DRI Catalogue. In DRI the metadata for a record is split between two systems, Tessella's SDB (Safety Deposit Box), recently renamed Tessella Preservica, and the DRI Catalogue. Although a goal of the DRI project in the past was to move all metadata into the DRI Catalogue, this has not yet been

achieved. The DRI Catalogue currently contains, and is the authoritative source of, metadata for the transfer and *n*-level arrangement of the record (DU, Manifestation, and Digital File), its GCR, and Access/Closure status. SDB currently contains, and is the authoritative source of, descriptive metadata for details of the record (e.g. Title, Covering Dates, Scope and Content, etc.), technical metadata about the record (both provided and computed), provenance metadata (both preservation and migration). Many of the descriptive metadata terms have a correspondence with TNA-CS13 Data Elements, however their expressions are likely more highly structured in DRI.

Below we describe the general metadata held for describing and organising digital records. DRI also has technical metadata, however the format of this is currently tightly coupled to the version of their Preservation System - SDB, and so we won't examine that further at this time. In addition, each Accession by DRI has custom metadata for the purposes of better describing those Series, that also will not be covered here.

DRI Record Metadata

Metadata	Used for	Mandatory	Type	TNA-CS13 equiv.
General Metadata				
dct:accessRights	DU, highest level	Yes	tnaxmdc:closure	Closure Type Closure Code Record Opening Date Closure Status
dct:coverage	DU, highest level	Yes	tnaxmdc:coverage	Covering Dates
dct:creator	DU, highest level	No		Creator Name(s)
dct:dateCopyrighted	DU, highest level	No	dct:W3CDTF	
dct:publisher	DU, highest level	Yes	Literal	
dct:relation (*DEPREDCTAED*)	DU, level 1	No	tnaxmdc:parentIA	
dct:rights	DU, highest level	Yes	Literal	
dct:rightsHolder	DU, highest level	Yes	Literal	
dct:spatial	DU, highest level	No	tnaxmdc:spatial	Dimensions Map Scale Number
Reference Metadata				

Metadata	Used for	Mandatory	Type	TNA-CS13 equiv.
dct:identifier	All	Yes	tnacdc:generatedCatalogueIdentifier tnaxmddc:digitalFileIdentifier tnacdc:departmentIdentifier tnacdc:divisionIdentifier tnacdc:seriesIdentifier tnacdc:subSeriesIdentifier tnacdc:subSubSeriesIdentifier tnacdc:pieceIdentifier tnacdc:itemIdentifier	Reference
dct:isPartOf	DU, level 1	Yes	tnacdc:batchIdentifier tnacdc:collectionIdentifier	
Digital Record Metadata				
dct:title	DU and DF	Yes	Literal or tnaxmddc:digitalFileTitle	Title
dct:alternative	DF	Yes	dct:URI	
dct:provenance	MA and DF	Yes	tnaxmddc:digitalFileProvenance or opm:opmGrap	
dct:subject	DU, any level	No	Various, e.g. hgdc:person	Index Terms: Corporate Names Index Terms: Personal Names Index Terms: Places Index Terms: Subjects
dct:references	DU, any level	No	Various, e.g. hgdc:battalion	Related Material Former Reference (Department) Former Reference (PRO)

- DU - Deliverable Unit
- MA - Manifestation
- DF = Digital File
- dct: - Dublic Core Terms namespace
- opm: - Open Provenance Model namespace
- tna*: - TNA specific sub-class namespaces
- Highest level - Mandatory at the highest level to which it applies.

5.2.3 Strengths of DRI Catalogue Model

1. The DRI Catalogue Model itself is implemented as a Graph in a triple store, and by following the Open World Assumption it is easily extended when there are new metadata requirements for records.

2. The DRI Catalogue Model uses a core model of Dublin Core and then extends it as necessary. It also incorporates other international standards. This makes it relatively easy to interoperate and describe archival records in a larger context than just TNA.
3. GCRs work well for digital records, they separate the larger arrangement of the record from its reference, and are concise, easily memorable and communicable by humans.
4. Unlike TNA-CS13, the DRI Catalogue Model does not restrict the number of levels of arrangement.
5. Unlike TNA-CS13, the DRI Catalogue Model has the concept of a *Manifestation of a Deliverable Unit (record)*, which allows for the realisation of multiple "versions" or "formats" of a record. This can support digital surrogates, redactions, and digital migrations.
6. Unlike TNA-CS13, the DRI Catalogue Model has the concept fo a Digital File. This allows it to describe technical metadata about the file, and directly link to the content of the file. Thus also making it suitable for digitised, and born-digital records.

5.2.4 Weaknesses of DRI Catalogue Model

1. The DRI Catalogue Model was designed relatively early in TNAs efforts to work with born-digital records and subsequently evolved over time, and as such makes some constraints and assumptions which would be better revised today.
2. Technical description (Metadata) of Public Records in DRI is split between SDB and the DRI Catalogue.
3. The DRI Catalogue Model as it stands only permits a single arrangement of a record within the catalogue, i.e. a mono-hierarchy. Digital Records originating from some systems can have more than one arrangement, i.e. a poly-hierarchy, this is not yet supported in the DRI Catalogue Model.
4. In DRI, for born-digital accessions, file-system folders are currently modelled as Deliverable Units which restricts multiple-arrangements of a Digital File, and restricts the possibility to de-duplicate identical files.

5.3 Business Information Architecture Model

The BIA (Business Information Architecture) model was developed for use in the Discovery project. The BIA model was originally developed to allow the Discovery website to facilitate access to both paper, and digital surrogate records. The BIA as initially developed was designed to hold a copy of the public presentation information about records from both the ILDB and ePRO databases, as such it was able to accommodate the classical hierarchical ordering of TNA-CS13, and a flexible set of metadata suitable for both TNA-CS13 paper records and digital-surrogates from DocsOnline. The BIA also held additional metadata for records for the purposes of visual placement and search indexing (e.g. names and places keywords) for the website.

The BIA centres around the concept of an IA (Information Asset). Each IA is an entity of interest that can be assigned an identifier and described by metadata. These IAs can be arranged into a hierarchy, for example, an IA could be a container of public records (e.g. a box, or Piece in TNA-CS13 terms) of records, or a record itself. Each IA may then have one or more Replica. Each replica is a manifestation of that IA, to which there may be attached multiple Digital File. The Replica serves the purpose of allowing Digital File(s) to be shared between manifestations of the IA as necessary.

For the arrangement of records, the BIA and DRI models are on the surface conceptually similar, both theoretically allow for *n*-levels of hierarchical arrangement, both model the concept of a descriptive entity (an IA in BIA, and a DU in DRI), both model the manifestation of a descriptive entity (a Replica in BIA, and a Manifestation in DRI), and both separate the Digital Files from the manifestations.

However, in practice the BIA model as implemented in Discovery is much less flexible than the DRI Catalogue model. Whilst the BIA model itself allows for the arrangement of records into a hierarchy of unlimited levels, the implementation of the BIA in program code in Discovery has restricted it to 7 levels which strictly mirror the TNA-CS13 hierarchy. This restriction, made it impossible for Discovery to accurately present born-digital records from DRI; records in DRI's *n*-level model lost additional context when they were sent to Discovery for display purposes, as Discovery had to both compress their arrangement into a 7 level model and discard extended metadata.

Internally the BIA model uses generated UUIDs (Universally Unique Identifiers) as identifiers for each IA, Replica, and Digital File, however these are never presented to user. For presentation purposes, each IA also holds either, a CCR which is taken from ILDB, or a GCR which is taken from the DRI Catalogue, as the authoritative catalogues.

5.3.1 BIA Metadata Fields

The BIA model defines a number of fields which hold metadata about each level, because of the reuse of TNA-CS13 levels, many of these have TNA-CS13 Data Element equivalents, others have been added to accommodate digital surrogates, or for presentation purposes. Below we describe the metadata fields used in the BIA for describing records.

Field	Description	Used for	Mandatory	Type	TNA-CS13 equiv.
AccDates	Accumulation Dates	IA, levels 3 to 5	When applicable	Text	Accumulation

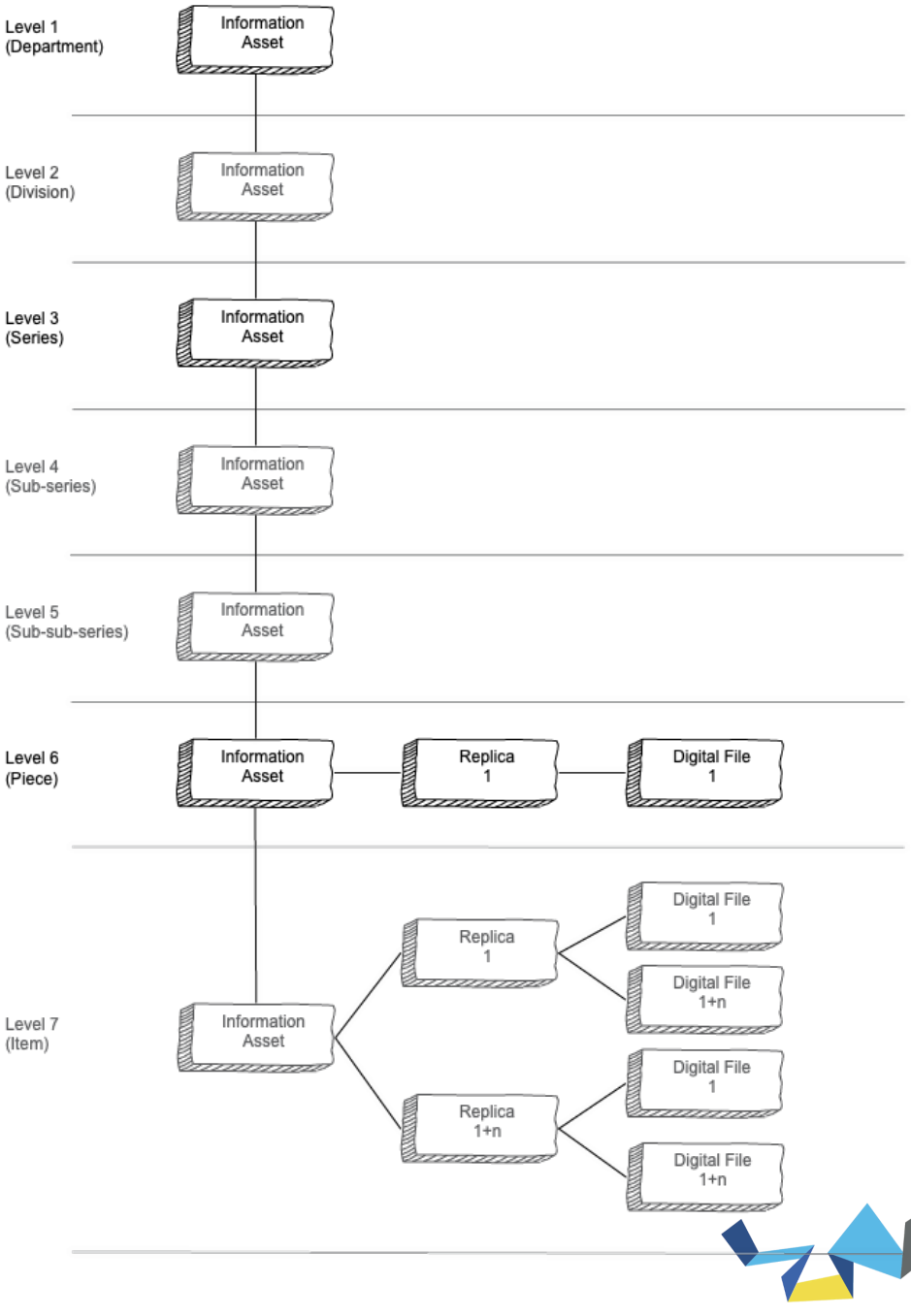
Field	Description	Used for	Mandatory	Type	TNA-CS13 equiv.
Accrls	Accruals	IA, level 3	When applicable	Text	
AcsConds	Access Conditions	IA, levels 1 to 5	Levels 1 and 3	Text	Access
AdmBgr	Administrative Background	IA, levels 1 to 7	Levels 1 and 2, when applicable	Text	Administrative/ Biographical Background
ApprInfo	Appraisal Information	IA, levels 1 to 7	When applicable	Text	Appraisal/ Destruction Information
Arrmnt	Arrangement	IA, levels 1 to 6	Yes, if diff. level arrangement	Text	
BatchId	BatchId	IA	No	Text	
BReflAID	Browse Parent IAID	IA	No	Foreign Key to IA	
CatId	Catalogue ID	IA	No	Foreign Key to ILDB	Reference
CFrmDt	Covering From Date	IA, levels 1 to 7	Yes	Text	Covering Dates
Clsr	Closure	IA, levels 6 and 7	Yes; Record Opening Date (if closed)	Foreign Key to Clsr	Closure Type Closure Code Record Opening Date Closure Status
CorpNs	Corporate Names	IA, levels 1 to 7	No	Foreign key(s) to CorpBody	Index Terms: Corporate
CovDts	Covering Dates	IA, levels 1 to 7	Yes	Text	Covering Dates
CpsInfo	Copies Information	IA, levels 1 to 7	No	Foreign key(s) to CpInfo	Copies Information
CrtrNames	Creator(s)	IA, levels 1 to 7	Highest level	Foreign key(s) to CrtrName	Creator Name(s)
CToDt	Covering To Date	IA, levels 1 to 7	Yes	Text	Covering Dates
CustHist	Custodial History	IA, levels 1 to 6	When applicable	Text	Custodial History
Dgtzd	Digitised	IA	No	Boolean	
Dim	Dimensions	IA, levels 6 and 7	No	Text	Dimensions
Files	Files	RE	Yes	Foreign key(s) to Digital File	

Field	Description	Used for	Mandatory	Type	TNA-CS13 equiv.
Format	Format	DF	Yes	Text	
FRefDep	Former Reference (Department)	IA, levels 3 to 7	When applicable	When applicable	Former Reference (Department)
FRefPro	Former Reference (PRO)	IA, levels 3 to 7	When applicable	When applicable	Former Reference (PRO)
HeldBys	Held By(s)	IA	No	Foreign key(s) to XRef	Place of Deposit
IAID	Information Asset ID	IA	Yes	UUID	Reference
ImmSrcOfAcs	Immediate Source of Acquisition(s)	IA, levels 1 to 6	Highest level	Foreign key(s) to ImmSrcOfAc	Immediate Source of Acquisition
Lang	Language	IA, levels 1 to 7	Highest level	Text	Language
LglStts	Legal Status	IA, levels 1 to 7	Highest/ Orderable levels	Enumeration	Legal Status
Links	Links	IA	No	Foreign key(s) to XRef	
LocOfOrigs	Location of Originals	IA, levels 1 to 7	When applicable	Foreign key(s) to LocOfOrig	Location of Originals
LvlId	Catalogue Level ID	IA	Yes	Number	
MapDes	Map Designation	IA, levels 6 and 7	When applicable	Text	Map Designation
MapScNum	Map Scale Number	IA, levels 3 to 7	When applicable	Number	Map Scale Number
MD5	MD5 Checksum	DF	Yes	Text	
Name	Name	DF	Yes	Text	
Note	Note	IA, levels 1 to 7	No	Text	Note
OriginalName	Original Name	DF	Yes	Text	
Origination	Origination	RE	Yes	Enumeration	
People	Personal Names	IA, levels 1 to 7	No	Foreign key(s) to Person	Index Terms: Personal Names
PhysCond	Physical Condition	IA, levels 3 to 7	No	Text	Physical Condition

Field	Description	Used for	Mandatory	Type	TNA-CS13 equiv.
PhysDescExtnt	Physical Description (Extent)	IA, levels 1 to 7	Levels 1 to 3	Text	Physical Description (Extent)
PhysDescFrm	Physical Description (Form)	IA, levels 1 to 7	Levels 1 to 3	Text	Physical Description (Form)
PIAID	Parent IA ID	IA	Levels 2 to 7	Foreign key to IA	
Places	Places	IA, levels 1 to 7	No	Foreign key(s) to Place	Index Terms: Places
PublNotes	Publication Note(s)	IA	No	Text(s)	Publication Note
Ref	Reference	IA, levels 1 to 7	Yes	Text	Reference
RelMats	Related Materials	IA, levels 1 to 7	No	Foreign key(s) to IA	Related Material
RepIIdents	Replica ID(s)	IA	No	Foreign key(s) to Replicas	
RRec	Record	IA	No	Foreign key(s0 to RegRecs	
RstrOnUse	Restrictions on Use	IA, levels 1 to 7	Highest level	Text	Restrictions on Use
SepMats	Separated Materials	IA, levels 1 to 7	No	Foreign key(s) to Its	Separated Materials
SC	Scope and Comntent	IA, levels 1 to 7	Levels 1 to 3, 6 and 7	Foreign key to SCont	Scope and Content
Size	Size	DF	Yes	Number	
Src	Data Source	IA	Yes	Enumeration	
SrtKey	Sort Key	IA	No	Text	
Subjects	Authority Terms	IA, levels 1 to 7	No	Foreign key(s) to Subject	Index Terms: Subjects
Summary	Summary	RE	No	Text	
TotalSize	Total Size	RE	Yes	Number	
Ttl	Title	IA, levels 1 to 7	Levels 1 to 5	Text	
UnpFindAids	Unpublished Finding Aids	IA, levels 1 to 7	No	Text	Unpublished Finding Aids
Version	Version	RE	Yes	Number	

- IA - Information Asset
- RE - Replica
- DF - Digital File
- Highest level - Mandatory at the highest level to which it applies.
- Highest/Orderable levels - Mandatory at the highest level to which it applies, and for every Orderable unit (whether Piece or Item).
- When applicable - if the information is known and/or accurate for that level.

Fig 5 - BIA (Discovery) - Record Levels



EVOLVED BINARY

5.3.3 Strengths of BIA Model

1. Unlike TNA-CS13, the BIA Model has the concept of a *Replica of an Information Asset (record)*, which allows for the realisation of multiple “versions” or “formats” of a record. This can support digital surrogates, redactions, and digital migrations.
2. Unlike TNA-CS13, the BIA Model has the concept fo a Digital File. This allows it to describe technical metadata about the file, and directly link to the content of the file. Thus making it also suitable for digitised, and born-digital records.

5.3.4 Weaknesses of BIA Model

1. The BIA internally uses UUIDs for record references. However, these have leaked through APIs and are visible externally, e.g. via the Discovery website. These are not suitable for use by humans.
2. The BIA as implemented in Discovery, externally uses TNA-CS13 CCRs and DRI GCRs, and so inherits the weaknesses of both of those.
3. The BIA as implemented in Discovery, is effectively restricted to the 7 levels of TNA-CS13, and so is not suitable for storing more complex arrangements of records, e.g. Born Digital.
4. Storing additional structured record metadata in Discovery is possible but expensive and inflexible. It requires pre-knowledge of the metadata structures, from which changes to both the MongoDB database and .NET application code have to be made.
5. The BIA’s *Replica* and *Digital File* entities are similar to the DRI Catalogue Model’s *Manifestation* and *Digital File* respectively. However the BIA implementations are less flexible than DRIs, as the BIA operates a closed-world model, vs DRI’s open-world model.

5.4 UK Government Web Archive Model

TBC

5.5 ePRO Model

TBC

6. REQUIREMENTS FOR A PAN-ARCHIVAL CATALOGUE MODEL

This section sets out the requirements of a data model for a pan-archival catalogue to form the basis of Project Omega. For evaluating potential data models, we also detail a set of test cases in section 7. Each test case describes a complex arrangement of some record type that must be expressible by any such potential model.

6.1 Conceptual Vocabulary

From our analysis of existing catalogue models at TNA, it can be seen that many of the models have similar concepts and properties. Most of the existing models have a strong ISAD(G) heritage, or otherwise could be mapped to an ISAD(G) like model. However, for all their similarities, one model sometimes uses the same term as another model to mean something different, or conversely two models use different terms for a concept that is semantically equivalent. Obviously such conflicts can easily lead to both confusion and ambiguity.

In an attempt to reduce future confusion, we define a vocabulary of core conceptual terms to create a common language around catalogue records. This serves for the purpose of discussing analogues between differing models. The definitions are left intentionally thin so as to allow wider meanings to be inferred.

6.1.1 Record

Formally, a *Public Record* which is the main concern of TNA is defined by the Public Records Act, 1958. This document however infers a more abstract concept of a *Record*, of which a *Public Record* could be considered a subclass.

A *Record* is an abstract concept, and the entity of main focus. It is something that can be referenced, and delivered upon request (in some manifest form) assuming that *Access Constraints* are met. The record's medium/format is not of importance, it could be of paper (physical), digital, or any other form.

NOTE: For our definition, a manifestation of "The Record" encompasses *both* the asset (i.e. a physical (paper) record or digital byte-stream), and any associated descriptive and technical metadata.

6.1.2 Record Set

A *Record Set* is just a grouping of records which are related by one or more common properties. Like a Record, a Record Set can be referenced, and may be deliverable (in some manifest form) upon request assuming that there is both, an Access Channel, and that Access Constraints are met.

Across TNA's existing catalogue systems, there are many different terms in use which allow the grouping records into a record set, including: Collection, Series, Piece, parent Deliverable Unit, and parent Information Asset.

6.1.3 Manifestation

A Record may have multiple manifestations. These manifestations may also be referenced. It is the manifestation of the record that is delivered upon request.

For paper, there is often only one manifestation which is the original paper record, although a redaction can lead to a second, or even third instance.

For a born-digital record, due to the ethereal nature of digital information there is no global consensus on what forms the *original record.* For the purposes of this document, the original born-digital record is the Digital Files and associated metadata making up a single record as accessioned into TNAs Digital Preservation system (DRI). For most purposes, the first manifestation of a born-digital record can be considered the *original* form.

6.1.3.1 Redacted Manifestation

Occasionally the activity of redaction has to be applied to a record. The activity itself is applied to an existing manifestation of a record to create a new redacted manifestation of that same record.

A redaction is a type of manifestation.

6.1.3.2 Surrogate Manifestation

It is often desirable to create a surrogate of a record through the activity of substitution. The surrogate is commonly a digital representation, common known as a "*digital surrogate*". The creation of such a surrogate could be for the purpose of delivering or preserving a paper record digitally, or transcoding a born-digital record for delivery or presentation.

A surrogate is a type of manifestation.

A digital surrogate could be an additional manifestation of a paper record. For born-digital records there could be many surrogate manifestations for presentation purposes.

6.1.4 Digital File

For original paper records there is no digital file, however a digital-surrogate *manifestation* of a paper record could have one or more digital files. For born-digital records, a single manifestation may be made up of several

digital files. These digital file may have complex technical metadata associated with them, this is especially pertinent for preservation purposes.

6.1.5 Activity

An activity describes a task or process that occurred; activities have a temporal dimension. An activity may be performed upon a Record, Record Set, Manifestation, or Digital File. The simplest process could for example be the creation of a Record Set.

6.1.5.1 Redaction

Redaction is an activity which is applied to a record. Redaction is the process of censoring or obscuring part(s) of a record for legal, political, or security purposes. In practice, redaction is applied to a manifestation of a record, to generate a new redacted manifestation. The process of redaction may cause other non-redacted manifestations to have access constraints imposed upon them.

The process of redaction has a temporal aspect, the redactions (or parts thereof) may only be imposed for a period of time, after which the non-redacted manifestation(s) again become authoritative.

The process of redaction has an access aspect, the access constraints of manifestations which exist before the process of redaction may need to be updated for certain access channels. Any such imposed access constraints may then again need to be updated in future inline with the temporal aspect of the redaction. Example of this include:

1. When there is a paper record with a digital surrogate (manifestation), and then at a later date the paper record is redacted, we must address the question - What is the access status of the non-redacted digital surrogate? Should that no longer be accessible, and in-response do we have to create a redacted digital surrogate as well?
2. When there is a born-digital record with additional manifestation(s) (for preservation or presentation purposes), and then at a later date the born-digital record is redacted, we must address the question - What is the access status of the additional manifestation(s)? Should that no longer be accessible, and in-response do additional redacted manifestations need to be generated?

6.1.5.2 Substitution

Substitution is an activity which is applied to a record. Substitution is the process of creating a copy of a record which will be used in place of the record itself for access purposes. In practice, substitution is applied to a manifestation of a record, to generate a new surrogate manifestation. The process of substitution may cause other non-surrogate manifestations to have access constraints imposed upon them.

The process of substitution has an access aspect, each surrogate may only be applicable to specific access channels.

6.1.5.3 Migration

Migration is an activity which is undertaken on digital records to ensure the digital preservation of a record in the face of file format obsolescence. In practice, substitution is applied to a manifestation of a record, to generate a new surrogate manifestation. Migration involves the transcoding of the digital files of a manifestation, to create a new manifestation with new digital files encoded in a file format which is more widely accessible by available technology at that point in time.

6.1.5.4 Closing

Closing is an activity which is applied to a record. Closing is the process of adding access constraints to a record for the purposes of FOI (Freedom of Information) exemption. There may be a variety of access constraints that can be imposed.

Historically, records had three levels of access constraint:

1. Closed
2. Open Description, Closed Content
3. Open

When a record was closed or had its content closed, there was a temporal aspect to the access constraint imposed on the record whereby at a date in the future, the record may either be: 1) considered open, or 2) have its access constraints reviewed.

Such access constraints may only apply to some access channels. It is worth noting that records may have their access constraints adjusted more than once, for example if the activity of Closing was undertaken in error (or part thereof), there may be a subsequent process of Opening. This alternation could also occur more than once due to an FOI review and subsequent challenges to that review.

6.1.5.5 Opening

Opening is an activity which is applied to a record. It is the inverse activity of Closing. Opening is the process of removing some or all access constraints from a record.

6.1.6 Agent

An agent is a person, organisation, computer program, or some such source, which is attributable as the instigator of an activity. An agent is used to attribute ownership of an activity within the model.

6.1.7 Access Channel

This is the route by which a record is accessed.

For paper records, there are several access channels at TNA, with the most common being user based record ordering via the Reading Rooms.

For digital records, there is a huge variety of possible access channels. Such channels may be restricted based on the size of the Record(s) or Record Set(s) and available transfer bandwidth.

6.1.8 Access Constraint

An access constraint intersects both whom is allowed to access a record, when they are allowed to access it, and the channel by which they may access it.

6.2 Functional Requirements

This section sets out a number of FR (Functional Requirements) for any potential new model.

6.2.1 FR1 - Record Scope

Any such future model MUST respect that a Record is composed of both: 1) the physical/digital record content (e.g. paper and/or digital files), and 2) the metadata about that record (both arrangement and description).

6.2.2 FR2 - Medium Independent

Any such future model MUST be able to describe any record regardless of its medium; At a minimum: Physical (e.g. Paper, Maps, Folders, Analogue Audio Cassette, Microfilm), Digitised (2D and 3D), and Born Digital mediums must be accommodated.

6.2.3 FR3 - Independent of Record Locality

Any such future model MUST be able to describe records independent of the location of their content. This may include records which are held by TNA, retained (Physical, and Born Digital) by the transferring organisation, or held by an external entity (e.g. Person, Organisation, or Department).

6.2.4 FR4 - Separate Arrangement and Description

Any such future model MUST be able to both describe and arrange records. However, the parts of the description of the record which are not arrangement dependent SHOULD be able to be held separately, and SHOULD not need to be repeated for differing arrangement purposes.

6.2.5 FR5 - Manifestations of Record

Any such future model MUST support multiple manifestations of a record. This implies a separation between the concept of a record and any realisation(s) of that record.

Any such future model SHOULD split properties of description and arrangement(s) of a record, between the concept of a record (which is likely immutable) and the manifestation(s) of that record.

Activities which may create Manifestations include Redaction and Creation of Digital Surrogates for Preservation and/or Presentation. An example of this is Multi-Spectral imaging of physical records that is being employed by TNA's Collection Care department to provide new insights into existing records. Each multi-spectral image could be considered a new manifestation of the record.

6.2.6 FR6 - Multiple Arrangements

Any such future model MUST be able to support one or more distinct arrangements of the same records. Records may be organised in more than one manner e.g. Physical vs. Logical, or that of the Transferring Organisation vs. archivist collection curation. In particular born digital records may come from systems which are not organised by filesystem hierarchies, but rather views; thus pertaining to multiple arrangements of the same records.

Any such future model MUST be able to able to arrange records into hierarchies of unlimited depth, and allow description of each level. An example of a TNA-CS13 Series that would benefit from a deeper hierarchical arrangement has CCR: HCA 32.

Any such future model MUST also be able to describe and arrange components of records with respect to their relationship of the record.

6.2.7 FR7 - Provenance of Record Evolution

Any such future model MUST be able to support the evolution of records over time. The content, description, and arrangement(s) of a record or record-set may evolve and/or accumulate either independently or in lock-step over time.

Any such future model MUST be able to support expressing the provenance of a record; explicitly it must support describing changes made to the record, by whom, and for what purpose. Activities of this include Redaction, Re-Arrangement, Preservation, and Presentation.

One such example is TNA-CS13 Piece with CCR: HCA 32/111, which was split into multiple pieces (as it could not otherwise be described), but no provenance information about this split was recorded, and so the resultant Pieces appear unrelated at present. Another example, is the process of multi-spectral imaging used by Collection Care which generates provenance data about the imaging process, but is currently uncatalogued.

Any such future model MUST support archivists and records users by facilitating access to the state of a record or record-set at any prior time, and also to its progression.

6.2.8 FR8 - Facilitate Controlled Access

Whilst the exact mechanisms for controlling access to any such future model are undefined and the implementation of such is outside of the model itself, the model SHOULD be able to describe access constraints upon records and/or record-sets for different access channels.

6.2.9 FR9 - Authoritative Entities

Any such future model MUST be able to support both defining and linking to Authoritative Entities (e.g. controlled vocabularies of People, Families, Corporate Bodies, Places and Manors). These Authoritative Entities MUST be able to be reused within the description of records.

Any such future model MUST also support auditing provenance of Authoritative Entities similarly to that of records in section 6.2.7.

6.2.10 FR10 - Extensible

Any such future model MUST support extensibility as an immediate concern. We recognise that how records are described, and arranged will change likely change over time. Any such future model MUST be flexible enough to

accept new vocabularies for description of records. Any such future model MUST be flexible enough to allow the definition and expression of new types of relationships between records and record sets.

Some examples of TNA-CS13 Pieces and Items that would have previously benefited from extensible description are those with CCR: ADM 101/168, RAIL 33/60/559, and C 1/464/59. Each of these has a great deal of unstructured information within its scope and content description, that would have been better stored as structure data.

6.2.11 FR11 - Exclusively for Records

Any such future model SHOULD NOT conflate the metadata (description and arrangement(s)) of records with data or metadata needed for performing business processes such as records management; the model SHOULD be solely concerned with describing and arranging records through time.

7. TEST CASES FOR A PAN-ARCHIVAL CATALOGUE MODEL

To ensure that a potential new model can at least express the current estate of records held by TNA, we define a number of RTC (Record Test Case).

These test cases are designed to show not just the common arrangements and associated metadata when cataloguing records, but also the less usual edge cases. Whilst TNA do not yet hold either, Born Digital records which have a poly-hierarchical structure, or that have been migrated due to preservation concerns, we include such records in the test cases as it is recognised that it is just a matter of time before such records will need to be catalogued.

The purpose of these test cases is to form part of an evaluation criteria for potential catalogue models.

7.1 Test Cases for Paper Records

These test cases ensure that any potential model can model the existing paper records held by TNA, and of which are currently structured according to TNA-CS13.

7.1.1 RTC1 - Piece (3 Levels)

TNA possess a large number of paper records which are arranged within a 3-level hierarchy inline with TNA-CS13.

In current TNA paper cataloguing parlance these would be described as: “paper records catalogued to Piece level”.

Any such future model must be able to express such records, their hierarchal arrangement, and in some form all of the data elements of each hierarchical level of that record as required by TNA-CS13.

7.1.2 RTC2 - Item (4 Levels)

Similarly to RTC1, TNA possess a lesser but still significant number of paper records which are arranged within a 4-level hierarchy inline with TNA-CS13.

In current TNA paper cataloguing parlance these would be described as: “paper records catalogued to Item level”.

Any such future model must be able to express such records, their hierarchal arrangement, and in some form all of the data elements of each hierarchical level of that record as required by TNA-CS13.

7.1.3 RTC3 - Item (7 Levels)

A more complex variant of RTC2, TNA also posses paper records which are arranged within the maximal 7-level hierarchy as defined in TNA-CS13.

In current TNA paper cataloguing parlance these would be described as: “paper records catalogued to Item level”, but compared to RTC3, they have additional levels preceding the Item level, of which additional descriptive metadata is required.

Any such future model must be able to express such records, their hierarchal arrangement, and in some form all of the data elements of each hierarchical level of that record as required by TNA-CS13.

7.1.4 RTC4 - Piece with Item components (3,4 Levels)

Like RTC1, TNA possess a large number of paper records which are arranged within a 3-level hierarchy inline with TNA-CS13, however whilst the record is considered to be the Piece, there may also be a sub-level of Items which form parts of the record.

In current TNA paper cataloguing parlance these would be described as: “paper records catalogued to Piece level, with Item components”.

Any such future model must be able to express such records, their hierarchal arrangement, and in some form all of the data elements of each hierarchical level of that record as required by TNA-CS13.

7.1.5 RTC5 - Series

A Series as defined by TNA-CS13 is not itself a Record, but is rather a tool for arranging records, it can be thought of as a Record Set. A Series is the 3rd hierarchical level of a TNA-CS13 arrangement and has two levels above it, each level has data elements required by TNA-CS13.

This test case is included, as it requires solely describing an abstract arrangement, of which there may be no corresponding records at the time of description; although records may later be added to such an arrangement.

Any such future model must be able to express the abstract concept of an arrangement of records which may be hierarchical. When hierarchical, each level of arrangement must be able to be related, and express in some form the data elements of that level defined by TNA-CS13.

7.1.6 RTC6 - Retained Piece (3 Levels)

Similarly to RTC1, TNA may occasionally catalogue records where the physical record itself has been *retained* by the transferring organisation for legal, political, or security purposes. Whilst this may affect access channels, it should not be conflated with access constraints that may have been imposed due to FOI exemptions or other purposes.

Any such future model must be able to express such records, their hierarchal arrangement, and in some form all of the data elements of each hierarchical level of that record as required by TNA-CS13. In addition, the model must also be able to express:

1. the physical location of the record
2. metadata about the purpose of retention
3. points of contact for enquires about the record at the retaining organisation
4. any changes to points of contact for the record
5. the activity of relocating the record one or more times at a future date

7.1.7 RTC7 - Retained Item (4 Levels)

The same test case as RTC6, expect that the starting point is RTC2 instead of RTC1.

7.1.8 RTC8 - Retained Item (7 Levels)

The same test case as RTC7, expect that the starting point is RTC3 instead of RTC2.

7.1.9 RTC9 - Redacted Piece (3 Levels)

Starting from RTC1, TNA at the request of the creator of a record catalogued to TNA-CS13 Piece level, occasionally has to perform the activity of redaction upon the record.

Any such future model must be able to express:

1. both the original record and its redacted form(s)
2. the activity of redaction itself
3. metadata about the request for redaction and its purpose
4. the relationship between the redacted form of the record and the original
5. that redaction may only apply over a temporal period, after which the non-redacted form of the record again becomes authoritative.
6. multiple redactions of a record; This may occur because because a future redaction is more or less restrictive than a prior redaction. One could even imagine that should a gradated access policy be adopted by TNA, then there could be differing appropriate redaction for various access channels.

7.1.10 RTC10 - Redacted Item (4 Levels)

The same test case as RTC9, expect that the starting point is RTC2 instead of RTC1.

7.1.11 RTC11 - Redacted Item (7 Levels)

The same test case as RTC10, expect that the starting point is RTC3 instead of RTC2.

7.1.12 RTC12 - Digital Surrogate Item (4 Levels)

Starting from RTC2, TNA through the activity of substitution may wish to create a digital surrogate of a paper record catalogued to TNA-CS13 Item level. This digital surrogate may be created for a variety of purposes such as, presentation, or reducing handling of the original paper record.

Any such future model must be able to express:

1. both the original record and its surrogate form(s)
2. the activity of substitution itself

3. metadata about the request for substitution and its purpose
4. the relationship between the surrogate form of the record and the original
5. multiple substitutions of a record; This may occur because different formats and/or qualities of substitution are needed for different purposes.

7.2 Test Cases for Born Digital Records

These test cases ensure that any potential model can model the existing and envisaged digital records held by TNA, most of which are currently structured according to varying ePRO, BIA, and DRI models.

7.2.1 RTC13 - Born Digital (Mono-hierarchical)

TNA have and will continue to receive an increasing volume of born-digital records. Born-digital records are typically organised in one of two ways, 1) a heavily curated arrangement of records, whereby the arrangement of the records may not reflect the file plan of the digital files and associated metadata, or 2) a loosely curated arrangement whereby the arrangement is in fact the file plan of the digital files. The arrangement of these born-digital records are often attached to a larger higher-level arrangement, typically this has been a Series level arrangement as defined to TNA-CS13.

Regardless of the curated arrangement of these born-digital records, the common expression is a mono-hierarchy of *n*-levels. The exact number of levels may vary on a per-record basis, and records may share common ancestors.

Depending on the curation strategy, each born-digital record (which includes the records descriptive metadata), may be composed of more than one digital file or byte-stream. It could even be the case that several born-digital records may share several common digital files.

Any such future model must be able to express:

1. a hierarchical arrangement of born-digital records, of which depth is determined on a per-record basis
2. for records within the hierarchy to have common ancestors
3. unrestricted structured metadata about the description a born-digital record
4. that a born-digital digital record is comprised of one or more digital files or byte-streams
5. the location of a digital file or byte-stream within the system or an external system

6. that several born-digital records, may in fact share some of the same digital files or byte-streams
7. unrestricted structured metadata about about the technical properties of each digital file or byte-stream

7.2.2 RTC14 - Born Digital (Poly-hierarchical)

Similar to the RTC13 test case, but where there is more than one arrangement of the born-digital records. This may occur predominantly for two reasons, 1) the born-digital records originated from a document management system whereby there were not organised by content location, meaning that the same born-digital record could appear in more than one location or view of the arrangement of records, 2) that there is a curated arrangement of the records as well as a digital ordering of the records.

Any such future model must be able to express:

1. that the same born-digital record, may appear in multiple distinct hierarchical arrangements.
2. that the same born-digital record, may appear in more than one position within the same hierarchical arrangement.
3. that an arrangement itself may have associated descriptive metadata, for example a name for a *view* of records.

7.2.3 RTC15 - Retained Born Digital (Mono-hierarchical)

Similarly to RTC13, TNA may occasionally catalogue records where the digital files comprising the record itself have been *retained* by the transferring organisation for legal, political, or security purposes. Whilst this may affect access channels, it should not be conflated with access constraints that may have been imposed due to FOI exemptions or other purposes.

In addition to the requirements of RTC13, any such future model must also be able to express:

1. the location of the digital files; this could be a URI, the geographical address of an organisation, or some sort of fine grained external system identifier.
2. metadata about the purpose of retention
3. points of contact for enquires about the digital files of the record at the retaining organisation
4. any changes to points of contact for the record
5. the activity of relocating the digital files one or more times at a future date

7.2.4 RTC16 - Retained Born Digital (Poly-hierarchical)

The same test case as RTC15, expect that the starting point is RTC14 instead of RTC13.

In addition to the requirements of RTC15, any such future model must also be able to express:

1. the relationship(s) between the redacted form of the record and the original. Where the original may appear in multiple hierarchies or multiple locations within the same hierarchy, the redacted form must related to the original record and additionally be visible from each appearance of the original.

7.2.5 RTC17 - Redacted Born Digital (Mono-hierarchical)

Starting from RTC13, TNA at the request of the creator of a catalogued born-digital record, occasionally has to perform the activity of redaction upon the record.

In addition to the requirements of RTC13, any such future model must also be able to express:

1. both the original record and its redacted form(s)
2. the activity of redaction itself
3. metadata about the request for redaction and its purpose
4. the relationship between the redacted form of the record and the original
5. that redaction may only apply over a temporal period, after which the non-redacted form of the record again becomes authoritative.
6. multiple redactions of a record; This may occur because because a future redaction is more or less restrictive than a prior redaction. One could even imagine that should a gradated access policy be adopted by TNA, then there could be differing appropriate redaction for various access channels.

7.2.6 RTC18 - Redacted Born Digital (Poly-hierarchical)

The same test case as RTC17, expect that the starting point is RTC14 instead of RTC13.

In addition to the requirements of RTC17, any such future model must also be able to express:

1. the relationship(s) between the redacted form of the record and the original. Where the original may appear in multiple hierarchies or multiple locations within the same hierarchy, the redacted form must related to the original record and additionally be visible from each appearance of the original.

7.2.7 RTC19 - Born Digital Migrated (Mono-hierarchical)

Starting from RTC13, TNA through the activity of migration may wish to create a digital surrogate of a born-digital record. This digital surrogate is created for the purpose of digital preservation due to file-format obsolescence. It should be recognised that during a migration, as a manifestation of a record may be composed of multiple digital files of differing formats, only some of the digital files will be migrated.

Any such future model must be able to express:

1. both the original record and its surrogate form(s)
2. the activity of migration itself
3. metadata about the request for migration, its purpose, and the required file formats and tools
4. the relationship between the surrogate form of the record and the original
5. when only some digital files are transcoded for migration to a new manifestation, the new manifestation must be relatable to both, those new digital files, and any digital files making up the previous manifestation which remained unchanged.
6. multiple migrations of a record; This may occur because over time multiple migrations are undertaken

7.2.8 RTC20 - Born Digital Migrated (Poly-hierarchical)

The same test case as RTC19, expect that the starting point is RTC14 instead of RTC13.

In addition to the requirements of RTC19, any such future model must also be able to express:

1. the relationship(s) between the migrated form of the record and the original. Where the original may appear in multiple hierarchies or multiple locations within the same hierarchy, the migrated form must related to the original record and additionally be visible from each appearance of the original.

7.2.9 RTC21 - Digital Surrogate of Born Digital (Mono-hierarchical)

Starting from RTC13, TNA through the activity of substitution may wish to create a digital surrogate of a born-digital record. This digital surrogate is created for the purpose of presentation.

Any such future model must be able to express:

1. both the original record and its surrogate form(s)
2. the activity of substitution itself

3. metadata about the request for substitution and its purpose
4. the relationship between the surrogate form of the record and the original
5. multiple substitutions of a record; This may occur because different formats and/or qualities of substitution are needed for different purposes.

7.2.10 RTC22 - Digital Surrogate of Born Digital (Poly-hierarchical)

The same test case as RTC21, expect that the starting point is RTC14 instead of RTC13.

In addition to the requirements of RTC21, any such future model must also be able to express:

1. the relationship(s) between the surrogate form of the record and the original. Where the original may appear in multiple hierarchies or multiple locations within the same hierarchy, the surrogate form must related to the original record and additionally be visible from each appearance of the original.

7.3 Test Cases for Access to Records

These test cases ensure that any potential model can model the access constraints upon records that may be required by TNA.

7.3.1 RTC23 - Restricting and Permitting Access to Records

There are many cases in which TNA may need to restrict or permit access to records, some of these activities such as Redaction, Closing and Opening have already been covered. In the more abstract sense, for any record or record set, TNA needs to be able to change both the access constraints, the available access channels, and the intersection of those two properties over time.

Any such future model must be able to express:

1. access constraints upon a record or record set, or manifestation thereof
2. metadata about the purpose of any access constraints
3. the access channels to a record or a record set, or manifestation thereof
4. changes to access constraints and channels over time

8. EXISTING APPLICABLE STANDARDS

Alongside the previously described catalogue data models in use at TNA, there are a number of new or emerging standards which could potentially be used for a new pan-archival catalogue data model. This section discusses each in turn, and examines their suitability.

8.1 TNA-CS13 Model

The TNA-CS13 Model, and its strengths and weaknesses have already been discussed in detail in section 5.1.

TNA-CS13 whilst in the most part well suited to physical records, is in its current form not suitable for a pan-archival catalogue model due to the following deficits:

1. lacking data elements for describing digital records (fails FR2, RTC13-22)
2. no facility to describe manifestations of a record (fails FR5)
    1. Digital Surrogates are uncatalogued (fails RTC12)
    2. Redactions are currently handled through adding distinct unrelated catalogue entries (fails RTC9-11)
3. arrangement of records is restricted (fails FR4)
    1. strictly mono-hierarchical (fails FR6, RTC14,16,18,20,22)
    2. adherent to a maximum of 7 levels (fails RTC13-22)
4. lack of extensibility (fails FR10)
    1. only extension point is EADv1 content for some data elements; other vocabularies are unsupported
5. ILDB implementation restrictions
    1. only one level of provenance; current and previous state is tracked (fails FR7)

Whilst one could envisage possibly extending or deriving a new model from TNA-CS13, it would likely be sufficiently different from TNA-CS13 as to be unrecognisable as such. We believe that other readily available and more modern options may be better suited, as such we will not consider TNA-CS13 further; although many of the principles of TNA-CS13 and ultimately its underlying ISAD(G) base will resurface.

8.2 DRI Catalogue Model

The DRI Catalogue Model, and its strengths and weaknesses have already been discussed in detail in section 5.2.

The DRI Catalogue Model was designed just for digital records within the DRI system, and in its current form is not suitable for a pan-archival catalogue model due to the following deficits:

1. lacking data elements for describing physical records (fails FR2, RTC1-12)
2. arrangement of records is restricted (fails FR4)
    1. strictly mono-hierarchical (fails FR6, RTC14,16,18,20,22)
3. no modelling of Authoritative Entities (fails FR9)
4. DRI implementation restrictions
    1. technical description of records is split between SDB and the DRI Catalogue (fails FR11)
    2. file-system folders are currently modelled as Deliverable Units which restricts multiple-arrangements of a Digital File, and restricts the possibility to de-duplicate identical files (fails FR4)
    3. provenance is only stored for digitisation of records (fails FR7)

It would not be particularly difficult to extend the DRI Catalogue Model to handle records of any medium or arrangement, to meet the functional requirements and record test cases. However, the vocabulary and ontologies used for the DRI Catalogue Model were designed at a time when the cataloguing of digital records was a new activity for TNA. The DRI Catalogue Model is also focused entirely on digital records, and to retrofit it for paper records would require some compromises. As such, a new clean model and ontology would likely better reflect a a modern pan-archival approach. However, it should be recognised that DRI and the DRI Catalogue Model demonstrated the importance of separating the concept of a record, from manifestations of that record.

8.3 Business Information Architecture Model

The BIA Model, and its strengths and weaknesses have already been discussed in detail in section 5.3.

The BIA Model attempted to bridge the differences of physical and digital (surrogate) records, holding a copy of the catalogue information required for visual presentation of TNA's catalogue on the web. The BIA Model is quite flexible, in so far as it is not formally defined in documentation, however in practice its implementation in Discovery is both rigid, but frequently changed as new applications are developed. In its current form the BIA Model is not suitable for a pan-archival catalogue model due to the following deficits:

1. lacking data elements for describing born-digital records (fails FR2, RTC13-22)
2. following TNA-CS13, redactions of physical records are currently handled through adding distinct unrelated catalogue entries (fails RTC9-11)

3. arrangement of records is restricted (fails FR4)
    1. strictly mono-hierarchical (fails FR6, RTC14,16,18,20,22)
4. lack of extensibility (fails FR10)
    1. for physical records only extension point is EADv1 content for some data elements; other vocabularies are unsupported
    2. for digital records there is no extension point for additional data elements
5. Discovery implementation restrictions
    1. adherent to a maximum of 7 levels (fails RTC13-22)
    2. file-system folders are currently modelled as Information Assets which restricts multiple-arrangements of a Digital File, and restricts the possibility to de-duplicate identical files (fails FR4)
    3. no provenance (fails FR7)

Whilst the BIA Model can accommodate both physical and digital surrogate records, in many places it closely mirrors TNA-CS13 and sacrifices the ability to describe digital records in favour of following established practices for physical records. Rather than building upon the BIA Model, we believe that other readily available and more modern options may be better suited, as such we will not consider the BIA Model further.

8.4 Encoded Archival Description

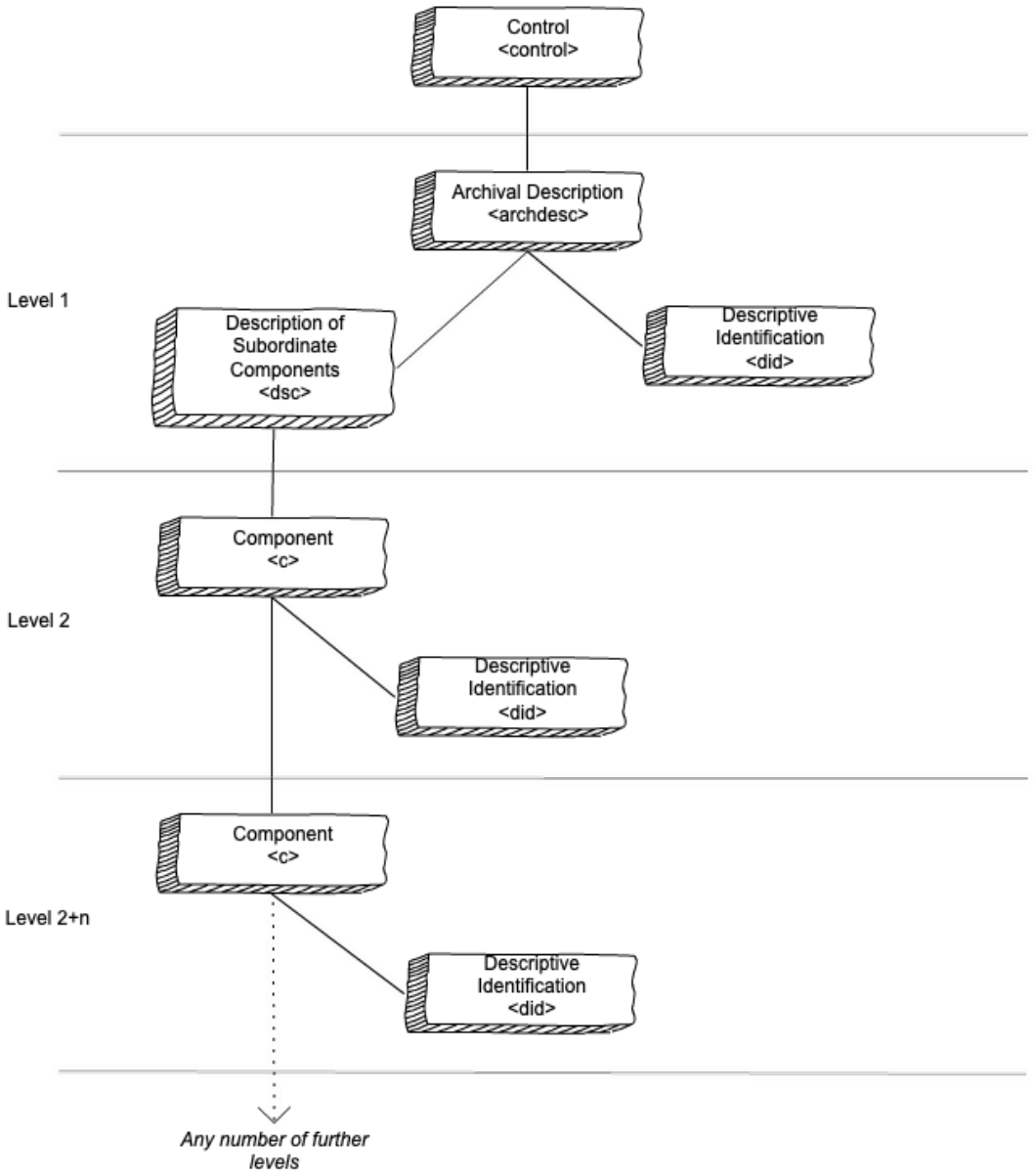
EAD (Encoded Archival Description) is predominantly an XML serialization format for describing archival records. It was developed by the SoAA (Society of American Archivists) and the LoC (Library of Congress) in 1998. EAD was developed to reflect the structure of ISAD(G), and due to its lineage it is often used as an implementation of DACS (Describing Archives a Content Standard; the US implementation of ISAD(G)).

Version 1 of EAD is currently used as both an input and export format for PROCat. The latest version, EADv3 was published in 2015, its goals were - to simplify EAD, remove presentation specific elements, and improve interoperability with other current standards.

EAD does not itself model authority entities as defined by ISAAR (International Standard Archival Authority Record), for such purposes it needs to be augmented with EAC-CPF (Encoded Archival Context - Corporate bodies, Persons and Families). However, unlike ISAD(G), EAD also offers first-class support for digital records, allowing descriptions to be linked to both digital surrogates and born digital records.

Due to its reflection of ISAD(G), EAD emphasises the hierarchical nature of archival description and inheritance of description, and it is very much designed for modelling archival records in mono-hierarchies. EAD does not impose a limit on the number of levels of description. Explicitly it models up to 12 levels through its nested `<c1>` to `<c12>` component elements, if further levels are needed, they can additionally be modelled implicitly through any number of `<c>` component elements.

Fig 6 - EADv3 - Record Levels



An example of the paper record WO 1/921/233 held by TNA record expressed in EADv3 XML is available in Appendix 1.1.

Whilst EADv3 combined with a suitable Native XML Database implementation could offer several improvements over TNA-CS13 as implemented in ILDB, the impact of such improvements for TNA would be incremental rather than transformational. EADv3 due to its ISAD(G) lineage takes a hierarchical approach in both its model and implementation (XML). It could be considered part of the *previous generation* of archival standards, other more recently developed archival standards, which we will examine shortly, all follow a graph approach in their model and an RDF (Resource Description Framework) implementation. Such graph based approaches allow to not only model hierarchies, but also more complex arrangements.

EADv3 is not suitable for a pan-archival catalogue model due to the following deficits:

1. lacking data elements for describing digital records (fails FR2, RTC13-22)
2. no facility to describe manifestations of a record (fails FR5)
    1. no explicit support for Digital Surrogates (fails RTC12)
    2. no explicit support for Redactions (fails RTC9-11)
3. arrangement of records is restricted (fails FR4)
    1. strictly mono-hierarchical (fails FR6, RTC14,16,18,20,22)
4. lack of extensibility (fails FR10)
5. no provenance (fails FR7)

8.5 Data Catalog Vocabulary

DCAT (Data Catalog Vocabulary) version 2 is a proposed W3C standard as of November 2019 which describes itself as:

> *"DCAT is an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web"*

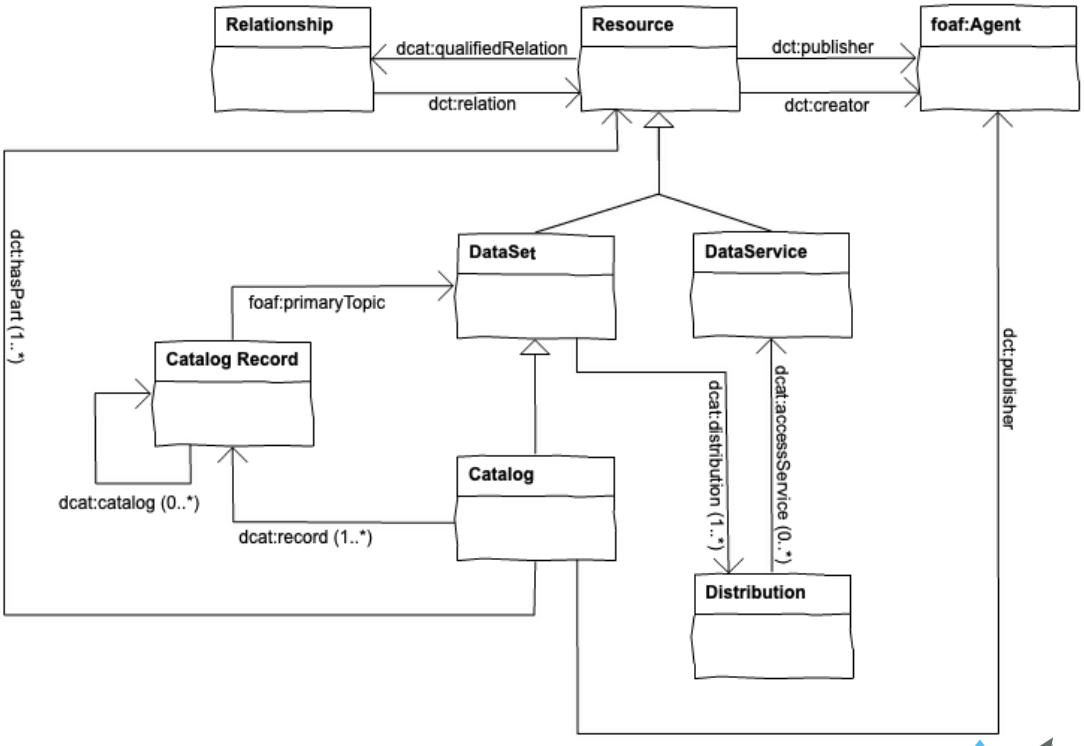
> *"DCAT enables a publisher to describe datasets and data services in a catalog using a standard model and vocabulary that facilitates the consumption and aggregation of metadata from multiple catalogs. This can increase the discoverability of datasets and data services. It also*

makes it possible to have a decentralized approach to publishing data catalogs and makes federated search for datasets across catalogs in multiple sites possible using the same query mechanism and structure. Aggregated DCAT metadata can serve as a manifest file as part of the digital preservation process."

DCAT at its core appears to be primarily concerned with governmental organisations being able to describe and exchange *data sets* over the web. The domain language of DCAT is centred around the concept of sets of data, whereas TNA's domain language is document centric and reflects records, and therefore tends to discuss sets of records instead of sets of data. Regardless, with some mapping of domain terms, it would appear possible to express a representation of TNA catalogue records in DCAT.

From its inception DCAT has been designed for extensibility. Unlike some other standards which define their own isolated vocabularies, DCAT reuses appropriate terms from other vocabularies, including: Dublin Core, EU ISA Programme Location Core Vocabulary, FOAF (Friend-of-a-Friend), W3C Open Digital Rights Language, and W3C PROV. Another aspect of extensibility is that for a particular domain, a DCAT AP (Application Profile) may be created. This AP may add or remove properties, enforce or prohibit the use of properties, or subclass core DCAT entities to better describe the domain. One option for TNA could be to establish a Archival Cataloguing AP, which both subclasses and extends DCAT, the former to refine DCAT entities into more suitable domain terms that are familiar to TNA, and the latter to add additional data properties which are specific and appropriate for records metadata.

Fig 7 - DCAT v2 Entities



To express TNA catalogue records in DCAT v2, the following mappings from our Conceptual Vocabulary to DCAT entities would be established:

Basic Mapping of Conceptual Vocabulary to DCAT

Conceptual Vocabulary Term	DCAT Entity/Property
Record	CatalogRecord and/or DataSet
Record Set	Catalog
Manifestation	Distribution
Digital File	?
Activity	Resource - dct:publisher and dct:creator
Agent	foaf:Agent
Access Channel	DataService
Access Constraints	DataService - dct:accessRights and ordl:hasPolicy)

There is no direct mechanism for describing Digital Files in DCAT. Whilst the Distribution entity of DCAT allows one to describe a *package format* and *download URL*, there is no mechanism for describing the constituent parts of that package. Of course, we could represent every digital file as its own DataSet or Distribution, but this seems like an impedance mismatch with the logical arrangement. TNA would likely need to add an additional entity in their own AP to describe the digital files that make up a distribution.

One advantage of DCAT is that it allows one to do some basic modelling of the role of an Agent through its reuse of the `foaf:Agent` entity, such an agent could be an organisation, person, group, or machine process. At present the only catalogue model in use at TNA that has a similar facility is the DRI Catalogue; which uses OPM to model agents and their actions in the act of digitisation. A novel feature of DCAT for which there is currently no equivalent in-use at TNA, is its DataService entity, which offers the ability to explicitly describe services for accessing datasets (e.g. record sets).

An example of the paper record WO 1/921/233 held by TNA record expressed in DCAT v2 RDF (Turtle) is available in Appendix 1.2.

DCATv2 is not suitable for a pan-archival catalogue model due to the following deficits:

1. lacking comprehensive data elements for describing physical and digital records (fails FR2, RTC1-22)

2. no modelling of Authoritative Entities (fails FR9)

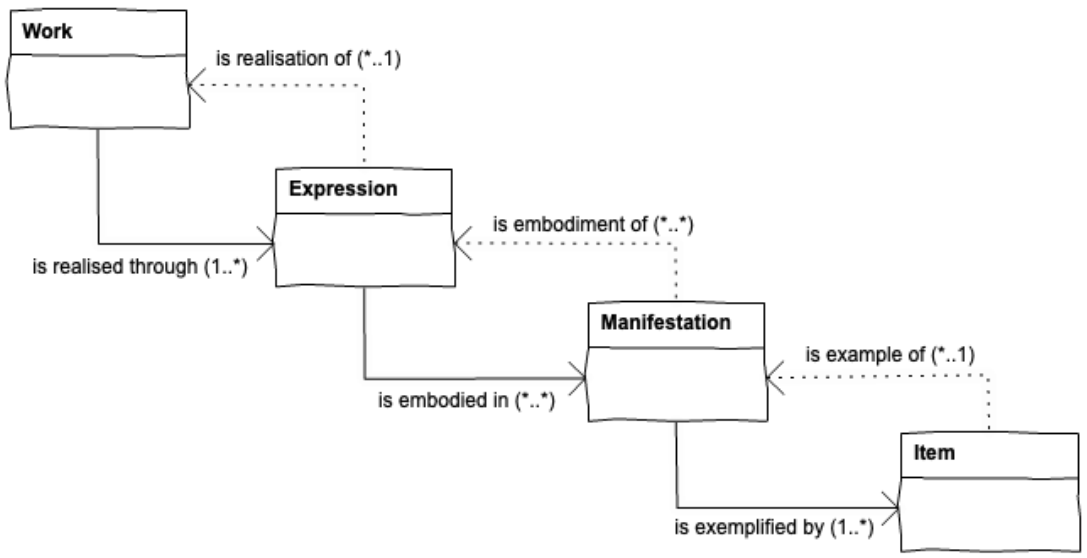
In general whilst a pan-archival catalogue could be expressed in DCATv2, the domain impedance mismatch should not be underestimated. In particular a TNA AP of DCAT would need to augment it with many properties for the description of records. Whilst DCAT could be suitable as an export format, DCAT will not be considered further for a pan-archival catalogue as there are other models which are more closely aligned or malleable to archival practices.

8.6 Functional Requirements for Bibliographic Records

FRBR (Functional Requirements for Bibliographic Records) is a conceptual model developed by the IFLA (International Federation of Library Associations and Institutions) in 1997, and updated through to February 2009.

Whilst FBRB is perhaps more closely aligned with the cataloguing activities of libraries as opposed to archives, one of the most interesting aspects of the FRBR model is its WEMI (Work, Expression, Manifestation, and Item) entities. The WEMI entities separate the abstract concept of what a record is, from different versions and realisations of that record. The relationships between those entities describe how different versions and realisations are connected. This separation of concept and realisation as promoted by FRBR is reused in several other models including: RDA and BIBFRAME.

Fig 8 - FRBR WEMI Entities

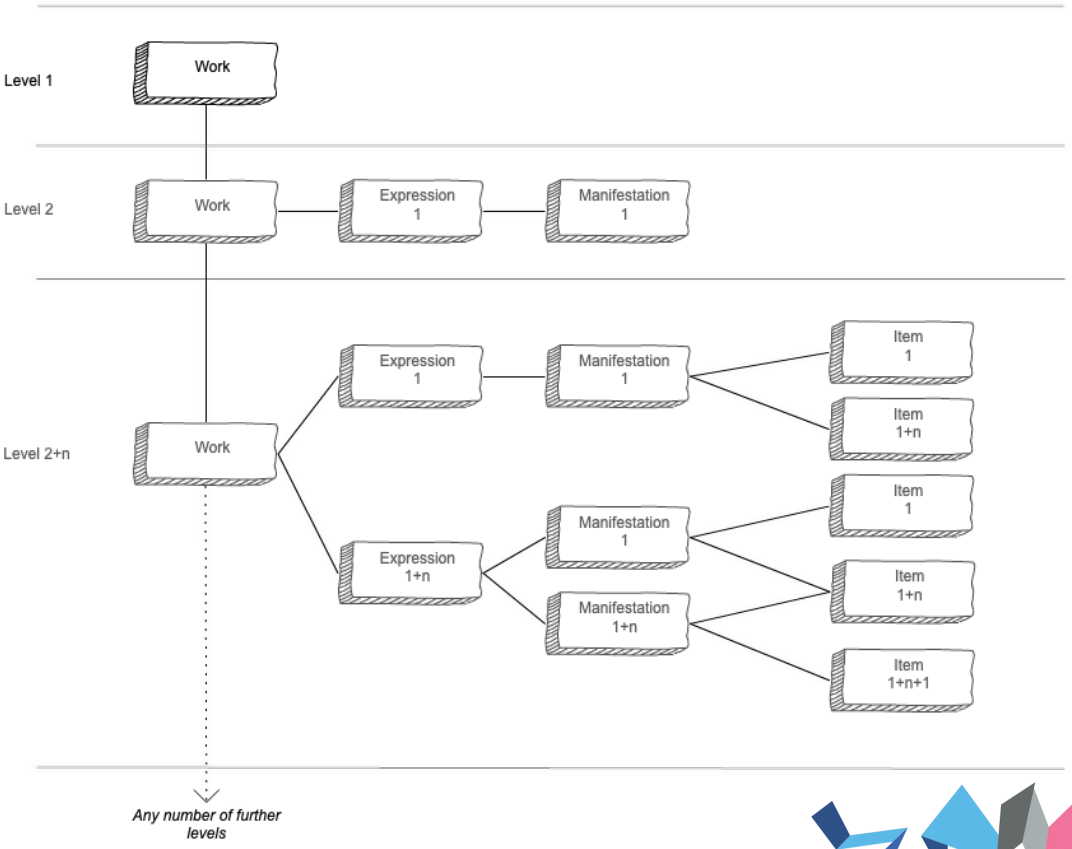


The latest standard from the IFLA is actually the LRM (Library Reference Model) which was released in 2017. The LRM incorporates FRBR, FRAD (Functional Requirements for Authority Data), and FRSAD (Functional Requirements for Subject Authority Data). However, we have focused on FRBR here, because its pioneering of the WEMI approach is of particular interest, and LRM is likely less familiar and adds little over FRBR for our purposes.

The FRBR model also defines authority entities for individuals and corporate bodies, from which relationships with WEMI entities can then be defined. Unlike any previous model we have discussed, FRBR then goes one step further and also defines authoritative entities for: concepts, objects, events and places, these can all have subject relationships with Work entities. The FRBR model additionally defines a set of attributes for each entity for the purposes of describing the entity.

Whilst the FRBR model is predominantly concerned with modelling a record, it also offers relationships for modelling arrangements of records. Between identical WEMI entities, relationships may be established both, horizontally to describe how one record relates to another, and also vertically (whole/part relationships) to describe groupings of records. For example, a record set could be modelled as a Work, with each record of that set also modelled as a Work with a *has part* relationship.

Fig 9 - FRBR - Record Levels



To express TNA catalogue records in FRBR, the following mappings from our Conceptual Vocabulary to FRBR entities would be established:

Basic Mapping of Conceptual Vocabulary to FRBR

Conceptual Vocabulary Term	FRBR Entity
Record	Work
Record Set	Work, and optionally Expression and Manifestation
Manifestation	Expression and/or Manifestation
Digital File	Item
Activity	Event
Agent	Person or Corporate Body
Access Channel	?
Access Constraints	?

There is no direct mechanism for describing Access Channels or Access Constraints in FRBR. Whilst there are some attributes that could be suitable, TNA would likely need to define an additional entities, attributes and relationships to support those requirements.

FRBR provides only a conceptual model, there is no concrete syntax suggested or otherwise for communicating records in a FRBR model. As such there is no appendix in this document for a FRBR syntax example. FRBRoo which was later developed in association with the ILFA in 2006, attempts to bring together FRBR and CIDOC-CRM (International Committee for Documentation - Conceptual Reference Model), and offering a concrete RDF syntax. We discount FRBRoo for the same reasons as FRBR detailed below.

FRBR is not suitable for a pan-archival catalogue model due to the following deficits:

1. lacking comprehensive data elements for describing digital records (fails FR2, RTC13-22)
2. lack of clearly defined extensibility points (fails FR10)
3. lacking comprehensive provenance; only current state is expressed (fails FR7)
4. no facilities for controlled access (fails FR8)

8.7 Resource Description and Access

RDA (Resource Description and Access) is a package of data elements, and guidelines, for creating resource metadata for Linked Data applications that are inline with models from other international standards. RDA was originally conceived in 2004 as AACR v3 (Anglo-American Cataloguing Rules), but renamed to RDA in 2005. RDA was later revised in 2007 to better integrate with other models - primarily the Dublin Core Abstract Model. RDA maintains step with the Statement of International Cataloguing Principles published by IFLA.

RDA is based upon the FRBR and FRAD (Functional Requirements for Authority Data) conceptual models. RDA reuses the WEMI entities from FRBR. Adopting the FRBR Person and Corporate body entities, it adds a Family entity, and refers to this entity set as PFC (Person, Family Corporate Body). RDA is very much focused on expressing relationships between entities, it augments all of the relationships from FRBR with several more - particularly around relationships between separate works.

RDA established a core-level record standard identifying a minimal set of required elements, this can optionally be extended by any implementing organisation. RDA itself is a content standard, it does not enforce a particular presentation or coding standard, often MARC21 or RDF is used; more recently RDA provides RDF vocabularies from its RDA Registry website.

Data elements and entities from RDA vocabularies are reused in BIBFRAME and the Matterhorn RDF Model. Unfortunately official documentation detailing the current RDA standard is not publicly available, accessing it requires committing to a subscription of for the RDA Toolkit, as such we cannot consider it further at this time.

8.8 BIBFRAME Lite + Archive

BIBFRAME initiated in 2012, was updated to BIBFRAME 2.0 in 2016 by the LoC (Library of Congress) to replace MARC (Machine-Readable Cataloging) and make use of Linked Data principles to increase the granularity and reuse of data. BIBFRAME promotes three key design principles:

1. Differentiate clearly between conceptual content and its physical/digital manifestation(s)
2. Unambiguously identify information entities (e.g., agents)
3. Leverage and expose relationships between and among entities

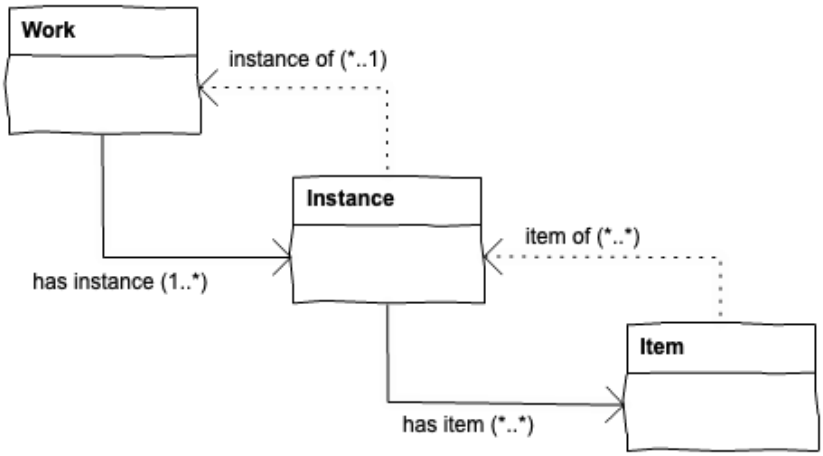
BIBFRAME is developed in two parts, the BIBFRAME Model, a (conceptual) model, and BIBFRAME Vocabulary which defines classes, entities and properties, and is used for implementation. The BIBFRAME Vocabulary adopts many elements from RDA, but places them into its own namespace. It is also worth mentioning that the entire BIBFRAME Vocabulary is extensive, as such an organisation called Zepheria have broken it into distinct

profiles, with the core profile being named BIBFRAME Lite. Our evaluation of BIBFRAME is based on the combination of BIBFRAME Lite and the additional BIBFRame Archive profile.

The separation of the abstract concept of a record from its manifestation in BIBFRAME 2.0 has undoubtedly been influenced by the ILFA's FRBR model. However compared with FRBR, BIBFRAME facilitates a simplified abstract record model, choosing to collapse the 4 levels of record model into just 3. In part this may have been an attempt to improve record description, due to the past confusion and disagreement that emerged about what accurately constitutes each level of the FRBR abstract record model.

BIBFRAME has 3 main entities of abstraction - Work, Instance, and Item; this is somewhat similar to FRBR, albeit BIBFRAME's Work entity is meant as a union of FRBR's Work and Expression entities. Like FRBR and RDA, BIBFRAME also defines additional core entities for describing interactions with records: Agents, Subjects, and Events.

Fig 10 - BIBFRAME 2.0 - Entities



BIBFRAME also provides a horizontal set of relationships for linking different Work(s), where one or more works can be modelled as *expressions* of other works.

To express TNA catalogue records in BIBFRAME 2.0, the following mappings from our Conceptual Vocabulary to BIBFRAME entities would be established:

Basic Mapping of Conceptual Vocabulary to BIBFRAME 2.0

Conceptual Vocabulary Term	BIBFRAME 2.0 Entity
Record	Work
Record Set	Work, and optionally Instance
Manifestation	Instance
Digital File	Item
Activity	Event
Agent	Agent
Access Channel	?
Access Constraints	?

BIBFRAME Lite + Archive is not suitable for a pan-archival catalogue model due to the following deficits:

1. due to its bibliographic heritage it lacks comprehensive data elements for describing digital records (fails FR2, RTC13-22)
2. lack of clearly defined extensibility points (fails FR10)
3. no facilities for controlled access (fails FR8)

8.9 Europeana Data Model

Europeana is the EU digital platform for cultural heritage, it operates as a centralised virtual library. Europeana operates in a portal like fashion, rather than store the digital objects themselves, it instead ingests metadata about the objects provided by a plethora of cultural organisations across Europe. The physical and digital objects themselves are retained by their host organisations, whilst Europeana collates and makes the metadata searchable, clicking though the Europeana website will eventually lead you to a digital representation of the object within the host organisation's own online presence.

The EDM (Europeana Data Model) created in 2010 is used to supply details of cultural artefacts to Europeana by the cultural heritage organisations. EDM was developed to replace the earlier ESE (Europeana Semantic Elements) model, which was deemed to be too constrained. Whereas ESE took a lowest-common-denominator approach to holding metadata from varying cultural heritage institutions, EDM takes the inverse approach and offers an open and extensible model that can accommodate varying community standards (such as EAD for archives). EDM is built upon a foundation of OAI-ORE (Open Archives Initiative - Object Reuse and Exchange), Dublin Core, SKOS, (Simple Knowledge Organisation System), and CIDOC-CRM. ESE was expressed using

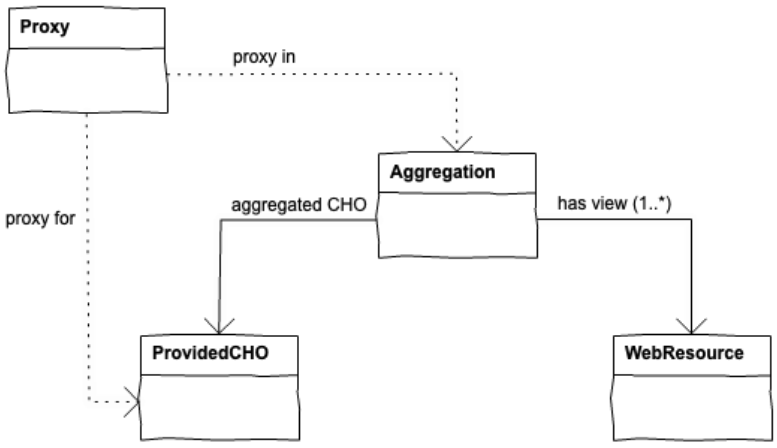
XML Schema, and provided an XML based tree model, whereas EDM uses an OWL Schema and provides an RDF based graph model.

EDM has a several interesting properties about its design, including:

1. distinction between objects and metadata records describing an object
2. allowing multiple records for the same object which may containing potentially contradictory statements about this object
3. compatibility with different levels of abstraction for description
4. allows objects supplied by one provider, to be enriched with metadata from other providers

EDM has 3 main entities of abstraction - ProvidedCHO (CHO - Cultural Heritage Object), WebResource, and Aggregation. EDM also separates the concept of a record from the realisation of a record, but its machinery for separation is much simpler. The ProvidedCHO entity represents the abstract concept of the object (record), in FRBR terms this would likely combine Work and Expression. The WebResource entity represents an accessible digital representation of the object, this is most closely related to FRBR's Instance entity. The Aggregation entity which links the ProvidedCHO and its WebResource(s) by activity. The Aggregation concept is directly reused from OAI-ORE, and models one logical unit which is constituted of the object (ProvidedCHO) together with its digital representations (WebResource(s)).

Fig 11 - Europeana Data Model- Entities

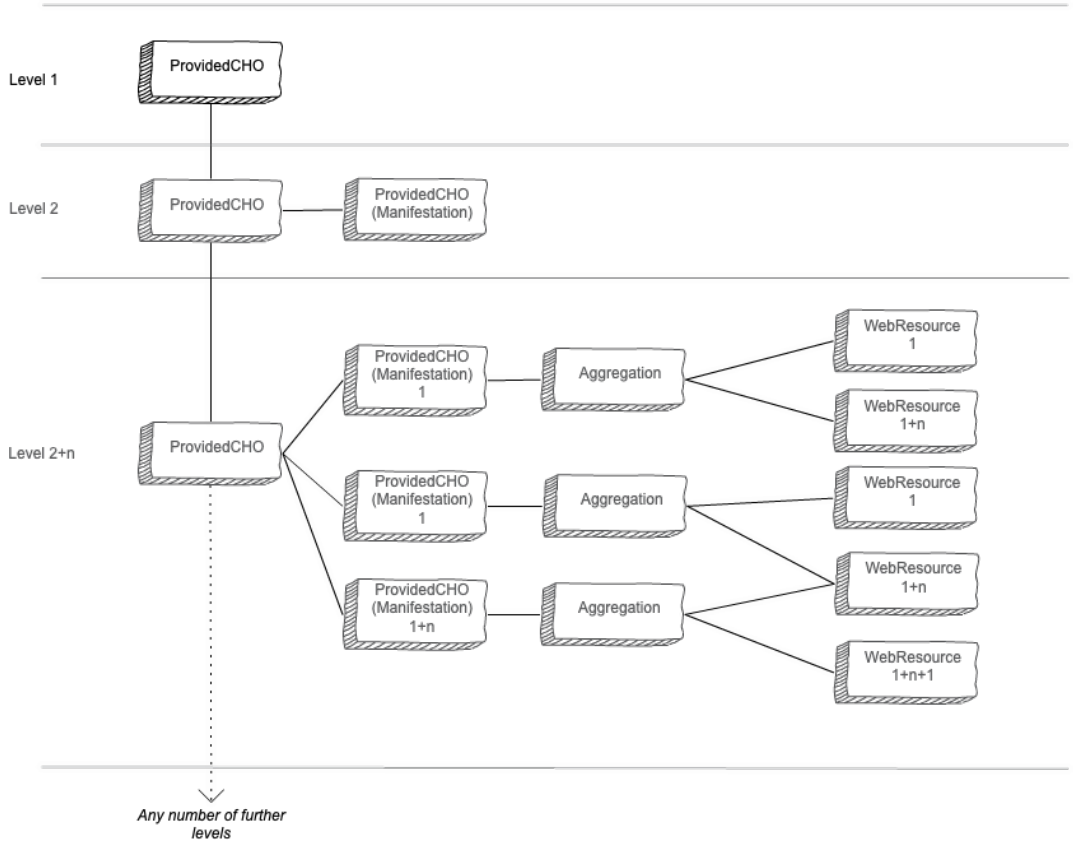


One may be tempted to think of an Aggregation entity as similar to FRBR's Expression entity, however this is not the case, whereas a Work (record) may have many Expressions in FRBR, a ProvidedCHO (record) in EDM may have only one Aggregation (per records provider), instead different expressions must be modelled as separate

ProvidedCHOs. However, a ProvidedCHO may be used to model the most abstract idea of a record (a Work in FBRR terms), and linked to expressions also modelled as ProvidedCHO by establishing the *isDerivativeOf* and/or *isRepresentationOf* relationship between them.

Like FRBR, RDA, and, BIBFRAME, EDM also defines additional core entities for describing interactions with records, it describes these as pertaining to an event-centric view: Agent, Event, Place, TimeSpan, and Concept.

Fig 12 - EDM - Record Levels



Finally it is worth briefly mentioning an additional entity in the EDM core model - Proxy. The Proxy object is used to represent an Aggregation of a ProvidedCHO from a specific provider. This entity is only necessary when records from multiple providing organisations are linked together within the larger Europeana system. When multiple providers wish to describe the same ProvidedCHO (record), they each have their own Aggregation of the record and associated WebResource(s), however instead of directly describing the ProvidedCHO, each

provider instead describes a Proxy which represents a composition of, their Aggregation and the shared ProvidedCHO.

Europeana requires mapping from a variety of specific schemas through to a *semantic interoperability core*. Such mappings are achieved through expressing sub-properties and sub-classes in RDFS and/or OWL, which allows concepts from different schemas to be aligned. To express TNA catalogue records in Europeana, the following mappings from our Conceptual Vocabulary to Europeana entities would be established:

Basic Mapping of Conceptual Vocabulary to EDM

Conceptual Vocabulary Term	EDM Entity
Record	ProvidedCHO
Record Set	ProvidedCHO
Manifestation	ProvidedCHO/WebResource
Digital File	WebResource
Activity	Event
Agent	Agent
Access Channel	?
Access Constraints	?

An example of the paper record WO 1/921/233 held by TNA record expressed in EDM RDF (Turtle) is available in Appendix 1.3.

EDM is not suitable for a pan-archival catalogue model due to the following deficits:

1. lacking comprehensive data elements for describing physical and digital records (fails FR2, RTC1-22)
2. assumes all records have a web-representation* (fails FR2)
3. no facility for controlling access, assumes all parts of all records are publicly available (fails FR8)
4. no modelling of Authoritative Entities (fails FR9)

*EDM does have a PhysicalObject entity, but its use is poorly defined and its experimental status is unclear.

Whilst the core of a pan-archival catalogue could be expressed in EDM, there is a noticeable domain mismatch that should not be underestimated. In particular, EDM's WebResource entity is semantically at odds with how TNA currently thinks about records. The name of the entity itself hints at its purpose, whilst the definition of a WebResource in EDM includes the text "*accessible digital representation*", by which is meant a dereferenceable URI that resolves to an online (Web) representation of the record. It is by no means clear or certain that at this time there is a requirement for all records that TNA describes to be accessible via the web (although this would be an admirable goal). For modelling paper records, EDM also has a PhysicalThing entity which could be used but the standard describes its use as "*less relevant now*". In addition, EDM lacks comprehensive properties for the technical description of paper and digital records, it is likely that we would need to extend EDM with additional properties to describe archival practice and TNAs records, which is perfectly valid.

Similarly to DCATv2, EDM might be better suited as an export format than an internal catalogue model.

8.10 Records in Context

RiC (Records in Context) is a new standard under development by the EGAD (Experts Group on Archival Description) group of ICA for the description of records based on archival principles. RiC builds upon the foundations of the previous ICA standards: ISAD(G), ISAAR (CPF), ISDF, and ISDIAH. RiC as delivered by ICA is split into two parts, RiC-CM (RiC Conceptual Model), and RiC-O (RiC Ontology). RiC-CM can be though of as the descriptive standard itself, whilst RiC-O prescribes one possible syntactic coding implementation of RiC-CM (albeit blessed by the ICA) in RDF.

The development of RiC was initiated in 2012, and is still in progress. To date two drafts of RiC-CM have been released RiC-CM v0.1 in September 2016, and RiC-CM v0.2 in December 2019. Unfortunately RiC-CM v0.2 is both incomplete and unfinished at the time of release. For RiC-CM v0.1, there was no official release of RiC-O, however an unofficial ontology was produced by Dunia Llanes-Padrón and Juan-Antonio Pastor-Sánchez in their 2017 paper: Records in Contexts: the road of archives to semantic interoperability. Subsequently, there has only been a single official release of RiC-O, v0.1 in December 2019, and somewhat confusingly that corresponds to RiC-CM v0.2. The current reported timeline from EGAD, is to have a final version 1.0 of both RiC-CM and RiC-O ready for release in November 2020 at the ICA Congress in Abu Dhabi; based on our observations of progress by EGAD thus far, we feel that this maybe optimistic.

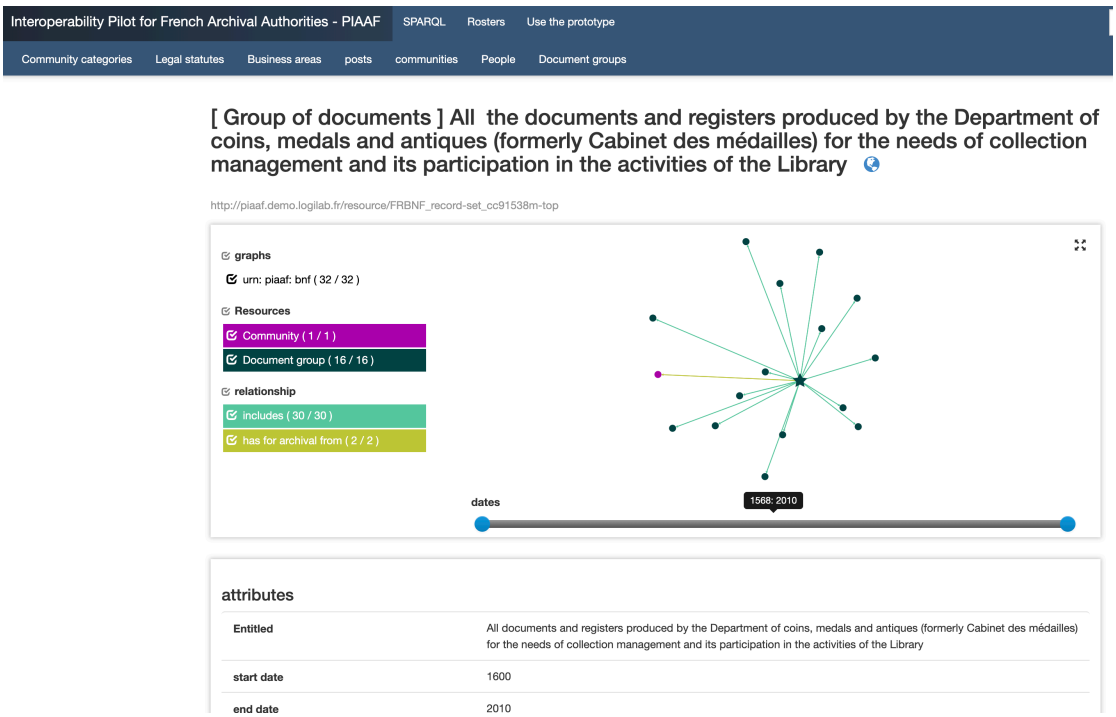
RiC-CM v0.1 was previously analysed by staff at TNA and feedback forwarded to EGAD. Many of the concerns raised by TNA appeared stylistic in nature, however the more serious concerns raised by TNA do in the most part seem to have been addressed in RiC-CM v0.2.

When considering whether RiC is suitable for a pan-archival catalogue model, it is informative to study work already undertaken by the French National Archives in this area. PIAAF (Interoperability Pilot for the French Archival Authorities), is an experimental project for the semanticisation and visualisation of archival metadata

conducted by the National Archives (of France), the Interministerial Service of the Archives of France, and the National Library of France. Resources of interest from PIAAF include:

- Project Issues, Objectives, and History - http://piaaf.demo.logilab.fr/editorial/historique
- Technical Conversion from EAC-CPF and EAD 2002 to RiC-O - http://piaaf.demo.logilab.fr/editorial/contexte-technique
- Demo website for visualising archival records - http://piaaf.demo.logilab.fr/ric/TopRecordSet

Fig 13 - PIAAF - Demo Website for Visualising Archival Records

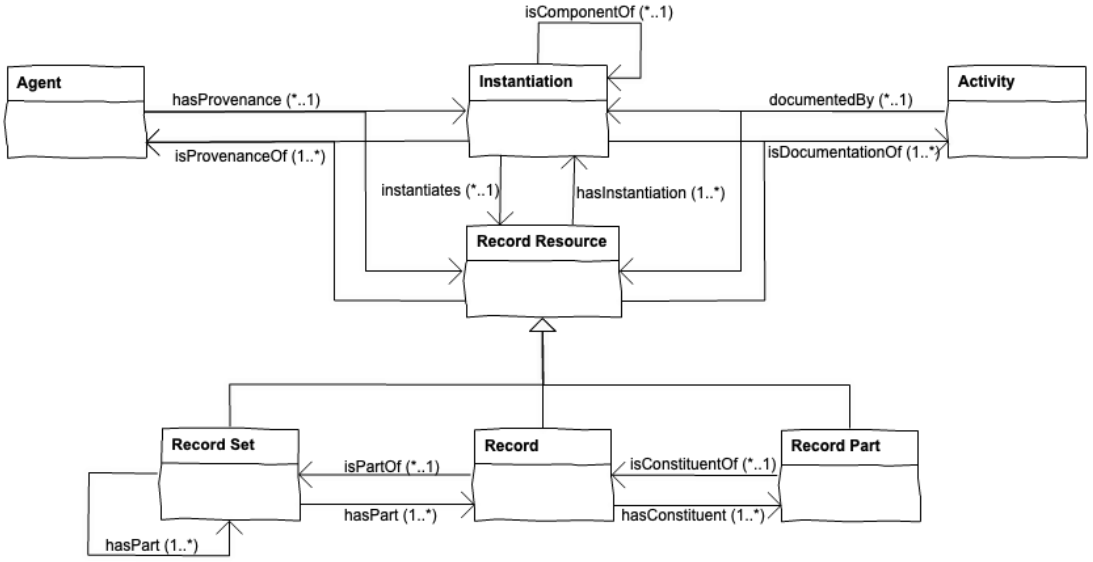


Unlike the ICA standards that predate it, and in particular ISAD(G), RiC-CM does not enforce a mono-hierarchical arrangement of records. RiC-CM allows for the ad-hoc arrangement of relationships between records, instead of a hierarchical model it promotes a more flexible graph based model. Of course, given appropriate relationships, it is still entirely possible to model hierarchical arrangements within a graph, thus allowing backward compatibility with records previously catalogued with ISAD(G). The key concept at the heart t of RiC-CM is that of context(s) (i.e. differing arrangement(s)), as communicated in the RiC-CM v0.1 document:

"… these observations and more lead to the conclusion that provenance is much more complex than the long established understanding of fonds: a fonds does not exist in isolation, but within layers of interconnected contexts, past, present, and future."

RiC-CM v0.2 both simplifies and improves upon the model proposed in RiC-CM v0.1. The are now just four core entities in RiC-CM: Record Resource (subclassed as Record Set Record, and Record Part), Instantiation, Agent, and Activity. Interestingly RiC-CM has a provenance first approach, and the Agent and Activity entities allow for discussing creation and changes to record description through time. Like FRBR, RiC-CM separates the concept of a record from its manifestation, albeit RiC-CM simplifies this even further than that of BIBFRAME, splitting the record into two aspects - its concept (Record Resource), and its manifestation (Instantiation).

Fig 14 - RiC-CM v0.2 Entities (Simplified)

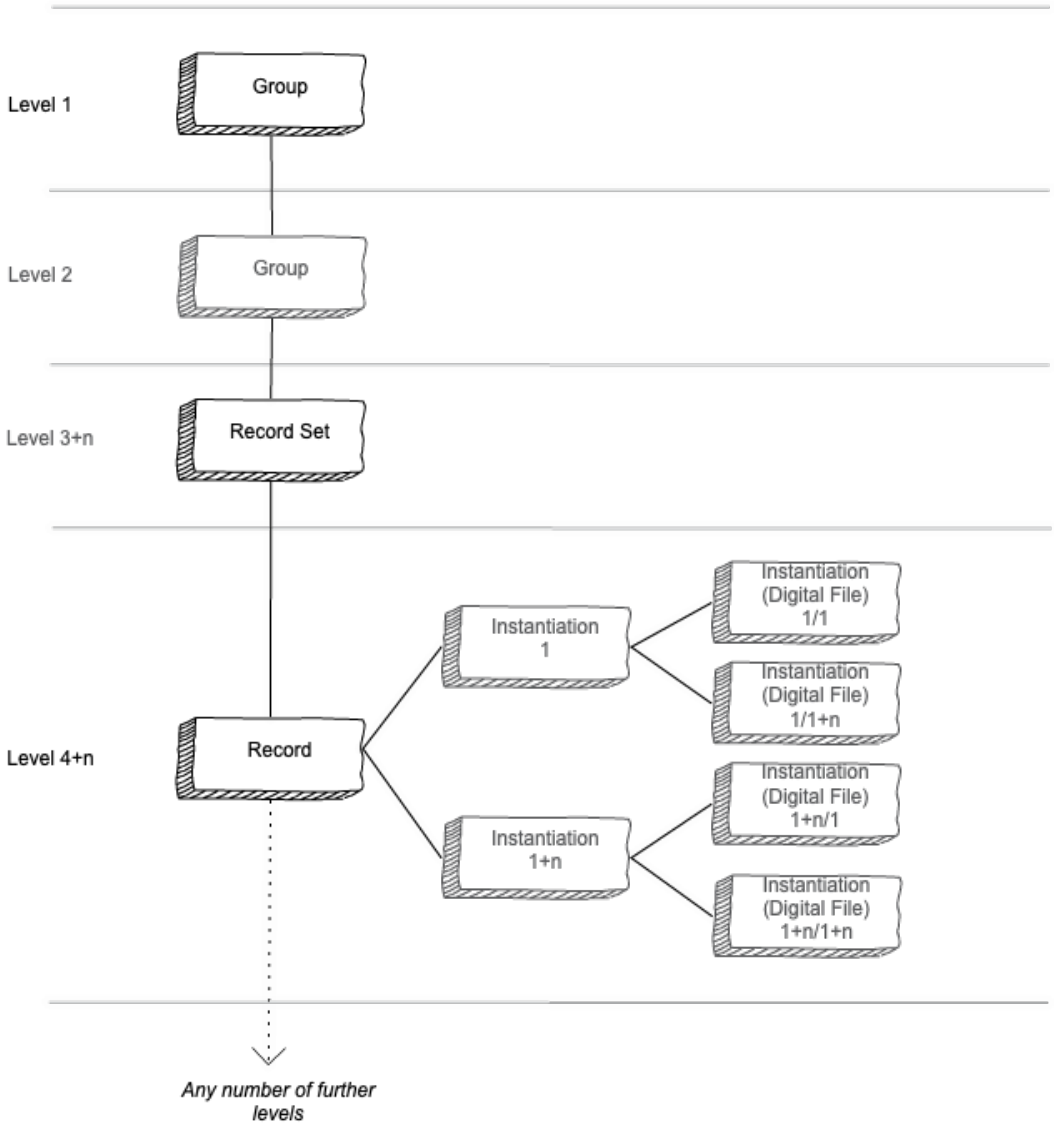


One potential modelling issue with RiC-CM for digital records, is that there is no distinct entity which maps to Digital File from our own conceptual model. There are two possible ways in which we could model an equivalence in RiC-CM:

1. a Record entity may have several *hasConstituent* related RecordPart entities, each of which may have their own instantiation.
2. an Instantiation entity may have several *hasComponent* related sub-Instantiation entities.

Guidance on the best way to model this should likely be sought from EGAD, however based on our current understanding we suspect that (1) would be better suited to describing parts of records (i.e. independent of their instantiations), and (2) would be better suited for when a logical digital instantiation is made up of many files (for example when a Record is a complex page, that needs to be digitised in parts thus generating several files).

Fig 15 - RiC-CM - Record Levels



To express TNA catalogue records in RiC-CM, the following mappings from our Conceptual Vocabulary to RiC-CM entities would be established:

Basic Mapping of Conceptual Vocabulary to RiC-CM

Conceptual Vocabulary Term	RiC-CM Entity
Record	Record
Record Set	Record Set
Manifestation	Instantiation (logical)
Digital File	Instantiation (physical)
Activity	Activity
Agent	Agent
Access Channel	?
Access Constraints	Record Resource's - Conditions of Access (limited)

An example of the paper record WO 1/921/233 held by TNA record expressed in RiC-O v0.1 RDF (Turtle) is available in Appendix 1.4.

RiC-CM is not entirely suitable for a pan-archival catalogue model due to the following deficits:

1. for a digital record it is unclear how the content of the digital file(s) composing that record should be linked to from RiC-CM. Additionally there is a lack of comprehensive data elements for describing digital records (fails FR2)
2. limited text based facilities for describing conditions of access, no explicit facility for computationally controlling access, assumes all parts of all records are publicly available (fails FR8)
3. RiC-CM does not describe any points for extensibility or developing organisation specific Application Profiles (fails FR11)

Whilst we have discussed RiC-CM in some detail, RiC-O has also made some design decisions which may not mesh well with TNA's aspirations of both, describing its records within a larger domain than just archives, and following W3C advice to reuse existing vocabularies to enable interoperability of Linked Data via the Web. RiC-O has chosen to establish its own distinct ontology and vocabulary, which is limited to archival practice. At present EGAD has provided no mapping between RiC-O and other popular ontologies, although their intention to do so has been documented. One other issue, is that much of RiC-O specifies that attributes (data elements) should be free-form text, where as we believe they would have been better defined as either text, URI (for establishing further relationships), or XML literal (to enable some existing applications of Markup).

8.11 Matterhorn RDF Data Model

The Matterhorn RDF Data Model is itself not a standard, rather it is an approach published as a collaborative effort between the Archives de l'Etat du Valais (State Archives of the Canton of Valais) Switzerland and the company Docuteam GmbH Switzerland. The paper - The Matterhorn RDF Data Model, Formalizing Archival Metadata With SHACL, Wildi and Dubois, was published in the proceedings of the 16th International Conference on Digital Preservation, iPRES 2019.

It should be noted that the Matterhorn RDF Data Model published as a research paper, reflects the experience and knowledge of a collaborative effort at a point-in-time. The material published should be considered incomplete, and does not form a finished product that could be adopted whole-sale as a catalogue model. Docuteam GmbH have suggested that Matterhorn RDF is unlikely to be developed further by its technical originators unless they see a demand from the community.

However, the approach described by Matterhorn RDF fits well with TNA's aspirations of both, describing its records within a larger domain than just archives, and following W3C advice to reuse existing vocabularies to enable interoperability of Linked Data via the Web. Matterhorn RDF has already been developed far enough that it can be used as a solid starting point for a catalogue model, although further refinements and/or extensions will likely be needed. As such, we explore the suitability of the Matterhorn RDF Data Model further below.

The Matterhorn RDF Data Model is built upon lessons learn from a previous approach called the Matterhorn METS Profile. As its name might imply, it was a profile for the LoC METS (Makes data Encoding and Transmission Standard) which was registered with, and published by, the LoC in 2012. It described itself as: "*a generic profile that can be used by memory institutions for the management of archival fonds*". Matterhorn METS is an XML model based on METS, PREMIS (Preservation Metadata: Implementation Strategy), EAD and EAC-CPF, which is now in use by around 30 institutions spread across France, Germany, and Switzerland. The key thing about Matterhorn METS, which is interesting here, is that it mixes together two different conceptual models:

1. the first a model around the contextualisation of records (i.e. description/arrangement through context(s): What?, Who?, How?, and When?,
2. and the second model around the preservation of records (i.e. that our understanding, description, and arrangement of records change over time).

Matterhorn RDF keeps the blend of contextualisation and preservation, but departs from a hierarchical model to use a more flexible graph model for describing relationships between records. Thus enabling more complex arrangements of records than just mono-hierarchical. Of course, Matterhorn RDF also provides appropriate relationships so that one can still model hierarchical arrangements within the graph, thus allowing backward

compatibility with records previously catalogued with ISAD(G). It covers the three ICA standards - ISAD(G), ISAAR(CPF), and ISDF, as well as the OAIS Information Model components for “Preservation Description Information” and “Representation Information”.

The Matterhorn RDF Model is designed to allow for the modelling of *Records in Context* but takes a different design approach than ICA EGAD's conceptual model - Records in Context (RiC), instead proposing an alternative way to contextualise records.

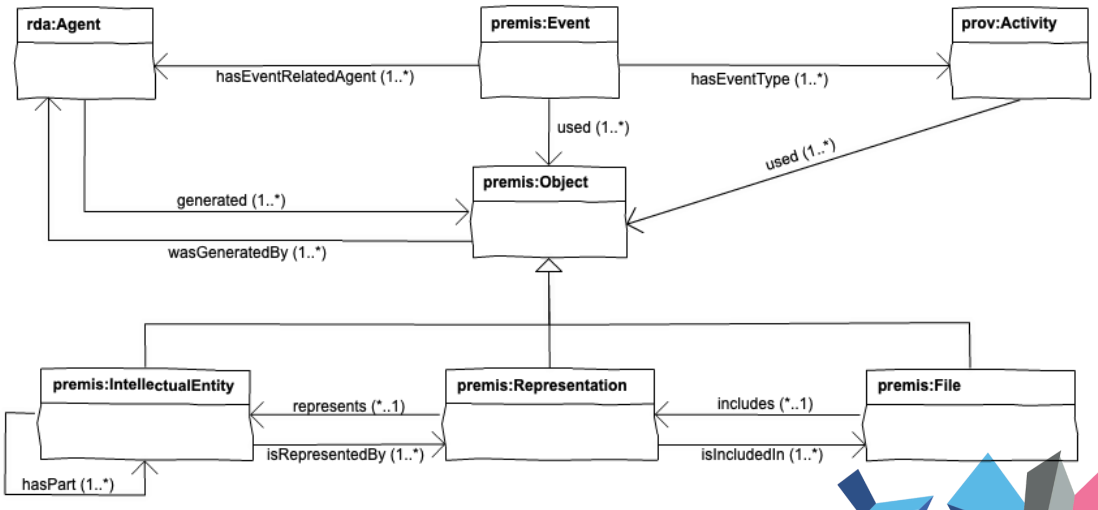
Matterhorn RDF is built to describe archival records as part of a wider world. This is in direct contrast to EGAD’s non-generic approach for RiC-O, of developing their own ontology and vocabulary with (future) gateways to other library and museum standards. Instead, Matterhorn’s RDF Data Model follows W3C’s recommendation to reuse existing standardised vocabularies. Matterhorn RDF composes its model by reusing entities and properties from several vocabularies including: PREMIS3 ontology, RDA-Ontology, W3C PROV-O, EBUCore, and Dublin Core Terms.

The Matterhorn RDF Data Model paper thus describes the relationship between RiC and Matterhorn RDF:

> *“Matterhorn RDF is not to be seen as an alternative to RiC, but rather seeks to elaborate the RiC concept model in a future version, taking into account, however, different design considerations to those which EGAD currently implements.”*

The core entities of the Matterhorn RDF Data Model are: Intellectual Entity, Representation, and File, all of which are taken from PREMIS3. For modelling provenance and contextualisation, the entities: Agent (from RDA), Event (from PREMIS3), and Activity (from W3C PROV), are reused by Matterhorn RDF. For access constraints it reuses the entities Rights and Rights Basis from PREMIS3.

Fig 16 - Matterhorn RDF Model Entities

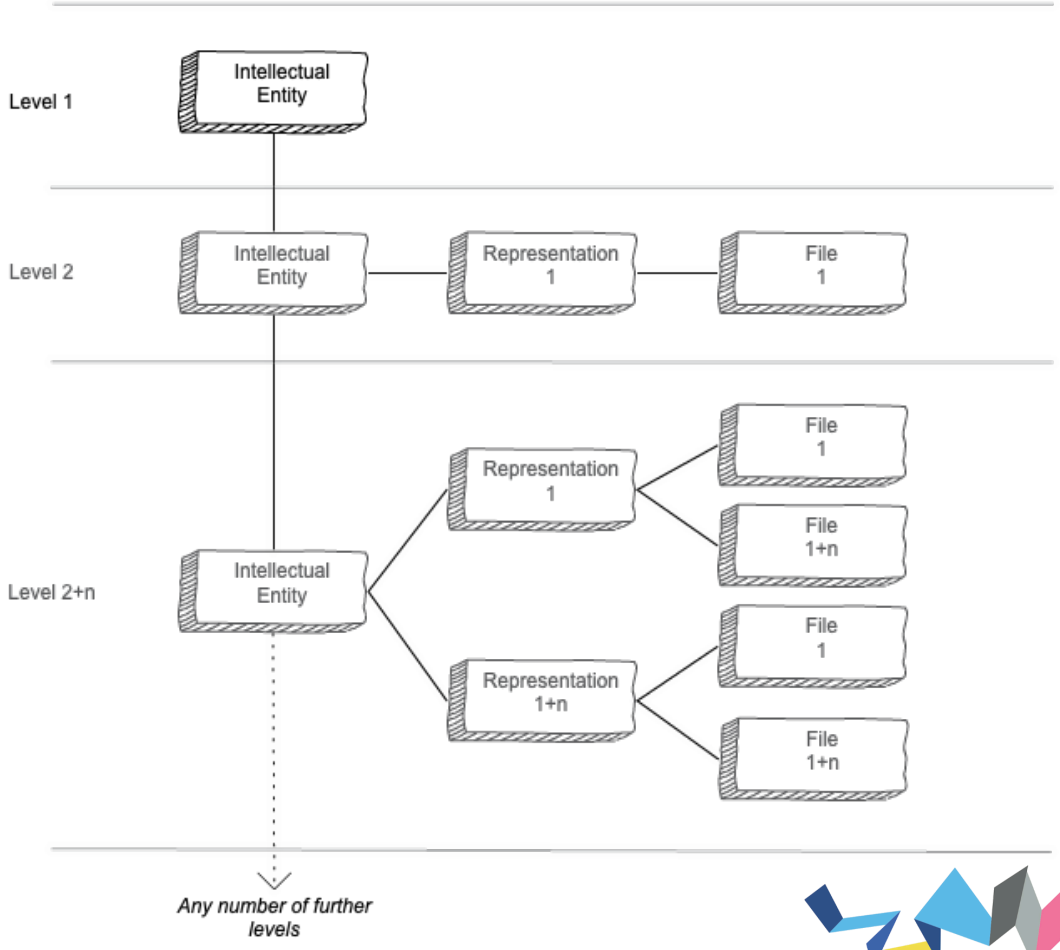


Somewhat like the FRBR WEMI concept, Matterhorn RDF separates the concept of a record from its manifestation, albeit more similarly to RiC-CM, the Matterhorn RDF Data Model splits the record into two aspects - its concept (Intellectual Entity), and its manifestation (Representation).

As the Matterhorn RDF Data Model reuses existing ontologies and vocabularies, there is no singular ontology to describe it. Instead the Matterhorn RDF Data Model, provides a series of SHACL (Shape Constraint Language) files, which describe its composite ontological graph. One nice feature that falls out from this is that one can the subsequently use a SHACL validator tool against data expressed in Matterhorn's RDF Data Model to validate adherence of the data to the model. For extensibility, it is also possible to define additional *shapes* to further ensure data quality.

Interestingly there are quite a few similarities between the DRI Catalogue Model and Matterhorn RDF, this is because they both share a common ancestor in the form of OAIS. The DRI Catalogue Model was influenced by aspects of XIP which is based upon OAIS, whilst Matterhorn RDF reuses PREMIS which was built atop OAIS.

Fig 17 - Matterhorn RDF Model - Record Levels



To express TNA catalogue records in the Matterhorn RDF Data Model, the following mappings from our Conceptual Vocabulary to Matterhorn RDF Data Model entities would be established:

Basic Mapping of Conceptual Vocabulary to Matterhorn RDF Data Model

Conceptual Vocabulary Term	Matterhorn RDF Entity
Record	Intellectual Entity
Record Set	Intellectual Entity
Manifestation	Representation
Digital File	File
Activity	Event/Activity
Agent	Agent
Access Channel	Agent and Rule (TBC)
Access Constraints	Rights Basis and Rights Status

An example of the paper record WO 1/921/233 held by TNA expressed in Matterhorn RDF (Turtle) is available in Appendix 1.5.

The Matterhorn RDF Data Model seems well suited to a pan-archival catalogue model, it appears to be able to handle the expression of all our record Test Cases and meets all of our Functional Requirements in the most part. There are however two areas of uncertainty around how suitable the Matterhorn RDF Data Model will be:

1. the extent of descriptive properties for both physical and digital records; Without purchasing access to various resources around RDA, our ability to examine the full descriptive facilities of RDA are restricted (challenges FR2 and FR4)
2. how flexible the model is for our requirements around Access Channels and Access Constraints; it certainly supports our existing records access policies, but TNA's thinking on Gradated Access is still in its infancy, and therefore it is hard to know if it will meet future requirements (challenges RTC23 and FR8)

The risk of these uncertainties can however likely be mitigated through the open-world assumption taken by the Matterhorn RDF Data Model, which allows it to freely extended.

One technical issue against the use of Matterhorn RDF is through its use of RDA in implementation. Unfortunately the canonical URI used by RDA are not suitable for consumption by humans, which makes reading and authoring RDF which uses RDA very difficult without higher-level tools. RDA does also provide a “readable" lexical form of URIs, but the use and mapping of these does not seem to be without issue. For example, to describe the "*system of organization*” for a record with RDA, one would use the canonical property: http://rdaregistry.info/Elements/u/P60348, whereas the equivalent lexical property would be: http://www.rdaregistry.info/Elements/u/systemOfOrganization.en. For further details see: https://github.com/RDARegistry/RDA-Vocabularies/issues/38

9. PAN-ARCHIVAL DATA MODEL SELECTION

From our analysis of the existing applicable standards there are a few common themes between the standards that we have observed:

1. whilst the older standards used a hierarchy model, the more modern standards have all adopted a graph model
2. the graph based models all provide facilities for describing the provenance of records
3. the graph based models allow both hierarchical arrangement of records and more complex ad-hoc arrangements
4. almost all of the models separate the concept of a record from its realisation, many also allow multiple realisations
5. the archival centric standards have all been influenced by ISAD(G), whilst the library centric standards have all been influenced by MARC

It is important to also remember that TNA has some philosophical values that should be considered when selecting a new catalogue model:

1. consider TNA's records within the larger world, i.e. their use/reuse/relatability outside of the archive
2. open-world assumption, extensibility is important, TNA does not know the variety of records that it will encounter in future or how it might enhance the description of existing records
3. follow established standards and best practice where possible
4. prefer pragmatic decisions instead of adhering rigidly to standards
5. prefer international standards from recognised standards groups over industry specific or de-facto standards

Ultimately none of the standards that we analysed were a perfect fit for a new pan-archival catalogue model for TNA, and although some are more suited than others, each involve various trade-offs. Whilst of course we could develop a new vocabulary/ontology for TNA, we believe the best approach would be to reuse and/or collaborate on an existing standard; the W3C states "It is best practice to use or extend an existing vocabulary before creating a new vocabulary".

From the standards that we analysed, as measured against our Functional Requirements and Record Test Cases, the three most applicable standards (in ascending order of preference) were:

1. BIBFRAME Lite+Archive
2. Records in Context
3. Matterhorn RDF

All three are graph based models, and interestingly both BIBFRAME Lite+Archive and Matterhorn RDF make use of elements from RDA, but each take a different approach. RDA is more oriented towards libraries than archives, and this shows through in BIBFRAME whilst Matterhorn RDF's model (which mixes in PREMIS) is arguably more suitable for archives.

Whilst BIBFRAME could form the major part of a new pan-archival catalogue model, we would need to augment it with, support for detailed technical description of digital records, and facilities for controlled access. However, we felt that the extent of the required augmentation and its libraries bibliographic influence did not make it the strongest contender.

Likewise, RiC could certainly be used for a new pan-archival catalogue model, and might indeed prove to be the best choice for adoption in the long-term, however at present it feels immature and unfinished. It is unclear how digital files should be modelled in RiC, and there is also a lack of data elements for detailed technical description of digital records; perhaps this is due to its ISAD(G) heritage! We perceive this lack of support for digital records to be a major shortcoming at the moment. Like BIBFRAME we would need to augment RiC with better support for digital records, and facilities for controlled access. Another concern is that RiC creates its own vocabulary and ontology, which is very much inwardly focused towards archives, and at present there is no interoperability with other widely known standards such as Dublin Core or W3C PROV.

Finally, we come to Matterhorn RDF, this approach passed all of our Functional Requirements and Record Test Cases, however it is unclear how much it offers in terms of controlled access. Matterhorn RDF also aligns well with TNA's philosophical values including - reusing existing vocabularies and ontologies, and interoperating with the wider world. However Matterhorn RDF is not a standard so much as an approach for reusing existing

standards, however it is an approach that could be adopted and used today. Matterhorn RDF certainly comes with risk, due to its unstable status, however we believe that it is most suitable as the starting point for a TNA pan-archival catalogue model. As Matterhorn RDF is designed for reuse, TNA are free to extend it if/as needed with additional vocabularies and ontologies, effectively deriving their own model from existing standards.

9.1 Recommendations

Our recommendations for TNA are a as follows:

1. create a single pan-archival catalogue for the organisation
2. for the proof-of-concept, adopt Matterhorn RDF as the pan-archival catalogue model
3. consider Matterhorn RDF to be a starting point, TNA should clearly document their use of it
4. document and provide SHACL shapes for any extensions of Matterhorn RDF
5. after the proof-of-concept, should there be a full project for a pan-archival catalogue, reconsider the suitability of RiC vs Matterhorn RDF at that time
6. as RiC is still in development and in future it may prove desirable to adopt it, TNA should immediately become a member of ICA's EGAD, providing both feedback and technical expertise to help shape the standard

GLOSSARY

Glossary Table

Term	Description
A2A	Access 2 Archives; predecessor of MYC
AACR	Anglo-American Cataloguing Rules
AP	Application Profile
ARK	Advice and Records Knowledge; historical name for the CEE department within TNA
BIA	Business Information Architecture
CCR	Classic Catalogue Reference; as defined by and used with TNA-CS13
CIDOC-CRM	International Committee for Documentation - Conceptual Reference Model
CEE	Collections Expertise and Engagement; department within TNA. Previously known as ARK
CSV	Comma Separated Value; a text file format for storing tabular data
CRUD	Create, Replace, Update, and Delete; actions performed upon a database
DACS	Describing Archives a Content Standard; US implementation of a content standard for ISAD(G) and ISAAR
DCAT	Data Catalog Vocabulary; W3C standard for facilitating interoperability between data catalogs published on the Web
DORIS	Document Ordering and Reader Information Service; System used by TNA staff to manage physical access requests for paper records
DPI	Dots Per Inch; is a measure of spatial printing, video or image scanner dot density, often denoting a quality factor
DRI	Digital Records Infrastructure; TNA's Digital Archive
DU	Deliverable Unit; DRI Catalogue Term for one or more things which are deliverable digitally
EAC-CPF	Encoded Archival Context - Corporate bodies, Persons and Families; related to EAD
EAD	Encoded Archival Description; standard for describing archival finding aids
EDM	Europeana Data Model
EGAD	Expert Group on Archival Description; group within ICA
ESE	Europeana Semantic Elements
ePRO	Electronic PRO; The database used by DocsOnline
FOAF	Friend of a Friend; an ontology for describing people and their relationships
FOI	Freedom of Information; a statutory legal act
FR	Functional Requirement; a description of a requirement that a system should be able to fulfil

Term	Description
FRAD	Functional Requirements for Authority Data; ILFA standard
FRBR	Functional Requirements for Bibliographic Records; IFLA standard
FRBRoo	FRBR-object oriented; unifies FRBR and CIDOC-CRM
FRSAD	Functional Requirements for Subject Authority Data; ILFA standard
GCR	Generated Catalogue Reference; the catalogue reference scheme employed for Digital Records in DRI
HMC	Historical Manuscripts Commission; holds catalogue information about records and papers in other repositories and private hands, and the MDR
IA	Information Asset; as defined in the BIA standard
ICA	International Council on Archives; standards body for archival standards
IFLA	International Federation of Library Associations and Institutions
ILDB	Inventory Lists Database; holds the catalogue records managed by PROCat
ISAAR	International Standard Archival Authority Record; ICA standard
ISAD(G)	General International Standard Archival Description; ICA standard
ISDF	International Standard for Describing Functions; ICA standard
ISDIAH	International Standard for Describing Institutions with Archival Holdings; ICA standard
JPEG	Joint Photographic Experts Group; typically a lower-resolution lossy image file format
LoC	Library of Congress (USA)
LRM	Library Reference Model; IFLA standard
MARC	Machine-readable cataloging; standards from the LoC
METS	Metadata Encoding and Transmission Standard
MDR	Manorial Documents Register
MODS	Metadata Object Description Schema; standard from the LoC
MYC	Manage Your Collections; catalogue of records held by other archives, usually non-authoritative for an archive, structure based on ISAD(G
OAI-ORE	Open Archives Initiative - Object Reuse and Exchange
OAIS	Open Archival Information System; ISO standard 14721:2012
OPM	Open Provenance Model; standard for describing provenance metadata, superseded by W3C PROV
OWL	Web Ontology Language; W3C standard for authoring ontologies, used with RDF
PIAAF	Interoperability Pilot for the French Archival Authorities

Term	Description
PET	Perl-based Extraction & Translation; tool for converting CSV files provided by the Transfer Team, CEE, volunteers and others into EADv1 for loading into PROCat
PREMIS	Preservation Metadata: Implementation Strategies; LoC standard for preservation
Project Omega	The proof-of-concept project to replace PROCat and potentially other TNA catalogues.
PRO	Public Record Office; predecessor to TNA
PROV	Provenance; W3C family of standards for provenance information
PROV-O	Provenance Ontology; part of the PROV family of standards
PROCat	PRO Catalogue; TNA's current paper records catalogue management applications
RDA	Resource Description and Access; standard for descriptive cataloguing
RDF	Resource Description Framework; W3C standard for describing resources on the web
RDFS	RDF Schema; W3C standard for data modelling ontology for RDF
RiC	Records in Context; ICA standard positioned to supercede ISAD(G), ISAAR (CPF), and ISDF
RiC-CM	RiC Conceptual Model
RiC-O	RiC Ontology; ICA's RDF implementation schema and vocabulary for RiC-CM
RTC	Record Test Case; a description of an arrangement and metadata of records that a system should be able to model
SAR	System for Access Regulation; holds and manages Closure Status information of records at TNA and applications to the Advisory Council
SID	Super Integrated System; the database used by the Collection Care department
SFA	Supplementary Finding Aid; paper based description of catalogued records
SHACL	Shapes Constraints Language; W3C standard for describing and validating RDF graphs
SKOS	Simple Knowledge Organisation System; W3C standard for representing conceptual vocabularies
SDB	Safety Deposit Box; the central component of the DRI Preservation System
SoAA	Society of American Archivists
TIFF	Tagged Image File Format; raster graphics image file format
TNA	The National Archives; successor of PRO
TNA-CS13	The National Archives, Cataloguing Standards - June 2013
UKGWA	UK Government Web Archive; web archive service operated by TNA
UI	User Interface
UUID	Universally Unique Identifier

Term	Description
WARC	Web ARChive; an archival file format standard for web archiving, as used by UKGWA
WEMI	Work, Expression, Manifestation, and Item; as defined by the FRBR conceptual model
W3C	World Wide Web Consortium; an international standards organisation
XIP	XML for Information Package; the Tessella SDB/Preservica XML Schema
XML	Extensible Markup Language; W3C standard for document markup
XMP	Extensible Metadata Platform; ISO standard for standardised and custom metadata for digital documents and data sets

APPENDIX 1. MODEL SYNTAX EXAMPLES

This appendix contains the expression of a TNA-CS13 Item catalogued at all 7 levels, in various model syntaxes.

A1.1 EADv3 XML Syntax Example

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-model href="https://www.loc.gov/ead/ead3.rng" type="application/xml" schematypens="http://relaxng.org/
ns/structure/1.0"?>
<ead xmlns="http://ead3.archivists.org/schema/">
    <control>
        <recordid>TNA-CS13/WO</recordid>
        <filedesc>
            <titlestmt>
                <titleproper>Records created or inherited by the War Office, Armed Forces, Judge
                    Advocate General, and related bodies</titleproper>
            </titlestmt>
        </filedesc>
        <maintenancestatus value="new"/>
        <maintenanceagency countrycode="UK">
            <agencyname>The National Archives</agencyname>
        </maintenanceagency>
        <maintenancehistory>
            <maintenanceevent>
                <eventtype value="created"/>
                <eventdatetime standarddatetime="1795-10-20">20 October 1795</eventdatetime>
                <agenttype value="human"/>
                <agent>Staff of The National Archives</agent>
            </maintenanceevent>
        </maintenancehistory>
    </control>
    <archdesc level="fonds">
        <did>
            <repository>
                <corpname>
                    <part>The National Archives</part>
                </corpname>
                <address>
                    <addressline>Bessant Drive</addressline>
                    <addressline>Kew</addressline>
                    <addressline>Surrey</addressline>
                    <addressline>TW9 4DU</addressline>
                </address>
            </repository>
            <unittitle>Records created or inherited by the War Office, Armed Forces, Judge Advocate
                General, and related bodies</unittitle>
            <unitid>WO</unitid>
            <langmaterial>
                <language langcode="en">English</language>
            </langmaterial>
            <physdescstructured physdescstructuredtype="materialtype" coverage="part">
                <quantity>417</quantity>
                <unittype>boxes</unittype>
            </physdescstructured>
        </did>
        <scopecontent>
            <head>Scope and Content</head>
            <p>Records created or inherited by the War Office, Armed Forces, Judge Advocate General,
```

```
                and related bodies relating to the administration of the armed forces.</p>
            <p>Comprises records of: <list>
                    <item>General War Office records</item>
                    <item>Chief of the (Imperial) General Staff and its directorates</item>
                    <item>Commander-in-Chief, Military Secretary and Army Council</item>
                    <item>Army Medical services</item>
                    <item>Other administrative departments of the War Office</item>
                    <item>Finance departments</item>
                    <item>Central Department of the Permanent Secretary of State</item>
                    <item>Land Branch</item>
                    <item>Board of General Officers</item>
                    <item>Royal Chelsea and Kilmainham Hospitals</item>
                    <item>Constable of the Tower of London</item>
                    <item>Research establishments</item>
                    <item>Armed forces records from commands, headquarters, regiments and
                        corps</item>
                    <item>Auxiliary forces</item>
                    <item>Army in Ireland</item>
                    <item>Armed forces service records</item>
                    <item>Private office papers and private collections</item>
                    <item>Commissions, committees and councils, etc</item>
                    <item>Ordnance Office and its War Office successors</item>
                    <item>Judge Advocate General</item>
                </list></p>
            <p>WO 56, WO 125, WO 200, WO 249 and WO 289 are numbers not used.</p>
            <p>WO 20 and WO 21 were transferred to other series before 1905.</p>
        </scopecontent>
        <dsc>
            <c level="subfonds">
                <did>
                    <unittitle>General records of the War Office and predecessors and
                        successor</unittitle>
                    <unitid>WO_1</unitid>
                    <langmaterial>
                        <language langcode="en">English</language>
                    </langmaterial>
                    <physdescstructured physdescstructuredtype="materialtype" coverage="part">
                        <quantity>28</quantity>
                        <unittype>boxes</unittype>
                    </physdescstructured>
                </did>
                <scopecontent>
                    <head>Scope and Content</head>
                    <p>General records of the War Office and predecessors and successors relating to
                        various aspects of its duties for the administration of the armed
                        forces.</p>
                    <p>Comprises: <list>
                            <item>Correspondence and papers of the Secretary-at-War, <ref
                                    href="http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14212"
                                    >WO 1</ref>, <ref
                                    href="http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14249"
                                    >WO 40</ref> and <ref
                                    href="http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14252"
                                    >WO 43</ref></item>
                            <item>Registered correspondence, <ref
                                    href="http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14241"
                                    >WO 32</ref>, <ref
```

```
                        href="http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14252"
                        >WO 43</ref> and <ref
                        href="http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14348"
                        >WO 141</ref></item>
                    <item>Indexes to correspondence, <ref
                        href="http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14346"
                        >WO 139</ref> and <ref
                        href="http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14601"
                        >WO 900</ref></item>
                    <item>Correspondence and papers relating to the First World War, <ref
                        href="http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14368"
                        >WO 161</ref></item>
                    <item>Entry books of out-letters, <ref
                        href="http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14215"
                        >WO 4</ref> with indexes in <ref
                        href="http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14213"
                        >WO 2</ref></item>
                    <item>Other entry books, <ref
                        href="http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14234"
                        >WO 25</ref>, <ref
                        href="http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14235"
                        >WO 26</ref> and <ref
                        href="http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14239"
                        >WO 30</ref></item>
                    <item>Out-letters of the Secretary of State for War and Secretary of
                        State for War and the Colonies, <ref
                        href="http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14217"
                        >WO 6</ref></item>
                    <item>Various departmental out-letters, <ref
                        href="http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14218"
                        >WO 7</ref></item>
                    <item>Personal files, <ref
                        href="http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14345"
                        >WO 138</ref></item>
                    <item>Civilian personal files, <ref
                        href="http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14539"
                        >WO 335</ref></item>
                    <item>Printed reports and papers including hand books and manuals in
                        <ref
                        href="http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14242"
                        >WO 33</ref>, <ref
                        href="http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14484"
                        >WO 279</ref> and <ref
                        href="http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14492"
                        >WO 287</ref></item>
```

```
                        <item>Maps and plans, <ref
                                href="http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14286"
                                >WO 78</ref></item>
                        <item>War diaries of War Office directorates for the Second World War,
                                <ref
                                href="http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14372"
                                >WO 165</ref></item>
                        <item>Minutes of directors' meetings, 1906 to 1913, <ref
                                href="http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14448"
                                >WO 242</ref></item>
                        <item>Establishment papers, <ref
                                href="http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14233"
                                >WO 24</ref></item>
                        <item>War Office: Military Intelligence section 9 (MI9), Escape and
                            Evasion maps, Second World War, <ref
                                href="http://discovery.nationalarchives.gov.uk/details/r/C16747711"
                                >WO 418</ref></item>
                    </list></p>
                </scopecontent>
                <c level="series">
                    <did>
                        <unittitle>War Office and predecessors: Secretary-at-War, Secretary of State
                            for War, and Commander-in-Chief, In-letters and Miscellaneous
                            Papers</unittitle>
                        <unitid>WO_1_1</unitid>
                        <langmaterial>
                            <language langcode="en">English</language>
                        </langmaterial>
                        <physdescstructured physdescstructuredtype="materialtype" coverage="part">
                            <quantity>1138</quantity>
                            <unittype>boxes</unittype>
                        </physdescstructured>
                    </did>
                    <scopecontent>
                        <head>Scope and Content</head>
                        <p>These are for the most part journals and indexes of letters in the letter
                            books of the Secretary-at-War and of War Office departments. There are
                            also journals with an index of the Board of General Officers, 1808 to
                            1816; an index of the Agent General of Militia, 1803 to 1804; and an
                            entry book of memoranda and precedents relating to the War Office
                            establishment in 1782. Many of the letter books to which these indexes
                            refer cannot be identified and probably do not exist.</p>
                        <p>The series includes material which would otherwise be found amongst the
                            Records of the Board of General Officers and records relating to
                            soldiers' clothing (<ref>WO</ref>).</p>
                    </scopecontent>
                    <arrangement>
                        <p>Papers of the Secretary-at-War and the Secretary of State for War are
                            arranged under the various military stations or public departments from
                            which they were sent, or under the subjects with which they deal, or
                            alphabetically by the name of the writer.</p>
                        <p>The papers are also grouped into two larger and two smaller subseries: </p>
                        <p>I: Records of the old War Office (Secretary-at-War), arranged into three
                            groups: <list>
                                <item>1. The books of original dispatches from various commanders, </item>
                                <item>2. Correspondence received from other government departments, </item>
                                <item>3. Correspondence from others, arranged roughly by the
                                    alphabetical order of the surnames of the senders, and including
```

```
                    certain correspondence which might be expected to be found in
                    the volumes of the first two groups. This has been divided into
                    two sequences: the first consisting of foolscap volumes, the
                    second of smaller volumes. </item>
                </list></p>
            <p>II: Records of the War Department (Secretary of State for War), arranged
                into four groups: </p>
            <p><list>
                    <item>1. Despatches, correspondence and papers from the French Wars
                        period (1794-1816). Some of these books start earlier than July
                        1794, before the separate War Department was created, </item>
                    <item>2. A few volumes covering the period 1817-1831, when the major
                        part of the work of the Colonial and War Department was
                        colonial, </item>
                    <item>3. Military in-letters from the 1830s to the 1850s, arranged
                        by colonies, </item>
                    <item>4. Crimean War correspondence and despatches. </item>
                </list></p>
            <p>III: Loose bundles of despatches addressed to the Commander-in-Chief. </p>
            <p>IV: Miscellaneous volumes. </p>
        </arrangement>
        <c level="subseries">
            <did>
                <unittitle>II. WAR DEPARTMENT IN-LETTERS AND PAPERS</unittitle>
                <unitid>WO_1_1_2</unitid>
            </did>
            <c level="otherlevel" otherlevel="subsubseries">
                <did>
                    <unittitle>1. Of the French Wars period</unittitle>
                    <unitid>WO_1_1_2_1</unitid>
                </did>
                <c level="file">
                    <did>
                        <unitid>WO_1_1_2_1_921</unitid>
                    </did>
                    <scopecontent>
                        <p>j. Intelligence: Prince de Bouillon: Correspondence.
                            (Described at item level) </p>
                    </scopecontent>
                    <accessrestrict>
                        <p>Open Document, Open Description. Normal closure before FOI
                            Act: 30</p>
                    </accessrestrict>
                    <c level="item">
                        <did>
                            <unitid>WO_1_1_2_1_921_233</unitid>
                            <didnote>Enclosed in ff 229-232 French.</didnote>
                        </did>
                        <scopecontent>
                            <p>Folio(s) 233-236: Bertin to Bouillon. Discussing arms
                                received by Vieuville and 'Rodolphe' and 19 Oct.
                                landings; traitors and informants; weakness of St. Malo
                                garrison and desirability of retaining control of
                                Chausey.</p>
                            <p>Date and Place: 1795 20 Oct. Jersey.</p>
                        </scopecontent>
                        <accessrestrict>
                            <p>Open Document, Open Description. Normal closure before
                                FOI Act: 30</p>
                        </accessrestrict>
                    </c>
                </c>
```

```
                        </c>
                    </c>
                </c>
            </c>
        </dsc>
    </archdesc>
</ead>
```

A1.2 DCAT v2 RDF (Turtle) Syntax Example

```
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix cat: <http://nationalarchives.gov.uk/catalogue> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix prov: <http://www.w3.org/ns/prov#> .
@prefix schema: <http://schema.org/> .
@prefix vcard: <http://www.w3.org/2006/vcard/ns#> .
@prefix wd: <http://www.wikidata.org/entity/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@base <https://www.w3.org/ns/dcat.rdf> .

cat:ccr rdf:type owl:DatatypeProperty ;
  rdfs:subPropertyOf dct:identifier ;
  rdfs:domain [
    a owl:Class
  ] ;
  rdfs:range rdfs:Literal ;
  rdfs:comment "Classic Catalogue Reference" ;
  rdfs:label "classic catalogue reference "@en
  .
dcat:Dataset rdfs:subClassOf
   [ a owl:Restriction ;
     owl:onProperty cat:ccr ;
     owl:maxCardinality 1;
   ] .
dcat:Catalog rdfs:subClassOf
   [ a owl:Restriction ;
     owl:onProperty cat:ccr ;
     owl:maxCardinality 1;
   ] .

cat:extent rdf:type owl:DatatypeProperty ;
  rdfs:subPropertyOf dct:extent ;
  rdfs:domain [
    a owl:Class ;
    owl:unionOf (
     dcat:Catalog
     dcat:Dataset
   )
  ] ;
  rdfs:range rdfs:Literal ;
  rdfs:comment "Physical or Logical extent of the records" ;
  rdfs:label "physical or logical extent "@en
  .

cat:arrangement rdf:type owl:DatatypeProperty ;
  rdfs:subPropertyOf dct:description ;
```

```
  rdfs:domain [
    a owl:Class ;
    owl:unionOf (
     dcat:Catalog
     dcat:Dataset
   )
  ] ;
  rdfs:range rdfs:XMLLiteral ;
  rdfs:comment "Physical or Logical arrangement of the records" ;
  rdfs:label "description of arrangement "@en
  .

cat:note rdf:type owl:DatatypeProperty ;
  rdfs:subPropertyOf dct:description ;
  rdfs:domain [
    a owl:Class ;
    owl:unionOf (
     dcat:Catalog
     dcat:CatalogRecord
     dcat:Dataset
     dcat:Distribution
   )
  ] ;
  rdfs:range rdfs:Literal ;
  rdfs:comment "General Note" ;
  rdfs:label "general note "@en
  .

cat:TNA rdf:type prov:Organization ;
        dct:identifier "http://nationalarchives.gov.uk"^^xsd:anyURI ;
        dct:description "The National Archives of the UK" ;
        foaf:name "The National Archives" ;
        vcard:hasAddress cat:TNA-ADDR ;
        schema:about wd:Q392703
        .

cat:TNA-ADDR rdf:type vcard:Address ;
        vcard:street-address "Bessant Drive" ;
        vcard:locality "Kew" ;
        vcard:region "Surrey" ;
        vcard:postal-code "TW9 4DU" ;
        vcard:country-name "UK"
        .

cat:departmentWO rdf:type dcat:Catalog ;
        dct:title "Records created or inherited by the War Office, Armed Forces, Judge Advocate General, and
related bodies" ;
        dct:description "<scopecontent><head>Scope and Content</head><p>Records created or inherited by the
War Office, Armed Forces, Judge Advocate General, and related bodies relating to the administration of the
armed forces.</p> <p>Comprises records of: <list><item>General War Office records</item><item>Chief of the
(Imperial) General Staff and its directorates</item><item>Commander-in-Chief, Military Secretary and Army
Council</item><item>Army Medical services</item><item>Other administrative departments of the War Office</
item><item>Finance departments</item><item>Central Department of the Permanent Secretary of State</
item><item>Land Branch</item><item>Board of General Officers</item><item>Royal Chelsea and Kilmainham
Hospitals</item><item>Constable of the Tower of London</item><item>Research establishments</item><item>Armed
forces records from commands, headquarters, regiments and corps</item><item>Auxiliary forces</item><item>Army
in Ireland</item><item>Armed forces service records</item><item>Private office papers and private
collections</item><item>Commissions, committees and councils, etc</item><item>Ordnance Office and its War
Office successors</item><item>Judge Advocate General</item></list></p> <p>WO 56, WO 125, WO 200, WO 249 and WO
289 are numbers not used.</p> <p>WO 20 and WO 21 were transferred to other series before 1905.</p></
scopecontent>" ;
        cat:ccr "WO" ;
```

```
        dct:identifier "http://catalogue.nationalarchives.gov.uk/WO"^^xsd:anyURI ;
        dct:type "fonds" ;
        dct:language "eng" ;
        cat:extent "417" ;
        dct:accrualPolicy "open" ;
        dct:publisher cat:TNA
        .

cat:divisionWO_1 rdf:type dcat:Catalog ;
        dct:title "General records of the War Office and predecessors and successor" ;
        dct:description "<scopecontent><head>Scope and Content</head><p>General records of the War Office and
predecessors and successors relating to various aspects of its duties for the administration of the armed
forces.</p> <p>Comprises: <list><item>Correspondence and papers of the Secretary-at-War, <extref
href=&#34http://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14212&#34>WO 1</extref>, <extref
href=&#34http://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14249&#34>WO 40</extref> and <extref
href=&#34http://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14252&#34>WO 43</extref></
item><item>Registered correspondence, <extref href=&#34http://discovery.nationalarchives.gov.uk/SearchUI/
details?Uri=C14241&#34>WO 32</extref>, <extref href=&#34http://discovery.nationalarchives.gov.uk/SearchUI/
details?Uri=C14252&#34>WO 43</extref> and <extref href=&#34http://discovery.nationalarchives.gov.uk/SearchUI/
details?Uri=C14348&#34>WO 141</extref></item><item>Indexes to correspondence, <extref href=&#34http://
discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14346&#34>WO 139</extref> and <extref href=&#34http://
discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14601&#34>WO 900</extref></item><item>Correspondence
and papers relating to the First World War, <extref href=&#34http://discovery.nationalarchives.gov.uk/
SearchUI/details?Uri=C14368&#34>WO 161</extref></item><item>Entry books of out-letters, <extref
href=&#34http://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14215&#34>WO 4</extref> with indexes
in <extref href=&#34http://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14213&#34>WO 2</extref></
item><item>Other entry books, <extref href=&#34http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14234&#34>WO 25</extref>, <extref href=&#34http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14235&#34>WO 26</extref> and <extref href=&#34http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14239&#34>WO 30</extref></item><item>Out-letters of the Secretary of State for War and Secretary of State
for War and the Colonies, <extref href=&#34http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14217&#34>WO 6</extref></item><item>Various departmental out-letters, <extref href=&#34http://
discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14218&#34>WO 7</extref></item><item>Personal files,
<extref href=&#34http://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14345&#34>WO 138</extref></
item><item>Civilian personal files, <extref href=&#34http://discovery.nationalarchives.gov.uk/SearchUI/
details?Uri=C14539&#34>WO 335</extref></item><item>Printed reports and papers including hand books and manuals
in <extref href=&#34http://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14242&#34>WO 33</extref>,
<extref href=&#34http://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14484&#34>WO 279</extref> and
<extref href=&#34http://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14492&#34>WO 287</extref></
item><item>Maps and plans, <extref href=&#34http://discovery.nationalarchives.gov.uk/SearchUI/details?
Uri=C14286&#34>WO 78</extref></item><item>War diaries of War Office directorates for the Second World War,
<extref href=&#34http://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14372&#34>WO 165</extref></
item><item>Minutes of directors' meetings, 1906 to 1913, <extref href=&#34http://
discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14448&#34>WO 242</extref></item><item>Establishment
papers, <extref href=&#34http://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14233&#34>WO 24</
extref></item><item>War Office: Military Intelligence section 9 (MI9), Escape and Evasion maps, Second World
War, <extref href=&#34http://discovery.nationalarchives.gov.uk/details/r/C16747711&#34>WO 418</extref></
item></list></p></scopecontent>"^^rdf:XMLLiteral ;
        cat:ccr "WO_1" ;
        dct:identifier "http://catalogue.nationalarchives.gov.uk/WO_1"^^xsd:anyURI ;
        dct:type "sub-fonds" ;
        dct:language "eng" ;
        cat:extent "28" ;
        dct:accrualPolicy "open" ;
        dcat:catalog cat:departmentWO
        .

cat:seriesWO_1_1 rdf:type dcat:Catalog ;
        dct:title "War Office and predecessors: Secretary-at-War, Secretary of State for War, and Commander-
in-Chief, In-letters and Miscellaneous Papers" ;
        dct:description "<scopecontent><head>Scope and Content</head><p>These are for the most part journals
and indexes of letters in the letter books of the Secretary-at-War and of War Office departments. There are
```

also journals with an index of the Board of General Officers, 1808 to 1816; an index of the Agent General of Militia, 1803 to 1804; and an entry book of memoranda and precedents relating to the War Office establishment in 1782. Many of the letter books to which these indexes refer cannot be identified and probably do not exist. The series includes material which would otherwise be found amongst the Records of the Board of General Officers and records relating to soldiers' clothing (<ref>WO</ref>). </scopecontent>"^^rdf:XMLLiteral ;
        cat:ccr "WO_1_1" ;
        dct:identifier "http://catalogue.nationalarchives.gov.uk/WO_1_1"^^xsd:anyURI ;
        dct:type "series" ;
        dct:language "eng" ;
        cat:extent "1138" ;
        cat:arrangement "<arrangement> Papers of the Secretary-at-War and the Secretary of State for War are arranged under the various military stations or public departments from which they were sent, or under the subjects with which they deal, or alphabetically by the name of the writer. The papers are also grouped into two larger and two smaller subseries: I: Records of the old War Office (Secretary-at-War), arranged into three groups: <list><item>1. The books of original dispatches from various commanders, </item><item>2. Correspondence received from other government departments, </item><item>3. Correspondence from others, arranged roughly by the alphabetical order of the surnames of the senders, and including certain correspondence which might be expected to be found in the volumes of the first two groups. This has been divided into two sequences: the first consisting of foolscap volumes, the second of smaller volumes. </item></list> II: Records of the War Department (Secretary of State for War), arranged into four groups: <list><item>1. Despatches, correspondence and papers from the French Wars period (1794-1816). Some of these books start earlier than July 1794, before the separate War Department was created, </item><item>2. A few volumes covering the period 1817-1831, when the major part of the work of the Colonial and War Department was colonial, </item><item>3. Military in-letters from the 1830s to the 1850s, arranged by colonies, </item><item>4. Crimean War correspondence and despatches. </item></list> III: Loose bundles of despatches addressed to the Commander-in-Chief. IV: Miscellaneous volumes. </arrangement>"^^rdf:XMLLiteral ;
        dct:accrualPolicy "open" ;
        dcat:catalog cat:divisionWO_1
        .

cat:subseriesWO_1_1_2 rdf:type dcat:Catalog ;
        dct:title "II. WAR DEPARTMENT IN-LETTERS AND PAPERS" ;
        cat:ccr "WO_1_1_2" ;
        dct:identifier "http://catalogue.nationalarchives.gov.uk/WO_1_1_2"^^xsd:anyURI ;
        dct:type "sub-series" ;
        dct:accrualPolicy "open" ;
        dcat:catalog cat:seriesWO_1_1
        .

cat:subsubseriesWO_1_1_2_1 rdf:type dcat:Catalog ;
        dct:title "1. Of the French Wars period" ;
        cat:ccr "WO_1_1_2_1" ;
        dct:identifier "http://catalogue.nationalarchives.gov.uk/WO_1_1_2_1"^^xsd:anyURI ;
        dct:type "sub-sub-series" ;
        dct:accrualPolicy "open" ;
        dcat:catalog cat:subseriesWO_1_1_2
        .

cat:pieceWO_1_1_2_1_921 rdf:type dcat:Catalog ;
        dct:description "<scopecontent> j. Intelligence: Prince de Bouillon: Correspondence. (Described at item level) </scopecontent>" ;
        cat:ccr "WO_1_1_2_1_921" ;
        dct:identifier "http://catalogue.nationalarchives.gov.uk/WO_1_1_2_1_921"^^xsd:anyURI ;
        dct:type "file" ;
        dct:accrualPolicy "open" ;
        dct:accessRights "<closure><status>O</status><type>N</type><code>30</code></closure>" ;
        dcat:catalog cat:subsubseriesWO_1_1_2_1 ;
        dcat:record cat:crWO_1_1_2_1_921_233
        .

cat:crWO_1_1_2_1_921_233 rdf:type dcat:CatalogRecord ;

```
        foaf:primaryTopic cat:itemWO_1_1_2_1_921_233
        .

cat:itemWO_1_1_2_1_921_233 rdf:type dcat:Dataset ;
        dct:description "<scopecontent><p>Folio(s) 233-236: Bertin to Bouillon. Discussing arms received by
Vieuville and 'Rodolphe' and 19 Oct. landings; traitors and informants; weakness of St. Malo garrison and
desirability of retaining control of Chausey.</p><p>Date and Place: 1795 20 Oct. Jersey.</p></scopecontent>" ;
        cat:ccr "WO_1_1_2_1_921_233" ;
        dct:identifier "http://catalogue.nationalarchives.gov.uk/WO_1_1_2_1_921_233"^^xsd:anyURI ;
        dct:type "item" ;
        dct:accrualPolicy "closed" ;
        cat:note "Enclosed in ff 229-232 French." ;
        dct:accessRights "<closure><status>O</status><type>N</type><code>30</code></closure>"
        .
```

A1.3 EDM RDF (Turtle) Syntax Example

NOTE: WebResource entities have been intentionally omitted as the example describes physical records. Potentially EDM's PhysicalObject entity could be used, but its status is uncertain.

```
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix cat: <http://nationalarchives.gov.uk/catalogue> .
@prefix edm: <http://www.europeana.eu/schemas/edm/> .
@prefix ore: <http://www.openarchives.org/ore/terms/> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix cldtype: <http://example.org/cld/type#> .
@prefix accpol: <http://purl.org/cld/accpol/> .
@prefix wd: <http://www.wikidata.org/entity/> .
@prefix vcard: <http://www.w3.org/2006/vcard/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .

@base <https://raw.githubusercontent.com/europeana/corelib/master/corelib-edm-definitions/src/main/resources/
eu/rdf/edm.ttl> .

cat:arrangement rdf:type owl:DatatypeProperty ;
  rdfs:subPropertyOf dcterms:description ;
  rdfs:domain [
    a owl:Class ;
    owl:unionOf (
     edm:ProvidedCHO
   )
  ] ;
  rdfs:range rdfs:XMLLiteral ;
  rdfs:comment "Physical or Logical arrangement of the records" ;
  rdfs:label "description of arrangement "@en
  .

cat:note rdf:type owl:DatatypeProperty ;
  rdfs:subPropertyOf dcterms:description ;
  rdfs:domain [
    a owl:Class ;
    owl:unionOf (
     edm:ProvidedCHO
   )
```

```
  ] ;
  rdfs:range rdfs:Literal ;
  rdfs:comment "General Note" ;
  rdfs:label "general note "@en
  .

cat:TNA rdf:type edm:Agent ;
        foaf:name "The National Archives" ;
        foaf:homepage "https://www.nationalarchives.gov.uk/" ;
        dcterms:relation wd:Q392703 ;
        vcard:street-address "Bessant Drive" ;
        vcard:locality "Kew" ;
        vcard:region "Richmond" ;
        vcard:postal-code "TW9 4DU" ;
        vcard:country-name "UK" .

cat:departmentWO rdf:type edm:ProvidedCHO ;
    dc:publisher cat:TNA ;
        dc:title "Records created or inherited by the War Office, Armed Forces, Judge Advocate General, and
related bodies" ;
        dc:description "<scopecontent><head>Scope and Content</head><p>Records created or inherited by the War
Office, Armed Forces, Judge Advocate General, and related bodies relating to the administration of the armed
forces.</p> <p>Comprises records of: <list><item>General War Office records</item><item>Chief of the
(Imperial) General Staff and its directorates</item><item>Commander-in-Chief, Military Secretary and Army
Council</item><item>Army Medical services</item><item>Other administrative departments of the War Office</
item><item>Finance departments</item><item>Central Department of the Permanent Secretary of State</
item><item>Land Branch</item><item>Board of General Officers</item><item>Royal Chelsea and Kilmainham
Hospitals</item><item>Constable of the Tower of London</item><item>Research establishments</item><item>Armed
forces records from commands, headquarters, regiments and corps</item><item>Auxiliary forces</item><item>Army
in Ireland</item><item>Armed forces service records</item><item>Private office papers and private
collections</item><item>Commissions, committees and councils, etc</item><item>Ordnance Office and its War
Office successors</item><item>Judge Advocate General</item></list></p> <p>WO 56, WO 125, WO 200, WO 249 and WO
289 are numbers not used.</p> <p>WO 20 and WO 21 were transferred to other series before 1905.</p></
scopecontent>" ;
        dc:identifier "WO" ;
        dc:type "department" ;
        dc:type cldtype:CollectionCollection ;
        edm:type "TEXT" ;
        dc:language "eng" ;
        dcterms:extent "417" ;
        dcterms:accrualPolicy accpol:passive .

cat:a0_departmentWO rdf:type ore:Aggregation ;
    edm:aggregatedCHO cat:departmentWO ;
    edm:dataProvider "The National Archives (UK)" ;
    edm:provider "The National Archives (UK)" ;
    edm:rights "http://rightsstatements.org/vocab/InC/1.0/" ;
    edm:isShownBy "https://discovery.nationalarchives.gov.uk/details/r/X1" .

cat:divisionWO_1 rdf:type edm:ProvidedCHO ;
    dc:publisher cat:TNA ;
        dc:title "General records of the War Office and predecessors and successor" ;
        dc:description "<scopecontent><head>Scope and Content</head><p>General records of the War Office and
predecessors and successors relating to various aspects of its duties for the administration of the armed
forces.</p> <p>Comprises: <list><item>Correspondence and papers of the Secretary-at-War, <extref
href=&#34;ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14212&#34;WO 1</extref>, <extref
href=&#34;ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14249&#34;WO 40</extref> and <extref
href=&#34;ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14252&#34;WO 43</extref></
item><item>Registered correspondence, <extref href=&#34;ttp://discovery.nationalarchives.gov.uk/SearchUI/
details?Uri=C14241&#34;WO 32</extref>, <extref href=&#34;ttp://discovery.nationalarchives.gov.uk/SearchUI/
details?Uri=C14252&#34;WO 43</extref> and <extref href=&#34;ttp://discovery.nationalarchives.gov.uk/SearchUI/
details?Uri=C14348&#34;WO 141</extref></item><item>Indexes to correspondence, <extref href=&#34;ttp://
```

discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14346"WO 139</extref> and <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14601"WO 900</extref></item><item>Correspondence and papers relating to the First World War, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14368"WO 161</extref></item><item>Entry books of out-letters, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14215"WO 4</extref> with indexes in <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14213"WO 2</extref></item><item>Other entry books, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14234"WO 25</extref>, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14235"WO 26</extref> and <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14239"WO 30</extref></item><item>Out-letters of the Secretary of State for War and Secretary of State for War and the Colonies, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14217"WO 6</extref></item><item>Various departmental out-letters, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14218"WO 7</extref></item><item>Personal files, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14345"WO 138</extref></item><item>Civilian personal files, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14539"WO 335</extref></item><item>Printed reports and papers including hand books and manuals in <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14242"WO 33</extref>, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14484"WO 279</extref> and <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14492"WO 287</extref></item><item>Maps and plans, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14286"WO 78</extref></item><item>War diaries of War Office directorates for the Second World War, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14372"WO 165</extref></item><item>Minutes of directors' meetings, 1906 to 1913, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14448"WO 242</extref></item><item>Establishment papers, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14233"WO 24</extref></item><item>War Office: Military Intelligence section 9 (MI9), Escape and Evasion maps, Second World War, <extref href="ttp://discovery.nationalarchives.gov.uk/details/r/C16747711"WO 418</extref></item></list> </scopecontent>" ;
        dc:identifier "WO_1" ;
        dc:type "division" ;
        dc:type cldtype:CollectionCollection ;
        edm:type "TEXT" ;
        dc:language "eng" ;
        dcterms:extent "28" ;
        dcterms:accrualPolicy accpol:passive ;
        dcterms:isPartOf cat:departmentWO .

cat:a0_divisionWO_1 rdf:type ore:Aggregation ;
    edm:aggregatedCHO cat:divisionWO_1 ;
    edm:dataProvider "The National Archives (UK)" ;
    edm:provider "The National Archives (UK)" ;
    edm:rights "http://rightsstatements.org/vocab/InC/1.0/" ;
    edm:isShownBy "https://discovery.nationalarchives.gov.uk/details/r/X12" .

cat:seriesWO_1_1 rdf:type edm:ProvidedCHO ;
    dc:publisher cat:TNA ;
        dc:title "War Office and predecessors: Secretary-at-War, Secretary of State for War, and Commander-in-Chief, In-letters and Miscellaneous Papers" ;
        dc:description "<scopecontent><head>Scope and Content</head> These are for the most part journals and indexes of letters in the letter books of the Secretary-at-War and of War Office departments. There are also journals with an index of the Board of General Officers, 1808 to 1816; an index of the Agent General of Militia, 1803 to 1804; and an entry book of memoranda and precedents relating to the War Office establishment in 1782. Many of the letter books to which these indexes refer cannot be identified and probably do not exist. The series includes material which would otherwise be found amongst the Records of the Board of General Officers and records relating to soldiers' clothing (<ref>WO</ref>). </scopecontent>" ;
        dc:identifier "WO_1_1" ;
        dc:type "series" ;
        dc:type cldtype:CollectionCollection ;
        edm:type "TEXT" ;
        dc:language "eng" ;
        dcterms:extent "1138" ;
        dcterms:accrualPolicy accpol:passive ;

```
        cat:arrangement "<arrangement><p>Papers of the Secretary-at-War and the Secretary of State for War are
arranged under the various military stations or public departments from which they were sent, or under the
subjects with which they deal, or alphabetically by the name of the writer.</p><p>The papers are also grouped
into two larger and two smaller subseries: </p><p>I: Records of the old War Office (Secretary-at-War),
arranged into three groups: <list><item>1. The books of original dispatches from various commanders, </
item><item>2. Correspondence received from other government departments, </item><item>3. Correspondence from
others, arranged roughly by the alphabetical order of the surnames of the senders, and including certain
correspondence which might be expected to be found in the volumes of the first two groups. This has been
divided into two sequences: the first consisting of foolscap volumes, the second of smaller volumes. </item></
list></p><p>II: Records of the War Department (Secretary of State for War), arranged into four groups: </
p><p><list><item>1. Despatches, correspondence and papers from the French Wars period (1794-1816). Some of
these books start earlier than July 1794, before the separate War Department was created, </item><item>2. A
few volumes covering the period 1817-1831, when the major part of the work of the Colonial and War Department
was colonial, </item><item>3. Military in-letters from the 1830s to the 1850s, arranged by colonies, </
item><item>4. Crimean War correspondence and despatches. </item></list></p><p>III: Loose bundles of despatches
addressed to the Commander-in-Chief. </p><p>IV: Miscellaneous volumes. </p></arrangement>" ;
        dcterms:isPartOf cat:divisionWO_1 .

cat:a0_seriesWO_1_1 rdf:type ore:Aggregation ;
    edm:aggregatedCHO cat:seriesWO_1_1 ;
    edm:dataProvider "The National Archives (UK)" ;
    edm:provider "The National Archives (UK)" ;
    edm:rights "http://rightsstatements.org/vocab/InC/1.0/" ;
    edm:isShownBy "https://discovery.nationalarchives.gov.uk/details/r/X123" .

cat:subseriesWO_1_1_2 rdf:type edm:ProvidedCHO ;
    dc:publisher cat:TNA ;
        dc:title "II. WAR DEPARTMENT IN-LETTERS AND PAPERS" ;
        dc:identifier "WO_1_1_2" ;
        dc:type "sub-series" ;
        dc:type cldtype:CollectionCollection ;
        edm:type "TEXT" ;
        dc:language "eng" ;
        dcterms:accrualPolicy accpol:passive ;
        dcterms:isPartOf cat:seriesWO_1_1 .

cat:a0_subseriesWO_1_1_2 rdf:type ore:Aggregation ;
    edm:aggregatedCHO cat:subseriesWO_1_1_2 ;
    edm:dataProvider "The National Archives (UK)" ;
    edm:provider "The National Archives (UK)" ;
    edm:rights "http://rightsstatements.org/vocab/InC/1.0/" ;
    edm:isShownBy "https://discovery.nationalarchives.gov.uk/details/r/X1234" .

cat:subsubseriesWO_1_1_2_1 rdf:type edm:ProvidedCHO ;
    dc:publisher cat:TNA ;
        dc:title "1. Of the French Wars period" ;
        dc:identifier "WO_1_1_2_1" ;
        dc:type "sub-sub-series" ;
        dc:type cldtype:CollectionCollection ;
        edm:type "TEXT" ;
        dc:language "eng" ;
        dcterms:accrualPolicy accpol:passive ;
        dcterms:isPartOf cat:subseriesWO_1_1_2 .

cat:a0_subsubseriesWO_1_1_2_1 rdf:type ore:Aggregation ;
    edm:aggregatedCHO cat:subsubseriesWO_1_1_2_1 ;
    edm:dataProvider "The National Archives (UK)" ;
    edm:provider "The National Archives (UK)" ;
    edm:rights "http://rightsstatements.org/vocab/InC/1.0/" ;
    edm:isShownBy "https://discovery.nationalarchives.gov.uk/details/r/X12345" .

cat:pieceWO_1_1_2_1_921 rdf:type edm:ProvidedCHO ;
```

```
    dc:publisher cat:TNA ;
        dc:description "<scopecontent><p>j. Intelligence: Prince de Bouillon: Correspondence. (Described at
item level) </p></scopecontent>" ;
        dc:identifier "WO_1_1_2_1_921" ;
        dc:type "piece" ;
        dc:type cldtype:CollectionPhysicalObject ;
        edm:type "IMAGE" ;
        dcterms:accrualPolicy accpol:passive ;
        dcterms:accessRights "<closure><status>O</status><type>N</type><code>30</code></closure>" ;
        dcterms:isPartOf cat:subsubseriesWO_1_1_2_1 .

cat:a0_pieceWO_1_1_2_1_921 rdf:type ore:Aggregation ;
    edm:aggregatedCHO cat:pieceWO_1_1_2_1_921 ;
    edm:dataProvider "The National Archives (UK)" ;
    edm:provider "The National Archives (UK)" ;
    edm:rights "http://rightsstatements.org/vocab/InC/1.0/" ;
    edm:isShownBy "https://discovery.nationalarchives.gov.uk/details/r/X123456" .

cat:itemWO_1_1_2_1_921_233 rdf:type edm:ProvidedCHO ;
    dc:publisher cat:TNA ;
        dc:description "<scopecontent><p>Folio(s) 233-236: Bertin to Bouillon. Discussing arms received by
Vieuville and 'Rodolphe' and 19 Oct. landings; traitors and informants; weakness of St. Malo garrison and
desirability of retaining control of Chausey.</p><p>Date and Place: 1795 20 Oct. Jersey.</p></scopecontent>" ;
        dc:identifier "WO_1_1_2_1_921_233" ;
        dc:type "item" ;
        dc:type cldtype:CollectionPhysicalObject ;
        edm:type "IMAGE" ;
        dcterms:accrualPolicy accpol:closed ;
        cat:note "Enclosed in ff 229-232 French." ;
        dcterms:accessRights "<closure><status>O</status><type>N</type><code>30</code></closure>" ;
        dcterms:isPartOf cat:pieceWO_1_1_2_1_921 .

cat:a0_itemWO_1_1_2_1_921_233 rdf:type ore:Aggregation ;
    edm:aggregatedCHO cat:itemWO_1_1_2_1_921_233 ;
    edm:dataProvider "The National Archives (UK)" ;
    edm:provider "The National Archives (UK)" ;
    edm:rights "http://rightsstatements.org/vocab/InC/1.0/" ;
    edm:isShownBy "https://discovery.nationalarchives.gov.uk/details/r/X1234567" .
```

A1.4 RiC-O v0.1 RDF (Turtle) Syntax Example

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix ric-rst: <https://www.ica.org/standards/RiC/vocabularies/recordSetTypes#> .
@prefix rico: <https://www.ica.org/standards/RiC/ontology#> .
@prefix cat-rst: <http://nationalarchives.gov.uk/catalogue/recordSetTypes#> .
@prefix cat: <http://nationalarchives.gov.uk/catalogue> .
@base <https://www.ica.org/standards/RiC/RiC-O_v0-1.rdf> .

cat:TNA rdf:type rico:CorporateBody ;
        rico:identifier "66" ;
        rico:name "The National Archives" ;
        .

cat:departmentWO rdf:type rico:RecordSet ;
        rico:name "Records created or inherited by the War Office, Armed Forces, Judge Advocate General, and
related bodies" ;
        rico:scopeAndContent "<scopecontent><head>Scope and Content</head><p>Records created or inherited by
the War Office, Armed Forces, Judge Advocate General, and related bodies relating to the administration of the
armed forces.</p> <p>Comprises records of: <list><item>General War Office records</item><item>Chief of the
```

(Imperial) General Staff and its directorates</item><item>Commander-in-Chief, Military Secretary and Army Council</item><item>Army Medical services</item><item>Other administrative departments of the War Office</item><item>Finance departments</item><item>Central Department of the Permanent Secretary of State</item><item>Land Branch</item><item>Board of General Officers</item><item>Royal Chelsea and Kilmainham Hospitals</item><item>Constable of the Tower of London</item><item>Research establishments</item><item>Armed forces records from commands, headquarters, regiments and corps</item><item>Auxiliary forces</item><item>Army in Ireland</item><item>Armed forces service records</item><item>Private office papers and private collections</item><item>Commissions, committees and councils, etc</item><item>Ordnance Office and its War Office successors</item><item>Judge Advocate General</item></list> WO 56, WO 125, WO 200, WO 249 and WO 289 are numbers not used. WO 20 and WO 21 were transferred to other series before 1905. </scopecontent>" ;
        rico:classifcation "WO" ;
        rico:identifier "http://catalogue.nationalarchives.gov.uk/WO" ;
        rico:hasRecordSetType ric-rst:Fonds ;
        rico:language "eng" ;
        rico:recordResourceExtent "417" ;
        rico:accrualStatus "open" ;
        rico:heldBy cat:TNA
        .

cat:divisionWO_1 rdf:type rico:RecordSet ;
        rico:name "General records of the War Office and predecessors and successor" ;
        rico:scopeAndContent "<scopecontent><head>Scope and Content</head> General records of the War Office and predecessors and successors relating to various aspects of its duties for the administration of the armed forces. Comprises: <list><item>Correspondence and papers of the Secretary-at-War, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14212"WO 1</extref>, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14249"WO 40</extref> and <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14252"WO 43</extref></item><item>Registered correspondence, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14241"WO 32</extref>, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14252"WO 43</extref> and <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14348"WO 141</extref></item><item>Indexes to correspondence, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14346"WO 139</extref> and <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14601"WO 900</extref></item><item>Correspondence and papers relating to the First World War, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14368"WO 161</extref></item><item>Entry books of out-letters, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14215"WO 4</extref> with indexes in <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14213"WO 2</extref></item><item>Other entry books, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14234"WO 25</extref>, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14235"WO 26</extref> and <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14239"WO 30</extref></item><item>Out-letters of the Secretary of State for War and Secretary of State for War and the Colonies, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14217"WO 6</extref></item><item>Various departmental out-letters, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14218"WO 7</extref></item><item>Personal files, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14345"WO 138</extref></item><item>Civilian personal files, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14539"WO 335</extref></item><item>Printed reports and papers including hand books and manuals in <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14242"WO 33</extref>, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14484"WO 279</extref> and <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14492"WO 287</extref></item><item>Maps and plans, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14286"WO 78</extref></item><item>War diaries of War Office directorates for the Second World War, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14372"WO 165</extref></item><item>Minutes of directors' meetings, 1906 to 1913, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14448"WO 242</extref></item><item>Establishment papers, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14233"WO 24</extref></item><item>War Office: Military Intelligence section 9 (MI9), Escape and Evasion maps, Second World War, <extref href="ttp://discovery.nationalarchives.gov.uk/details/r/C16747711"WO 418</extref></item></list> </scopecontent>" ;
        rico:classification "WO_1" ;
        rico:identifier "http://catalogue.nationalarchives.gov.uk/WO_1" ;

```
        rico:hasRecordSetType ric-rst:SubFonds ;
        rico:language "eng" ;
        rico:recordResourceExtent "28" ;
        rico:accrualStatus "open" ;
        rico:includedIn cat:departmentWO
        .

cat:seriesWO_1_1 rdf:type rico:RecordSet ;
        rico:name "War Office and predecessors: Secretary-at-War, Secretary of State for War, and Commander-
in-Chief, In-letters and Miscellaneous Papers" ;
        rico:scopeAndContent "<scopecontent><head>Scope and Content</head><p>These are for the most part
journals and indexes of letters in the letter books of the Secretary-at-War and of War Office departments.
There are also journals with an index of the Board of General Officers, 1808 to 1816; an index of the Agent
General of Militia, 1803 to 1804; and an entry book of memoranda and precedents relating to the War Office
establishment in 1782. Many of the letter books to which these indexes refer cannot be identified and probably
do not exist.</p><p>The series includes material which would otherwise be found amongst the Records of the
Board of General Officers and records relating to soldiers' clothing (<ref>WO</ref>).</p></scopecontent>" ;
        rico:classification "WO_1_1" ;
        rico:identifier "http://catalogue.nationalarchives.gov.uk/WO_1_1" ;
        rico:hasRecordSetType ric-rst:Series ;
        rico:language "eng" ;
        rico:recordResourceExtent "1138" ;
        rico:accrualStatus "open" ;
        rico:structure "<arrangement><p>Papers of the Secretary-at-War and the Secretary of State for War are
arranged under the various military stations or public departments from which they were sent, or under the
subjects with which they deal, or alphabetically by the name of the writer.</p><p>The papers are also grouped
into two larger and two smaller subseries: </p><p>I: Records of the old War Office (Secretary-at-War),
arranged into three groups: <list><item>1. The books of original dispatches from various commanders, </
item><item>2. Correspondence received from other government departments, </item><item>3. Correspondence from
others, arranged roughly by the alphabetical order of the surnames of the senders, and including certain
correspondence which might be expected to be found in the volumes of the first two groups. This has been
divided into two sequences: the first consisting of foolscap volumes, the second of smaller volumes. </item></
list></p><p>II: Records of the War Department (Secretary of State for War), arranged into four groups: </
p><p><list><item>1. Despatches, correspondence and papers from the French Wars period (1794-1816). Some of
these books start earlier than July 1794, before the separate War Department was created, </item><item>2. A
few volumes covering the period 1817-1831, when the major part of the work of the Colonial and War Department
was colonial, </item><item>3. Military in-letters from the 1830s to the 1850s, arranged by colonies, </
item><item>4. Crimean War correspondence and despatches. </item></list></p><p>III: Loose bundles of despatches
addressed to the Commander-in-Chief. </p><p>IV: Miscellaneous volumes. </p></arrangement>" ;
        rico:includedIn cat:divisionWO_1
        .

cat:subseriesWO_1_1_2 rdf:type rico:RecordSet ;
        rico:name "II. WAR DEPARTMENT IN-LETTERS AND PAPERS" ;
        rico:classification "WO_1_1_2" ;
        rico:identifier "http://catalogue.nationalarchives.gov.uk/WO_1_1_2" ;
        rico:hasRecordSetType ric-rst:SubSeries ;
        rico:accrualStatus "open" ;
        rico:includedIn cat:seriesWO_1_1
        .

cat:subsubseriesWO_1_1_2_1 rdf:type rico:RecordSet ;
        rico:name "1. Of the French Wars period" ;
        rico:classification "WO_1_1_2_1" ;
        rico:identifier "http://catalogue.nationalarchives.gov.uk/WO_1_1_2_1" ;
        rico:hasRecordSetType cat-rst:SubSubSeries ;
        rico:accrualStatus "open" ;
        rico:includedIn cat:subseriesWO_1_1_2
        .

cat:pieceWO_1_1_2_1_921 rdf:type rico:RecordSet ;
```

```
        rico:scopeAndContent "<scopecontent><p>j. Intelligence: Prince de Bouillon: Correspondence. (Described
at item level) </p></scopecontent>" ;
        rico:classification "WO_1_1_2_1_921" ;
        rico:identifier "http://catalogue.nationalarchives.gov.uk/WO_1_1_2_1_921" ;
        rico:hasRecordSetType ric-rst:File ;
        rico:accrualStatus "open" ;
        rico:includedIn cat:subsubseriesWO_1_1_2_1
        .

cat:itemWO_1_1_2_1_921_233 rdf:type rico:Record ;
        rico:scopeAndContent "<scopecontent><p>Folio(s) 233-236: Bertin to Bouillon. Discussing arms received
by Vieuville and 'Rodolphe' and 19 Oct. landings; traitors and informants; weakness of St. Malo garrison and
desirability of retaining control of Chausey.</p><p>Date and Place: 1795 20 Oct. Jersey.</p></scopecontent>" ;
        rico:classification "WO_1_1_2_1_921_233" ;
        rico:identifier "http://catalogue.nationalarchives.gov.uk/WO_1_1_2_1_921_233" ;
        rico:hasRecordSetType ric-rst:Item ;
        rico:accrualStatus "closed" ;
        rico:descriptiveNote "Enclosed in ff 229-232 French." ;
        rico:conditionsOfAccess "<closure><status>O</status><type>N</type><code>30</code></closure>" ;
        rico:includedIn cat:pieceWO_1_1_2_1_921
        .

cat:itemWO_1_1_2_1_921_233_0 rdf:type rico:Instantiation ;
        rico:classification "WO_1_1_2_1_921_233_0" ;
        rico:identifier "http://catalogue.nationalarchives.gov.uk/WO_1_1_2_1_921_233_0" ;
        rico:descriptiveNote "Default paper instantiation" ;
        rico:instantiates cat:itemWO_1_1_2_1_921_233
        .
```

A1.5 Matterhorn RDF (Turtle) Syntax Example

```
@prefix accpol: <http://purl.org/cld/accpol/> .
@prefix cat: <http://nationalarchives.gov.uk/catalogue> .
@prefix cat-rst: <http://nationalarchives.gov.uk/catalogue/recordSetTypes#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix evRelAgRole: <http://id.loc.gov/vocabulary/preservation/linkingAgentRoleEvent/> .
@prefix evType: <http://id.loc.gov/vocabulary/preservation/eventType> .
@prefix premis: <http://id.loc.gov/vocabulary/preservation/> .
@prefix prov: <http://www.w3.org/ns/prov#> .
@prefix relSubType: <http://id.loc.gov/vocabulary/preservation/relationshipSubType/> .
@prefix rdac: <http://rdaregistry.info/Elements/c/> .
@prefix rdau: <http://rdaregistry.info/Elements/u/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix pra: <http://nationalarchives.gov.uk/public-records-act/1967/#> .
@prefix xsd:  <http://www.w3.org/2001/XMLSchema#> .


cat:departmentWO rdf:type premis:IntellectualEntity ;
        dc:title "Records created or inherited by the War Office, Armed Forces, Judge Advocate
General, and related bodies" ;
        dc:description "<scopecontent><head>Scope and Content</head><p>Records created or
inherited by the War Office, Armed Forces, Judge Advocate General, and related bodies relating to
the administration of the armed forces.</p> <p>Comprises records of: <list><item>General War
Office records</item><item>Chief of the (Imperial) General Staff and its directorates</
item><item>Commander-in-Chief, Military Secretary and Army Council</item><item>Army Medical
services</item><item>Other administrative departments of the War Office</item><item>Finance
```

departments</item><item>Central Department of the Permanent Secretary of State</item><item>Land Branch</item><item>Board of General Officers</item><item>Royal Chelsea and Kilmainham Hospitals</item><item>Constable of the Tower of London</item><item>Research establishments</item><item>Armed forces records from commands, headquarters, regiments and corps</item><item>Auxiliary forces</item><item>Army in Ireland</item><item>Armed forces service records</item><item>Private office papers and private collections</item><item>Commissions, committees and councils, etc</item><item>Ordnance Office and its War Office successors</item><item>Judge Advocate General</item></list> WO 56, WO 125, WO 200, WO 249 and WO 289 are numbers not used. WO 20 and WO 21 were transferred to other series before 1905. </scopecontent>" ;
dc:identifier "WO" ;
dc:type cat-rst:Department ;
dct:language "eng" ;
rdau:P60550 "417" ;
dct:accrualPolicy accpol:passive ;
dct:publisher cat:TNA
.

cat:divisionWO_1 rdf:type premis:IntellectualEntity ;
dc:title "General records of the War Office and predecessors and successor" ;
dc:description "<scopecontent><head>Scope and Content</head> General records of the War Office and predecessors and successors relating to various aspects of its duties for the administration of the armed forces. Comprises: <list><item>Correspondence and papers of the Secretary-at-War, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14212"WO 1</extref>, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14249"WO 40</extref> and <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14252"WO 43</extref></item><item>Registered correspondence, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14241"WO 32</extref>, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14252"WO 43</extref> and <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14348"WO 141</extref></item><item>Indexes to correspondence, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14346"WO 139</extref> and <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14601"WO 900</extref></item><item>Correspondence and papers relating to the First World War, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14368"WO 161</extref></item><item>Entry books of out-letters, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14215"WO 4</extref> with indexes in <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14213"WO 2</extref></item><item>Other entry books, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14234"WO 25</extref>, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14235"WO 26</extref> and <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14239"WO 30</extref></item><item>Out-letters of the Secretary of State for War and Secretary of State for War and the Colonies, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14217"WO 6</extref></item><item>Various departmental out-letters, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14218"WO 7</extref></item><item>Personal files, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14345"WO 138</extref></item><item>Civilian personal files, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14539"WO 335</extref></item><item>Printed reports and papers including hand books and manuals in <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14242"WO 33</extref>, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14484"WO 279</extref> and <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14492"WO 287</extref></item><item>Maps and plans, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14286"WO 78</extref></item><item>War diaries of War Office directorates for the Second World War, <extref href="ttp://

discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14372"WO 165</extref></item><item>Minutes of directors' meetings, 1906 to 1913, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14448"WO 242</extref></item><item>Establishment papers, <extref href="ttp://discovery.nationalarchives.gov.uk/SearchUI/details?Uri=C14233"WO 24</extref></item><item>War Office: Military Intelligence section 9 (MI9), Escape and Evasion maps, Second World War, <extref href="ttp://discovery.nationalarchives.gov.uk/details/r/C16747711"WO 418</extref></item></list> </scopecontent>" ;
dc:identifier "WO_1" ;
dc:type cat-rst:Division ;
dct:language "eng" ;
rdau:P60550 "28" ;
dct:accrualPolicy accpol:passive ;
dct:publisher cat:TNA ;
relSubType:isp cat:departmentWO
.

cat:seriesWO_1_1 rdf:type premis:IntellectualEntity ;
dc:title "War Office and predecessors: Secretary-at-War, Secretary of State for War, and Commander-in-Chief, In-letters and Miscellaneous Papers" ;
dc:description "<scopecontent><head>Scope and Content</head> These are for the most part journals and indexes of letters in the letter books of the Secretary-at-War and of War Office departments. There are also journals with an index of the Board of General Officers, 1808 to 1816; an index of the Agent General of Militia, 1803 to 1804; and an entry book of memoranda and precedents relating to the War Office establishment in 1782. Many of the letter books to which these indexes refer cannot be identified and probably do not exist. The series includes material which would otherwise be found amongst the Records of the Board of General Officers and records relating to soldiers' clothing (<ref>WO</ref>). </scopecontent>" ;
dc:identifier "WO_1_1" ;
dc:type cat-rst:Series ;
dct:language "eng" ;
rdau:P60550 "1138" ;
dct:accrualPolicy accpol:passive ;
dct:publisher cat:TNA ;
rdau:P60348 "<arrangement> Papers of the Secretary-at-War and the Secretary of State for War are arranged under the various military stations or public departments from which they were sent, or under the subjects with which they deal, or alphabetically by the name of the writer. The papers are also grouped into two larger and two smaller subseries: I: Records of the old War Office (Secretary-at-War), arranged into three groups: <list><item>1. The books of original dispatches from various commanders, </item><item>2. Correspondence received from other government departments, </item><item>3. Correspondence from others, arranged roughly by the alphabetical order of the surnames of the senders, and including certain correspondence which might be expected to be found in the volumes of the first two groups. This has been divided into two sequences: the first consisting of foolscap volumes, the second of smaller volumes. </item></list> II: Records of the War Department (Secretary of State for War), arranged into four groups: <list><item>1. Despatches, correspondence and papers from the French Wars period (1794-1816). Some of these books start earlier than July 1794, before the separate War Department was created, </item><item>2. A few volumes covering the period 1817-1831, when the major part of the work of the Colonial and War Department was colonial, </item><item>3. Military in-letters from the 1830s to the 1850s, arranged by colonies, </item><item>4. Crimean War correspondence and despatches. </item></list> III: Loose bundles of despatches addressed to the Commander-in-Chief. IV: Miscellaneous volumes. </arrangement>" ;
relSubType:isp cat:divisionWO_1
.

cat:subseriesWO_1_1_2 rdf:type premis:IntellectualEntity ;

```
        dc:title "II. WAR DEPARTMENT IN-LETTERS AND PAPERS" ;
        dc:identifier "WO_1_1_2" ;
        dc:type cat-rst:SubSeries ;
        dct:accrualPolicy accpol:passive ;
        dct:publisher cat:TNA ;
        relSubType:isp cat:seriesWO_1_1
        .

cat:subsubseriesWO_1_1_2_1 rdf:type premis:IntellectualEntity ;
        dc:title "1. Of the French Wars period" ;
        dc:identifier "WO_1_1_2_1" ;
        dc:type cat-rst:SubSubSeries ;
        dct:accrualPolicy accpol:passive ;
        dct:publisher cat:TNA ;
        relSubType:isp cat:subseriesWO_1_1_2
        .

cat:pieceWO_1_1_2_1_921 rdf:type premis:IntellectualEntity ;
        dc:description "<scopecontent><p>j. Intelligence: Prince de Bouillon: Correspondence.
(Described at item level) </p></scopecontent>" ;
        dc:identifier "WO_1_1_2_1_921" ;
        dc:type cat-rst:Piece ;
        dct:accrualPolicy accpol:passive ;
        dct:publisher cat:TNA ;
        relSubType:isp cat:subsubseriesWO_1_1_2_1
        .

cat:itemWO_1_1_2_1_921_233 rdf:type premis:IntellectualEntity ;
        dc:description "<scopecontent><p>Folio(s) 233-236: Bertin to Bouillon. Discussing arms
received by Vieuville and 'Rodolphe' and 19 Oct. landings; traitors and informants; weakness of
St. Malo garrison and desirability of retaining control of Chausey.</p><p>Date and Place: 1795 20
Oct. Jersey.</p></scopecontent>" ;
        dc:identifier "WO_1_1_2_1_921_233" ;
        dc:type cat-rst:Item ;
        dct:accrualPolicy accpol:closed ;
        dct:publisher cat:TNA ;
        rdau:P60470 "Enclosed in ff 229-232 French." ;
        premis:rightsStatus [
           premis:basis pra:30_year_rule ;
           premis:determinationDate: "1234-01-01"^^xsd:date ;
        ] ;
        relSubType:isp cat:pieceWO_1_1_2_1_921 ;
        prov:wasGeneratedBy cat:TNA
        .

cat:itemWO_1_1_2_1_921_233_0 rdf:type premis:Representation ;
    dc:title "Default paper manifestation" ;
        dc:identifier "WO_1_1_2_1_921_233_0" ;
        relSubType:rep cat:itemWO_1_1_2_1_921_233
        .

# ====== agents
cat:TNA rdf:type rdac:C10005 ;
        dc:identifier "66" ;
        rdau:P60368 "The National Archives" ;
        dc:alternative "TNA";
```

```
        dct:language "eng" ;
        .

# ====== events
cat:accession1 rdf:type evType:accession ;
    evRelAgRole:aut cat:TNA ;
    premis:hasOutcome: "Success" ;
    prov:used cat:itemWO_1_1_2_1_921_233
    .

# ====== closure
pra:30_year_rule rdf:type premis:Statute ;
    premis:terms "Records become open thirty years after the 1st of January in the year after they
were created" ;
    premis:citation "until the expiration of the period of thirty years beginning with the first
day of January in the year next after that in which they were created, or of such other period" ;
    premis:jurisdiction <http://publications.europa.eu/resource/authority/country/GBR>
    .
```