Implementing Archaeological Time Periods Using CIDOC CRM and SKOS

Ceri Binding

University of Glamorgan, Pontypridd, UK
cbinding@glam.ac.uk

Abstract. Within the archaeology domain, datasets frequently refer to time periods using a variety of textual or numeric formats. Traditionally controlled vocabularies of time periods have used classification notation and the collocation of terms in the printed form to represent and convey tacit information about the relative order of concepts. The emergence of the semantic web entails encoding this knowledge into machine readable forms, and so the meaning of this informal ordering arrangement can be lost. Conversion of controlled vocabularies to Simple Knowledge Organisation System (SKOS) format provides a formal basis for semantic web indexing but does not facilitate chronological inference - as thesaurus relationship types are an inappropriate mechanism to fully describe temporal relationships. This becomes an issue in archaeological data where periods are often described in terms of (e.g.) named monarchs or emperors, without additional information concerning relative chronological context.

An exercise in supplementing existing controlled vocabularies of time period concepts with dates and temporal relationships was undertaken as part of the Semantic Technologies for Archaeological Resources (STAR) project. The general aim of the STAR project is to demonstrate the potential benefits in cross searching archaeological data conforming to a common overarching conceptual data structure schema - the CIDOC Conceptual Reference Model (CRM). This paper gives an overview of STAR applications and services and goes on to particularly focus on issues concerning the extraction and representation of time period information.

Keywords: CIDOC CRM, SKOS, semantic interoperability, time periods, archaeology.

1 Introduction

The work described here draws on work carried out for the AHRC funded Semantic Technologies for Archaeology Resources (STAR) project [1]. This is a 3 year project in collaboration with English Heritage, the broad aim of the research being to investigate the utility of mapping different archaeological datasets to a common overarching ontology, where the datasets are indexed using domain specific thesauri and glossaries. The goal is to demonstrate effective search across multiple different archaeological datasets and associated grey literature documents.

L. Aroyo et al. (Eds.): ESWC 2010, Part I, LNCS 6088, pp. 273–287, 2010.
© Springer-Verlag Berlin Heidelberg 2010

The current situation within archaeology is one of fragmented datasets and applications, with different terminology systems. The interpretation of a find (or free text report of an excavation) may not employ the same terms as the underlying dataset. Similarly searchers from different scientific perspectives may not use the same terminology. Separate datasets employ distinct schema for semantically equivalent information. Entities and relationships may have different names but be semantically equivalent. Even when datasets are made available on the Web, effective cross search is not possible due to these semantic interoperability issues [2].

STAR has aimed to address these concerns by exploiting the potential of a standard ontology for cultural heritage (extended for the archaeology excavation and analysis process) to link digital archive databases, vocabularies and the associated grey literature. This paper gives an overview of STAR applications and services and goes on to particularly focus on issues concerning the extraction and representation of time period information.

1.1 Extending CIDOC CRM for Use in Archaeology

Within the cultural heritage domain, the CIDOC Conceptual Reference Model (CRM) has emerged as a core ontological model [3]. The CRM is the outcome of more than a decade of work by the CIDOC Documentation Standards Working Group, and has more recently become an ISO Standard (ISO 21127:2006). The scope of the model encompasses the general cultural heritage domain and it is envisaged as ‘semantic glue’ useful for mediating between sometimes diverse information sources.

CIDOC CRM deals with concepts at a high level of generality. For working with archaeological datasets at a more detailed level, the English Heritage Centre for Archaeology developed an ontological model (CRM-EH) [4] as a CRM extension covering the archaeological excavation and analysis workflow. This model only existed in document form so in collaboration with staff at English Heritage an RDF implementation was first produced by Glamorgan [5], referencing and complementing the existing published (v4.2) RDFS implementation of the CRM [6]. Using this model, selections were extracted from multiple excavation datasets via SQL queries, and stored as RDF files. This work was significantly assisted by a semi-automatic mapping and data extraction tool. Further details of this process are discussed in [7].

1.2 STAR Architecture

The STAR system enables cross-search on multiple excavation datasets including Raunds Roman (RRAD), Raunds Prehistoric (RPRE), Museum of London (MoLAS), Silchester Roman (LEAP) and Stanwick sampling (STAN). The final system will extend this cross-search to data extracted from excavation reports originating from the OASIS index of grey literature, operated by the Archaeology Data Service (ADS).

The general architecture of the STAR system is shown in Fig. 1. A common RDF data store holds the CRM-EH ontology, thesauri and glossaries, and amalgamated data extracted from the previously separate databases. The data store will eventually also hold annotations extracted from the grey literature documents. Applications communicate to the server via web services (see section 1.3). Search result items offer

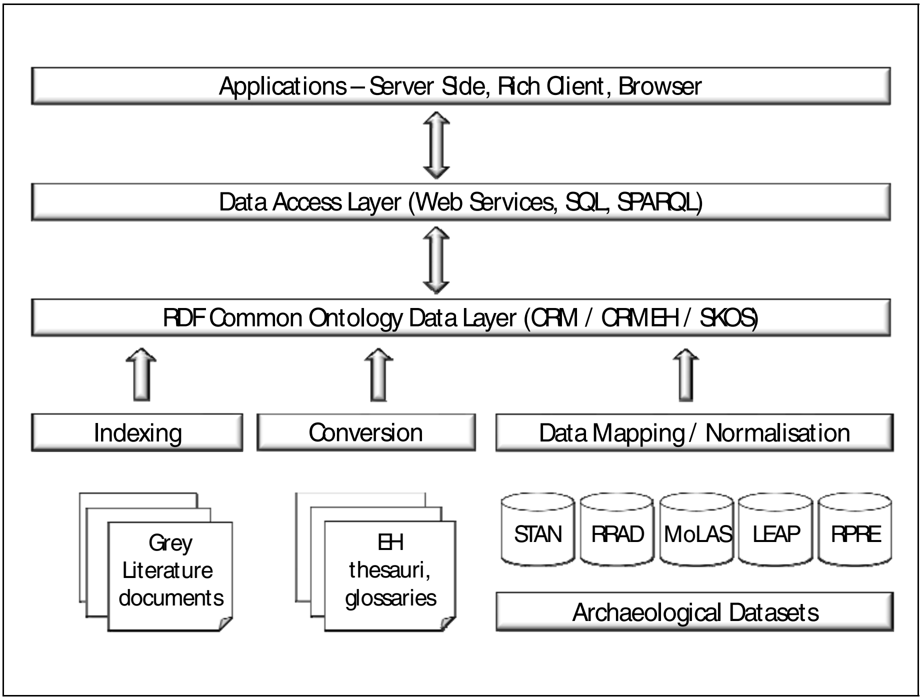

Fig. 1. General architecture of the STAR system

entry points into the structured data; allowing a user to browse to related data items via CRM relationship chains.

1.3 SKOS Based Terminology Services

STAR employs SKOS (Simple Knowledge Organization System) [8] as the representation format for domain specific thesauri and glossaries associated with the archaeological datasets. Seven English Heritage thesauri have been converted to SKOS RDF representation, along with various glossary term lists [9]. SKOS is a W3C Recommendation based on a formal data model intended as an RDF representation standard for the family of knowledge organization systems, with a lightweight semantics designed primarily for information retrieval purposes. This offers a cost effective approach for dealing with thesauri for STAR purposes. For other recent work employing SKOS see, for example, Isaac et al. [10] on aligning thesauri and Tuominen et al. [11] on various SKOS based semantic web services.

A set of terminology web services has been developed for the project [12] based on a subset of the SWAD Europe SKOS API [13] with extensions for concept expansion. The services provide term look up across the thesauri held in the system, allowing search systems to be augmented by SKOS-based vocabulary resources. Queries are often expressed at a different level of generalization from document content or metadata, so concepts may be expanded by synonyms or by semantically related concepts [14]. In addition to STAR, the services have been used by the DelosDLMS prototype next-generation Digital Library management system and the ADS ArcheoTools project. The services can be made applicable to other specialist domains by loading them with different SKOS thesauri. Further details of the services are given in [15].

2 Modelling Temporal Relationships

Traditionally controlled vocabularies have used classification notation and the collocation of terms in the display form to represent and convey tacit information about their relative order. [BS 8723-2:2005] states:

"*The terms in an array may be arranged either alphabetically or systematically. Alphabetical sequence should be used when there is no other obvious way to arrange a group of concepts. Systematic sequence should be used when it is likely to be familiar to most users, or when the arrangement helps to clarify the scope of the terms. In the example for electromagnetic radiation, the types of radiation are presented in order of increasing wavelength, as this might help some indexers in selecting the correct term(s).* "

This approach is used for various kinds of sequence:

- Temporal sequences (e.g. chronological progression: earliest → latest).
- Property continua (e.g. visual spectrum, wavelength, vocal range).
- Order of magnitude (smallest → largest).
- Derivative sequences (e.g. an evolutionary or developmental order).

In archaeology datasets the relative temporal context of objects and events is useful and important both for indexing and display purposes. Alphabetical ordering can be illogical - a chronological arrangement often being more appropriate. A particular application for this would be displaying archaeological finds within their chronological context and establishing linear paths for navigation through the data.

The Getty Thesaurus of Geographic Names (TGN) [16] facilitates the presentation of multiple historical variants for place names. The online version lists these variants in reverse chronological order, to present the likely most sought after information at the top of the list, whilst also illustrating a historical progression. However thesaurus relationships are not the most comprehensive mechanism to fully describe temporal relationships.

2.1 Types of Temporal Relationship

Previous research effort has been devoted to the modelling of specialised operators to define relationships between periods and events. Knight & Ma [17] defined a set of relationships including *before, after, during, pre, post, circa*. Interval Temporal Logic (a.k.a. Allen Algebra) [18] described a method for reasoning about temporal intervals. The system defined 13 possible types of relationship between periods – the equality condition plus 6 reciprocal property pairs. OWL-Time [19] (formerly DAML-Time) includes a practical RDF encoding of these properties for the Semantic Web. The properties (*intervalEquals, intervalBefore, intervalAfter* etc.) occur between

ProperInterval OWL classes. CIDOC CRM similarly models an equivalent set of relationships (*P114F.is_equal_in_time_to, P120F.occurs_before, P120B.occurs_after* etc.) as occurring between *E2.TemporalEntity* conceptual entities, as listed in Table 1. These relationships give no indication of scale so without supplementary information we cannot know the degree of temporal affinity between resources - only their relative chronological context.

Table 1. Relationships between time periods, with their associated inverse property

CIDOC CRM Property	OWL-Time Property	Transitive?
P114F.is_equal_in_time_to	intervalEquals	✓
P115F.finishes P115B.is_finished_by	intervalFinishes intervalFinishedBy	✓
P116F.starts P116B.is_started_by	intervalStarts intervalStartedBy	✓
P117F.occurs_during P117B.includes	intervalDuring intervalContains	✓
P118F.overlaps_in_time_with P118B.is_overlapped_in_time_by	intervalOverlaps intervalOverlappedBy	✗
P119F.meets_in_time_with P119B.is_met_in_time_by	intervalMeets intervalMetBy	✗
P120F.occurs_before P120B.occurs_after	intervalBefore intervalAfter	✓

```
A is_equal_in_time_to B :- (A.start = B.start AND A.end = B.end)
A finishes B :- (A.start > B.start AND A.end = B.end)
A is_finished_by B :- (A.start < B.start AND A.end = B.end)
A starts B :- (A.start = B.start AND A.end < B.end)
A is_started_by B :- (A.start = B.start AND A.end > B.end)
A occurs_during B :- (A.start > B.start AND A.end < B.end)
A includes B :- (A.start < B.start AND A.end > B.end)
A overlaps_in_time_with B :-
(A.start < B.start AND A.end > B.start AND A.end < B.end)
A is_overlapped_in_time_by B :-
(A.start > B.start AND A.start < B.end AND A.end > B.end)
A meets_in_time_with B :- (A.end = B.start)
A is_met_in_time_by B :- (A.start = B.end)
A occurs_before B :- (A.end < B.start)
A occurs_after B :- (A.start > B.end)
```

Fig. 2. Pseudo-code logic for temporal relationships, assuming valid(X) :- (X.start <= X.end)

Note: *transitive*[1] in Table 1 refers to relationships that are logically transitive; transitivity is not formally stated for the interval relations in the RDF implementations of CIDOC CRM and OWL-Time (the latter does however include separate transitive versions of the *before* and *after* properties). Fig. 2 shows how each of these formal relationships between periods can be deduced based on start/end dates.

[1] If a *transitive* relationship applies between successive members of a sequence, then that same relationship must also apply between *any* two members of the sequence taken in order. I.e. if A *occurs_before* B, and B *occurs_before* C, then A *occurs_before* C.

2.2 Using Temporal Relationships to Extend Controlled Vocabularies

As previously discussed, an established technique for modelling sequences within thesauri has been the use of classification notation to define a specific order within contiguous sequences of sibling terms. However collocation of sequential homogeneous terms is not always guaranteed within mono-hierarchical thesaurus structure due to *subsumption*, as described by Doerr [20]. Some examples of temporal sequences defined by classification notation can be found within the Getty Art & Architecture Thesaurus (AAT) [16] within the *Styles and Periods* facet. Fig. 3 illustrates how it is possible for a logical chronological sequence to span boundaries imposed by the hierarchical structure.

The primary temporal sequences encountered in Fig. 3 are:

1. Tudor → Stuart → Georgian
2. Regency → Victorian → Edwardian

A cursory analysis of the date spans represented by these terms (obtained via the associated scope notes), coupled with some knowledge of British reigns indicates that sequence 2 overlaps (and continues) sequence 1. However the evidence of term proximity alone is not sufficient in this case and the classification notation cannot describe the continuation across the imposed hierarchical boundary. The inclusion of temporal relationships would clarify that an item indexed using the term *Queen Anne* has an *earlier* historical context to an item indexed using the term *Regency*, even though these terms occur at different hierarchical positions.

As the STAR project is using the CIDOC CRM as an overarching ontology we represent the required additional relationships in terms of CRM properties in Fig. 4.

This additional information now formally links the 2 previously separate hierarchies and clarifies the relative ordering of items within the hierarchies. It states that *Caroline* and *Restoration* are represented by overlapping time periods, and that there is an interval between the end of *Georgian* and the start of *Victorian* (this is because the reign of William IV is not represented in the data). We have implied by using CRM relationships that all the periods referenced can also be regarded as CRM *E2.TemporalEntity* elements. However *E2* is an abstract class having no direct instances so it would be appropriate to also include statements declaring the type to be a specific subclass of *E2*, e.g. *E4.Period.* Other pertinent relationships between these time periods (*occurs_before*, *occurs_after* etc.) can then be derived from this initial set of assertions.

These relationships were all created by examining the dates mentioned in the scope notes - so why not just use dates directly? This is certainly a viable approach in cases where dates are commonly known and agreed, however in archaeology often absolute dates may be unknown or disputed, while relative ordering may be better agreed and established through the process of grouping and phasing. Section 3 discusses in more detail how we approached the alignment of data records with controlled terminology.

```
Hierarchy              Classification notation (truncated)  Date span

<British Renaissance-Baroque styles by reign>
. Tudor                .ALO.ARI.BIQ.BIQ.AFU.ALO.AFU          [1485-1603]
. . Elizabethan        .ALO.ARI.BIQ.BIQ.AFU.ALO.AFU.AFU      [1551-1603]
. Stuart               .ALO.ARI.BIQ.BIQ.AFU.ALO.ALO          [1603-1714]
. . Jacobean           .ALO.ARI.BIQ.BIQ.AFU.ALO.ALO.ALO      [1603-1625]
. . Caroline           .ALO.ARI.BIQ.BIQ.AFU.ALO.ALO.AFU      [1625-1685]
. . Restoration        .ALO.ARI.BIQ.BIQ.AFU.ALO.ALO.ARI      [1660-1688]
. . William and Mary   .ALO.ARI.BIQ.BIQ.AFU.ALO.ALO.AXC      [1688-1702]
. . Queen Anne         .ALO.ARI.BIQ.BIQ.AFU.ALO.ALO.BCW      [1702-1714]
. Georgian             .ALO.ARI.BIQ.BIQ.AFU.ALO.ARI          [1714-1830]
---------------------------------------------------------------------

Hierarchy              Classification notation (truncated)  Date span

<modern British styles by reign>
. Regency              .ALO.BCW.ALO.AXC.DIO.ALO.BCW.AFU      [1811-1830]
. Victorian            .ALO.BCW.ALO.AXC.DIO.ALO.BCW.ALO      [1837-1901]
. . Early Victorian    .ALO.BCW.ALO.AXC.DIO.ALO.BCW.ALO.AFU  [1837-1850]
. . High Victorian     .ALO.BCW.ALO.AXC.DIO.ALO.BCW.ALO.ALO  [1850-1870]
. . Late Victorian     .ALO.BCW.ALO.AXC.DIO.ALO.BCW.ALO.ARI  [1870-1901]
. Edwardian            .ALO.BCW.ALO.AXC.DIO.ALO.BCW.ARI      [1901-1910]
```

Fig. 3. Temporal sequences spanning hierarchical boundary (dates obtained from scope notes)

```
@prefix crm: < http://cidoc.ics.forth.gr/rdfs/cidoc_v4.2.rdfs#>
@prefix : <http://tempuri/concept#>

<Elizabethan> crm:P115F.finishes <Tudor> .
<Tudor> crm:P119F.meets_in_time_with <Stuart> .
<Jacobean> crm:P116F.starts <Stuart> ;
       crm:P119F.meets_in_time_with <Caroline> .
<Caroline> crm:P118F.overlaps_in_time_with <Restoration> .
<Restoration> crm:P119F.meets_in_time_with <William_and_Mary> .
<William_and_Mary> crm:P119F.meets_in_time_with <Queen_Anne> .
<Queen_Anne> crm:P115F.finishes <Stuart> .
<Stuart> crm:P119F.meets_in_time_with <Georgian> .
<Regency> crm:P115F.finishes <Georgian> .
<Victorian> crm:P120B.occurs_after <Georgian> ;
       crm:P119F.meets_in_time_with <Edwardian> .
<Early_Victorian> crm:P116F.starts <Victorian> ;
       crm:P119F.meets_in_time_with <High_Victorian> .
<High_Victorian> crm:P117F.occurs_during <Victorian> ;
       crm:P119F.meets_in_time_with <Late_Victorian> .
<Late_Victorian> crm:P115F.finishes <Victorian> .
```

Fig. 4. TURTLE syntax triples describing relationships between time periods

The boundaries of *terminus ante quem* ("limit before which") and *terminus post quem* ("limit after which") commonly used in the archaeological dating process can be modelled in CRM, although in our case the data available did not always fully support this finer level of reasoning. Refer to [21] and [22] for related work discussing the potential for chronological reasoning supported by the CRM.

3 Establishing Known Time Periods

Five archaeological datasets had been previously identified as suitable for use within the main STAR project:

- Raunds Roman Analytical Database (RRAD)
- Stanwick sampling data (STAN)[2]
- Raunds Prehistoric Database (RPRE)
- Silchester Roman Database (LEAP)
- Museum of London (MoLAS)

Within these datasets archaeological entities were typically associated with a date range rather than an absolute date. Time spans were expressed in a variety of different textual forms e.g. centuries, AD/BC years, named Roman Emperors / British Monarchs, the Three Age System[3]:

- MLC2-C3
- AD341-6
- Iron Age
- First half 1st century?
- Antonine
- LC2/EC3
- MLA

Even within an individual database field the formats used could vary, and sometimes multiple fields of the same record contained conflicting dating information. Obviously dates were important within archaeological datasets but in order to use the dates represented in any meaningful way, we had to undertake a process of data cleansing to convert this data to a more regular form.

Firstly for reference a controlled list of known time periods was collated to ensure a consistent approach across all databases. The English Heritage *Timelines* thesaurus [23] developed by the English Heritage Data Standards Unit was adopted for this purpose. The *Timelines* thesaurus integrates historic, cultural, political and geological chronological terminology, and has the scope of "*the whole span of human occupation of the United Kingdom*". It is currently only made available on request in draft format, as it has not been formally published. The thesaurus data received was first converted to SKOS format and then manually supplemented with dates deduced from the scope notes and from certain online historical resources.

---

2 The Stanwick sampling data represented the environmental sampling part of the Raunds Roman project, so the two databases were merged to enable easier subsequent data extraction.

3 A chronological classification system originally attributed to Nicholas Mahudel (and later to Christian Jürgensen Thomsen) as a way to describe historical epochs based on the predominant tool making materials of the time.

3.1 Adding Century Subdivisions to Known Time Periods

Prior to supplementing the thesaurus with dates we first established a convention for century subdivision and boundaries with reference to advice received from English Heritage. Centuries start at year 1 and end at year 100. In some cases in the datasets centuries AD were also observed prefixed with subdivision terms such as *Early, Mid, Late*. For consistency we established the following split to apply to all centuries AD:

- Early = 01→32
- Mid = 33→66
- Late = 67→100

Possibly an overlapping split of 01→50, 25→75, 50→100 respectively would also not be unreasonable, given the inherent uncertainty in this style of dating. The use of quarter and half subdivisions of centuries AD was also occasional practice observed in the datasets:

- 1^{st} Half = 01→50
- 2^{nd} Half = 51→100
- 1^{st} Quarter = 01→25
- 2^{nd} Quarter = 26→50
- 3^{rd} Quarter = 51→75
- 4^{th} Quarter = 76→100

The various subdivisions described were not represented in the original *Timelines* thesaurus so terms were manually added for each century AD (i.e. *Early 1^{st} Century, 1^{st} quarter 1^{st} century AD*, *1^{st} half 1^{st} century AD* etc.).

3.2 Aligning Data Records with Known Time Periods

Records containing date information were first semi-automatically processed to give 2 numeric values representing the approximate lower and upper bounds of the time periods indicated by the data. This process involved some data cleansing and the identification of common textual patterns (e.g. "MLC2-C3", "AD341-6" etc.) in fields describing periods. The resultant records were next processed to assign known time period identifiers to each record. This allows clustering and searching for records, and also facilitates matching between periods mentioned in database records and within the grey literature documents. A semantic closeness calculation for time periods used in previous work at Glamorgan (described more fully in [24]) was incorporated into a custom application (STAR.TIMELINE) to batch process the cleansed data records, comparing the derived start/end dates against our collated list of known periods. Periods frequently overlapped or were contained within others, so the matching method accommodated these issues to suggest the most appropriate match. The matching calculation used is reproduced below:

$$\text{Match}(P1, P2) = W1\,(MP / IU) + W2\,(IU / (NM + IU)) + W3\,(IU / (D + IU)).$$

- P1 & P2 are the periods being compared.
- D is the time elapsed between one period ending and another starting (expressed in years). Where the two periods overlap D will be 0.
- MP is the matching portion (overlap) between two periods – the number of years that the two periods have in common.
- NM is the non-matching portion between two periods – the number of years that the two periods *do not* have in common.
- IU is the duration in years of the period being used as the basis for the comparison.
- W1, W2 & W3 represent weightings for the appropriate factors. Following initial experimentation these weightings were set to 0.400, 0.200 and 0.400 respectively, resulting in a match value that is always within the range (0..1). Modification of these weights relative to each other could give higher precedence to e.g. overlapping terms.

Table 2. Calculation of the degree of match between periods P1 and P2

P1	P2	Relationship Type	D	MP	NM	IU	Match
0→150	200→300	P1 *occurs before* P2	50	0	250	150	0.375
0→150	150→250	P1 *meets* P2	0	0	250	150	0.475
0→150	100→200	P1*overlaps* P2	0	50	200	150	0.619
0→150	50→150	P1 *includes* P2	0	100	50	150	0.817

The calculated match value is then used for ranking the results in decreasing order of match. An optional minimum match threshold can also be set to prevent results with a lower degree of match being returned.

As a practical example, comparing a period P1 [175→190] to the concepts from the Timelines thesaurus yields the (top 10) results as shown in Table 3.

The relationship types are calculated as previously described in Fig. 2, using the start and end dates of the periods being compared. All of the relationships listed in

Table 3. Calculation of top 10 closest matches for P1 [175→190]

P2	Relationship Type	D	MP	NM	IU	Match
LATE 2ND CENTURY [167→200]	P1 *occurs during* P2	0	15	18	15	0.891
4TH QUARTER 2ND CENTURY AD [176→200]	P1 *overlapped by* P2	0	14	11	15	0.889
2ND HALF 2ND CENTURY AD [151→200]	P1 *occurs during* P2	0	15	34	15	0.861
2ND CENTURY AD [101→200]	P1 *occurs during* P2	0	15	84	15	0.830
ROMAN [43→410]	P1 *occurs during* P2	0	15	352	15	0.808
COMMODUS [180→192]	P1 *overlapped by* P2	0	10	7	15	0.803
AURELIUS [161→180]	P1 *overlaps* P2	0	5	24	15	0.610
3RD QUARTER 2ND CENTURY AD [151→175]	P1 *met by* P2	0	0	39	15	0.456
PERTINAX [193→193]	P1 *occurs before* P2	3	0	16	15	0.430
DIDIUS JULIANUS [193→193]	P1 *occurs before* P2	3	0	16	15	0.430

Table 3 are factually correct, however to reduce the potential number of new assertions made; only the maximal match for each relationship type is retained. E.g. "P1 *occurs during* LATE 2nd CENTURY" would render the subsequent *occurs during* relationships superfluous.

The STAR.TIMELINE process was run against data records extracted from a number of tables in the archaeological datasets to output the closest known period matches. A small selection of processed results when aligned with the English Heritage *Timelines* thesaurus, listing only maximal matches for each relationship type and applying a minimum match threshold of 0.500, are shown in Table 4.

Table 4. Sample of data aligned with English Heritage Timelines Thesaurus

Data Record	Relationship Type	Period Matched	Match
"AD 69-79"	*equals*	VESPASIAN [69→79]	1.000
	occurs during	LATE 1st CENTURY [67→100]	0.861
	overlapped by	3rd QUARTER 1st CENTURY AD [51→75]	0.703
	met by	OTHO [69→69]	0.545
"AD 270-4"	*equals*	TETRICUS I [270→274]	1.000
	starts	AURELIAN [270→275]	0.960
	occurs during	3rd QUARTER 3rd CENTURY AD [251→275]	0.833
	met by	QUINTILLUS [270→270]	0.614
	finished by	TETRICUS II [274→274]	0.614
"AD 275-402"	*includes*	4TH CENTURY AD [301→400]	0.876
	occurs during	ROMAN [43→410]	0.869
	overlapped by	LATE 3rd CENTURY [267→300]	0.586
	started by	TACITUS [275→276]	0.504
"AD 268-70"	*equals*	CLAUDIUS II GOTHICUS [268→270]	1.000
	occurs during	3rd QUARTER 3rd CENTURY AD [251→275]	0.817
	met by	LUCIUS AELIANUS [268→268]	0.733
	includes	MARCUS AURELIUS MARIUS [269→269]	0.733
	finished by	VICTORINUS [269→270]	0.733
	overlapped by	POSTUMUS [260→269]	0.636
"AD 270-84"	*occurs during*	LATE 3RD CENTURY [267→300]	0.885
	overlaps	4th QUARTER 3rd CENTURY AD [276→300]	0.706
	includes	PROBUS [276→282]	0.699
	started by	AURELIAN [270→275]	0.665
	overlapped by	3rd QUARTER 3rd CENTURY AD [251→275]	0.610
	met by	QUINTILLUS [270→270]	0.532

As a result of this process we created a set of records originating from multiple datasets that could be effectively cross searched either directly by date or via thesaurus concept. It was noted that in certain cases the alignment process rediscovered links to specific Roman emperors that had previously only been implicit in the dates used (see e.g. *VESPASIAN*, *TETRICUS I* and *CLAUDIUS II GOTHICUS* in Table 4).

The matched thesaurus terms were limited specifically to the geographic and cultural scope of the associated thesaurus used; however the STAR.TIMELINE application may be loaded with alternative period lists to align the data with periods specific to other locations. The overall quantity of records processed for date alignment is shown in Table 5.

Table 5. Overall quantity of records processed for date alignment

Table.Column	Records processed
LEAP.FINDS	2,719
MOLAS.FND_DATE	1,834
MOLAS.FINDS_INVENTORIES	5,229
MOLAS.FND_PRPOT	1,674
MOLAS.FND_RF	1,814
RRAD.CONTEXT_PERIOD	5,291
RRAD.OBJECT_PERIOD	3,765
RRAD.CERAMICS_PERIOD	5,401
RRAD.SAMPLE_PERIOD	369
RPRE.OBJECTS	4,828
Total records processed	32,924

In the next stage of the work the processed data will be converted to RDF conforming to the CRM model for representing period information as shown in Fig. 5, for import to the main STAR data store.

This import will facilitate querying the aligned records using SPARQL in the context of other extracted STAR data, and manipulation using existing STAR interface controls, e.g. the experimental timeline component illustrated in Fig. 6.

```
@prefix crm: <http://cidoc.ics.forth.gr/rdfs/cidoc_v4.2.rdfs#>
@prefix thes: <http://tempuri/star/concept#>
@prefix : <http://tempuri/star/base#>

# Supplement existing SKOS Concept ('LATE 3rd CENTURY') with CRM
# properties describing a known period relative to other periods.
thes:135952 a crm:E4.Period;
  crm:P115F.finishes thes:900086;
  crm:P117F.occurs_during thes:134738;
  crm:P118F.overlaps_in_time_with thes:136180;
  crm:P4F.has_time-span [
    crm:P81F.ongoing_throughout '+267/+300'
  ] .

# Data record related to known periods using CRM relationships.
<dataX> a crm:E4.Period;
  crm:P117F.occurs_during thes:135952;
  crm:P118F.overlaps_in_time_with thes:900011;
  crm:P117B.includes thes:136172;
  crm:P116B.is_started_by thes:136162;
  crm:P118B.is_overlapped_in_time_by thes:900010;
  crm:P119B.is_met_in_time_by thes:136160;
  crm:P4F.has_time-span [
    crm:P82F.at_some_time_within '+270/+284'
  ] .
# [etc.]
```

Fig. 5. Processed data expressed as a series of CRM *E4.Period* entities

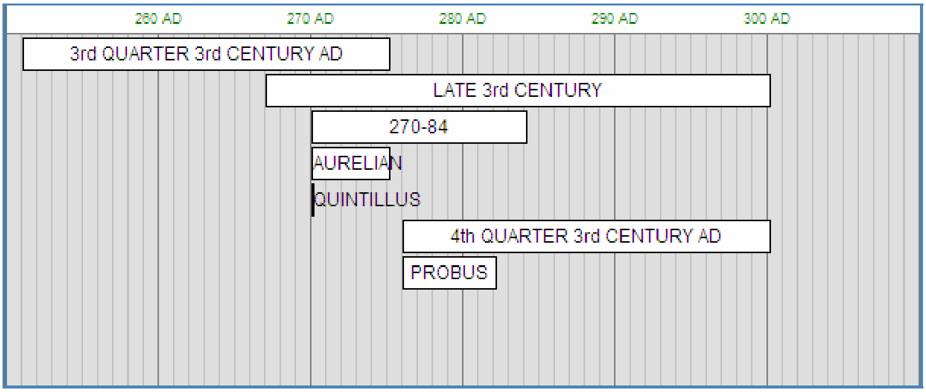

Fig. 6. Experimental STAR timeline interface component

4 Conclusions

This paper gave an overview of the STAR project and went on to discuss a particular aspect of the project in dealing with chronological information. We described suitable methods of modelling temporal relationships and demonstrated the use of CRM entities and properties to supplement existing controlled vocabularies, enabling temporal reasoning. We then described STAR.TIMELINE; a custom application that aligns archaeological records with a controlled set of known time periods, in the process rediscovering links to specific named periods that were only implicit in the original data. We acknowledge that archaeological dating can be uncertain while relative ordering may be better agreed and established. The processed records can be searched either by date range or by named period, and they facilitate temporal matching between database records and grey literature document content.

This work has potential uses beyond the immediate project, and we envisage reusing the implemented STAR.TIMELINE functionality to complement our existing suite of web services with a service for suggesting a ranked list of suitable named archaeological periods for any given dates, possibly making use of the distinction between period types e.g. century names, British Monarchs, Roman Emperors and extending to give further consideration to periods being specific to particular geographical locations.

Acknowledgements

The STAR project is funded by the UK Arts and Humanities Research Council (AHRC). Thanks are due to Phil Carlisle & Keith May (English Heritage), Douglas Tudhope and Andreas Vlachidis (University of Glamorgan), Renato Souza (Universidade Federal de Minas Gerais, Brazil).

References

1. STAR: Semantic Technologies for Archaeological Resources, http://hypermedia.research.glam.ac.uk/kos/star
2. Patel, M., Koch, T., Doerr, M., Tsinaraki, C.: Report on Semantic Interoperability in Digital Library Systems. DELOS Network of Excellence, WP5 Deliverable D5.3.1 (2005)
3. CIDOC Conceptual Reference Model (CRM), http://cidoc.ics.forth.gr
4. Cripps, P., Greenhalgh, A., Fellows, D., May, K., Robinson, D.: Ontological Modelling of the work of the Centre for Archaeology, CIDOC CRM Technical Paper (2004), http://cidoc.ics.forth.gr/technical_papers.html
5. CRM-EH: English Heritage Extension to CRM for the archaeology domain, http://hypermedia.research.glam.ac.uk/kos/CRM/
6. RDFS Encoding of the CIDOC CRM, http://cidoc.ics.forth.gr/rdfs/cidoc_v4.2.rdfs
7. Binding, C., Tudhope, D., May, K.: Semantic Interoperability in Archaeological Datasets: Data Mapping and Extraction via the CIDOC CRM. In: Christensen-Dalsgaard, B., Castelli, D., Ammitzbøll Jurik, B., Lippincott, J. (eds.) ECDL 2008. LNCS, vol. 5173, pp. 280–290. Springer, Heidelberg (2008)
8. SKOS: Simple Knowledge Organization Systems - W3C Semantic Web Deployment Working Group, http://www.w3.org/2004/02/skos
9. Tudhope, D., Binding, C., May, K.: Semantic interoperability issues from a case study in archaeology. In: Kollias, S., Cousins, J.(ed.) Semantic Interoperability in the European Digital Library, Proceedings of the First International Workshop (SIEDL) 2008, associated with 5th European Semantic Web Conference, Tenerife, pp. 88–99 (2008)
10. Isaac, A., Matthezing, H., van der Meij, L., Schlobach, S., Wang, S., Zinn, C.: Putting ontology alignment in context: Usage scenarios, deployment and evaluation in a library case. In: Bechhofer, S., Hauswirth, M., Hoffmann, J., Koubarakis, M. (eds.) ESWC 2008. LNCS, vol. 5021, pp. 402–417. Springer, Heidelberg (2008)
11. Tuominen, J., Frosterus, M., Kim Viljanen, K., Eero Hyvönen, E.: ONKI SKOS Server for Publishing and Utilizing SKOS Vocabularies and Ontologies as Services. In: Aroyo, L., Traverso, P., Ciravegna, F., Cimiano, P., Heath, T., Hyvönen, E., Mizoguchi, R., Oren, E., Sabou, M., Simperl, E. (eds.) ESWC 2009. LNCS, vol. 5554, pp. 768–780. Springer, Heidelberg (2009)
12. Terminology Services for the STAR Project, http://hypermedia.research.glam.ac.uk/resources/terminology/
13. SKOS API: SWAD EUROPE Thesaurus Project Output (2004), http://www.w3.org/2001/sw/Europe/reports/thes/skosapi.html
14. Binding, C., Tudhope, D.: KOS at your Service: Programmatic Access to Knowledge Organisation Systems. Journal of Digital Information 4(4) (2004), http://journals.tdl.org/jodi/article/view/110/109
15. Binding, C., Tudhope, D.: Terminology Services. Knowledge Organization, vol. 37. Ergon-Verlag (2010) (forthcoming) ISSN 0943-7444
16. Getty vocabulary databases, Paul Getty Trust, J.: http://www.getty.edu/research/tools/vocabulary/
17. Knight, B., Ma, J.: Time Representation: A Taxonomy of Temporal Models. Artificial Intelligence Review 7, 401–419 (1994)
18. Allen, J.F.: Maintaining knowledge about temporal intervals. Communications of the ACM 26 (1983)
19. OWL-Time, http://www.w3.org/TR/owl-time/

20. Doerr, M.: Semantic problems of thesaurus mapping. Journal of Digital Information 1(8) (2001), http://journals.tdl.org/jodi/article/view/31/32
21. Eide, O., Holmen, J., Ore, C.: Deducing event chronology in an archaeological documentation system. In: Proceedings Computer Applications and Quantitative Methods in Archaeology (CAA 2009), Williamsburg (2009)
22. Doerr, M., Plexousakis, D., Kopaka, K., Bekiari, C.: Supporting Chronological Reasoning in Archaeology. In: Proceedings Computer Applications and Quantitative Methods in Archaeology (CAA 2004), Prato (2004)
23. English Heritage Timelines Thesaurus, http://www.fish-forum.info/i_time.htm
24. Tudhope, D., Taylor, C.: Navigation via Similarity: automatic linking based on semantic closeness. Information Processing and Management 33(2), 233–242 (1997)