Linked Data for the Historic Environment

Ceri Binding, Tim Evans, Jo Gilham, Douglas Tudhope and Holly Wright

This article discusses the outcomes of research undertaken by the Hypermedia Research Group at the University of South Wales in collaboration with the OASIS team at the Archaeology Data Service in the *Linked Data for the Historic Environment* (LD4HE) project. The new OASIS system allows stakeholders to record information/events accurately and consistently, using established heritage and spatial vocabularies, including those currently available via the Heritage Data platform. The LD4HE project explores one avenue of enhancing the potential re-use of information recorded by OASIS and making connections with other online data collections. LD4HE enables the creation and export of RDF from the new OASIS V system, a major step towards the production of Linked Data. The main outcomes comprise a conceptual mapping between mandatory OASIS fields and the standard CIDOC-CRM ontology, together with a template-based tool to convert records exported from OASIS to a corresponding RDF representation. A set of SPARQL queries demonstrates the outcomes of the data conversion. New specialised vocabularies required by OASIS have been published as linked data on the Heritage Data platform. The methods are described and illustrated with examples. Reflections on the case study and cost/benefit considerations for Linked Data conversion are discussed, together with possible strategies for reducing the costs of producing Linked Data.

1. Introduction

This article discusses the outcomes of research undertaken by the Hypermedia Research Group (HRG) at the University of South Wales, in collaboration with the OASIS team at the Archaeology Data Service (ADS), in the Linked Data for the Historic Environment (LD4HE) project. The aim of the project was to investigate the creation of RDF from exports of the new OASIS V system. The OASIS project dates to 1999, when a consortium from ADS and English Heritage (now Historic England) developed a vision for an online index of fieldwork events and their unpublished reports that could be updated by the data producers, and thereafter fed back into the regional Historic Environment Records (HERs) and national heritage organisations. The first pilot form was launched in 2001, and has been updated ever since to reflect the changing needs of its users. At the time of writing, the form is used extensively in England and Scotland – where it encourages the continued reporting of fieldwork to the wider public through Archaeology Scotland's annual publication *Discovery and Excavation in Scotland* (DES) – and the maritime zone of Wales. During that period, over 94,000 records have been collected and over 50,000 unpublished reports have been made publicly available in the Open Access ADS Library (for further discussion of the use of archaeological grey literature, see Evans 2015). The ADS Library record contains a link to the final report, a Digital Object Identifier for citation, and a subset of the original OASIS metadata. It is important to highlight the difference between the contents of OASIS and the ADS Library. The latter is based on bibliographic metadata and resource discovery, while the former also contains rich metadata on why the project took place, with heritage-specific methodologies described.

The LD4HE project originated during the most recent OASIS redevelopment project, with aims that included the potential for innovation. During consultation for the specification for the new OASIS form, it was noted that some users of OASIS and the ADS Library would like to query OASIS data, such as details about the type of event, the rationale, the results and the literary outputs, for example as a means to producing business-level data about the projects and interventions recorded within the system. LD4HE explores one avenue of enhancing the potential re-use of information recorded by OASIS outside of traditional channels. Conversion to RDF is a major step in the production of Linked Data, which would open possibilities for connecting with other collections similarly made available online.

1.1 Related work

LD4HE utilises technical standards used by various projects and initiatives within cultural heritage (see for example the ARIADNE archaeological infrastructure, Aloia *et al.* 2017). Through mapping to the standard CIDOC CRM (Conceptual Reference Model) and

use of national archaeological vocabularies, OASIS will be aligned with international developments for the creation and potential re-use of data, such as FAIR Data (Wilkinson *et al*. 2016), the ongoing ARIADNEplus H2020 archaeological infrastructure project and the Arches platform for management of heritage inventories (Myers *et al*. 2016).

Linked Data (Bizer *et al*. 2009) offers a set of standards and technologies for making data of all kinds available online to encourage connection and re-use. This includes archaeology and heritage datasets and reports. For example, LD4HE builds on a previous collaboration (Binding *et al*. 2015) that investigated the conversion of datasets from the ADS Repository (including the Channel Tunnel Rail Link and the Aggregates Levy Sustainability Fund) to ADS Linked Data. Central to many of these projects are Linked Data publication of standard 'value vocabularies' (such as authority lists, thesauri and classifications – see Isaac *et al*. 2011) that serve as hubs in a linked data web by offering definitive versions of standard vocabularies with persistent identifiers (i.e. URIs). The Pelagios Linked Data initiative (Isaksen *et al*. 2014) has developed tools to facilitate the connection of web resources based on location (places in the ancient world, drawing on the Pleiades gazetteer). Facilitating temporal connections, the PeriodO platform allows the open publication of collections of named periods with persistent identifiers for defined geographical areas (Shaw *et al*. 2016). The Nomisma Linked Data initiative provides a set of standard persistent identifiers for numismatic concepts that allow the inter-connection of data relating to ancient monetary objects (see Gruber 2016 for background). Drawing on experience with the Open Context repository of archaeological datasets, Kansa and Kansa (2013) promote the routine web publication of archaeological datasets and documentation, richly interconnected via common concepts. For European heritage data, the Europeana project has devoted considerable effort to enriching object metadata with linked data vocabularies, thus offering a multilingual capability (see for example, Charles *et al*. 2014). Binding *et al*. (2015) discuss how the enrichment of ARIADNE partner data with the Getty Art and Architecture Thesaurus allowed a multilingual capability based on that vocabulary hub.

Archaeological archives, such as OASIS and the ADS Library, are increasingly turning to principles for long-term preservation and well-structured metadata, such as the FAIR Data initiative. Standard vocabularies with persistent unique identifiers play an important role as hubs in the web of data connecting archives. The new OASIS V system has been designed with a high level of semantic interoperability in mind, primarily to facilitate communication between systems (e.g. HERs). This interoperability is achieved via the use of cultural heritage thesauri and vocabularies made available as Linked Open Data (LOD), including those available via the HeritageData platform through a previous collaborative project (SENESCHAL) that produced SKOS standard versions of national heritage vocabularies (Binding and Tudhope 2016). As part of the work for LD4HE, new specialised vocabularies required by OASIS V have been published on the HeritageData platform. These were based on existing wordlists, which were transformed into SKOS representation and Linked Data. The new vocabularies are shown in Table 1.

Table 1: New OASIS vocabularies created as Linked Open Data

Scheme	Description
OASIS Funder	Funder
OASIS Associated ID	Groups together IDs specific to the Historic Environment within Northern Ireland
OASIS Development Type	Development type
OASIS Paper and Digital Archive Component	Paper and digital archive component
OASIS Protection Status	Protection status
OASIS Reason for Investigation	Reason for investigation

The possibility of OASIS Linked Data would enhance opportunities for inter-connection of OASIS content with international archaeological datasets and reports and allow greater possibilities for re-use.

2. Data Fields

The source data for the LD4HE project is a specific subset of fields originating from the overall OASIS dataset. These are mandatory fields recording details about the type of event/intervention, the rationale for the activity, the results and the literary outputs resulting from the work.

Conceptual mappings to the CIDOC CRM were designed from a set of sample records from ADS plus a spreadsheet indicating the positions of the fields of interest within the JSON structure of the test data files. The sample records consisted of two JSON format text files to indicate the contrast in the level of detail contained in different OASIS records. Although the location of the mandatory fields within the JSON test data structures had been communicated via an informal path syntax, it was necessary to explicitly define these paths in a machine-readable and programmatically actionable format. Each required field was identified within the sample data structure and a JSONPath expression was derived to define the location precisely. These derived paths were then tested against the example data using the JSONPath online evaluator application to ensure correctness.

Table 2 shows the subset of OASIS fields used for the project (more information on the data fields is provided in the LD4HE GitHub repository described in Section 3).

Table 2: OASIS V fields used as test data

OASIS Field	Field name	Brief field description
Field 1	OASIS ID	Contains the unique identifier for the OASIS record. There is one OASIS identifier per record.
Field 2	Event type	The type of investigation activity undertaken. There may be multiple event types per record.
Field 3	Reason for investigation	The reason for undertaking an investigation.
Field 4	Country	Location of a site.

Field 5	Site name	Free text local name for a site.
Field 9a	Grid reference (geom_ngr)	Location of a site. This stores either a point, line or polygon in a single geometry field (using PostGIS), using the OSGB36 crs.
Field 9b	Grid reference (geom_ll)	Location of a site. This stores either a point, line or polygon in a single geometry field (using PostGIS), using the WGS84 crs.
Field 15	County	Location that a site falls within.
Field 16	District	Location that a site falls within.
Field 17	Parish	Location that a site falls within.
Field 18	HER	Name of the Historic Environment Record organisation responsible for an area encompassing the location of a site.
Field 19	National body	National body as an organisation responsible for the area encompassing the location of a site.
Field 22	Project title	Descriptive title of the investigation.
Field 23a	Start date	Start of the overall timespan for an investigation.
Field 23b	End date	End of the overall timespan for an investigation.

Field 24b	Description	Brief textual description of an investigation.
Field 28	Planning application ID	Planning Identifier associated with an investigation.
Field 35	Publication type	The form of publication the report takes.
Field 36	Title of report	Textual report title.
Field 39	Author/editor	Author/editor of the report.
Field 44	Report date	Year the report was published/issued.
Field 45	Publisher	Name of the organisation responsible for publishing the report.
Field 46	Place of issue	Place of publication for the report.
Field 50	URL	URL of the report document - an identifier.
Field 50a	DOI	Digital Object Identifier of the report - an identifier.
Field 58	Name of organisation	Name of the organisation who undertook the work.

Field 62	Monument type	URIs from monument type vocabulary corresponding to site location: [England](#)/[Scotland](#)/[Wales](#).
Field 63	Monument period	URIs from period vocabulary corresponding to site location: [England](#)/[Scotland](#)/[Wales](#).
Field 64	Artefact type	URIs from object types vocabulary according to site location: [England & Wales](#)/[Scotland](#).
Field 65	Artefact period	URIs from period vocabulary corresponding to site location: [England](#)/[Scotland](#)/[Wales](#).
Field 70	Research outcomes	Uses selection from Research Frameworks list (non-LOD at time of writing).

3. Mapping from OASIS to the CIDOC CRM Ontology

The LD4HE data model is based on a subset of the [CIDOC Conceptual Reference Model](#) (CRM). A [LD4HE GitHub Repository](#) was created to post the outputs of the data mapping exercise for referencing, communication and discussion. An initial set of modular interconnecting data patterns was produced and uploaded to this platform. These online data patterns consist of a short description, a diagram illustrating the modelled entities and the relationships between them, and a practical example (expressed in a number of RDF serialisation formats) of instances of data conforming to the model. These patterns were then further refined and extended, following a review of issues raised and feedback from the project team. Figure 1 gives an indication of how these aspects are interlinked in the overall model. The dataset is composed of records referring to investigations. Investigations are carried out by organisations at sites during timespans. Report production describes reports that document the investigations.
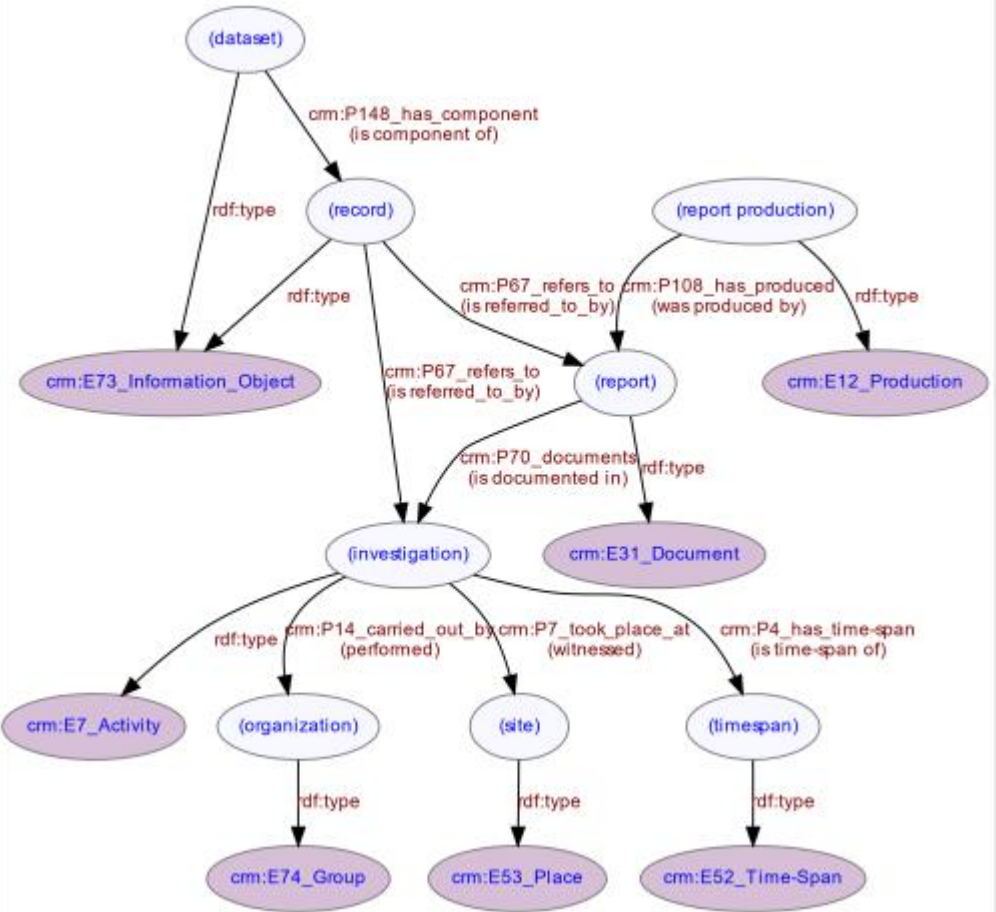
Figure 1: Main entities and properties

Each record is a component of the full dataset. Records (see Figure 2) refer to investigations, reports, HERs, national bodies, artefact types and monument types. Records may have both local identifiers and Digital Object Identifiers (DOIs - which are also used to identify reports).
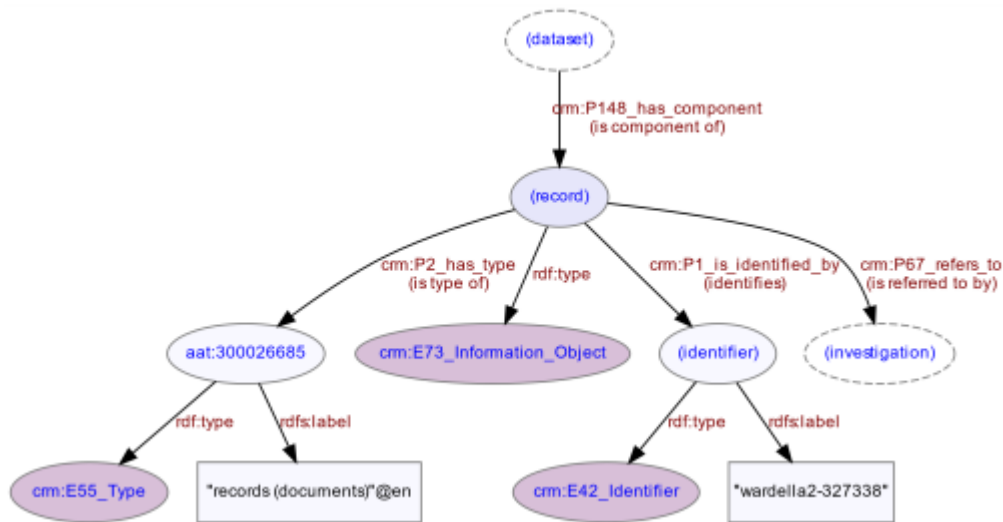
Figure 2: Record with associated entities and properties

Figure 3 illustrates one particular detail of the model relating to the use of standard HeritageData vocabularies that are themselves Linked Data. Monuments are larger immovable structures identified as being of potential interest to archaeological activities. Specific instances of monuments are not included in OASIS metadata records - instead the records refer to general monument types (and monument periods). Monument types are concepts originating from the monument types thesaurus corresponding to the location of the site. For England this will be the FISH Thesaurus of Monument Types, for Scotland it will be the Monument Type Thesaurus (Scotland) and for Wales it will be the MONUMENT TYPE (WALES) thesaurus. Where records refer to monument periods, these will be concepts originating from the appropriate periods list corresponding to the location of the site. For England this will be the Historic England Periods List, for Scotland it will be (potentially) ScAPA: Scottish Archaeological Periods & Ages and for Wales it will be PERIOD (WALES). There is no direct connection between the monument type and the period concept in the dataset record. However, reference to the HeritageData vocabularies allows a controlled, concept-based search on terms such as 'Early Medieval' or 'Lime Kiln' rather than a literal string search that might not take account of alternate terms. Currently the different national UK vocabularies are not inter-connected so concept-based search is not possible across the different national vocabularies. Section 6.3 discusses possible future work that would map corresponding concepts from the national vocabularies together to enable concept-based search across English, Scottish and Welsh OASIS data.
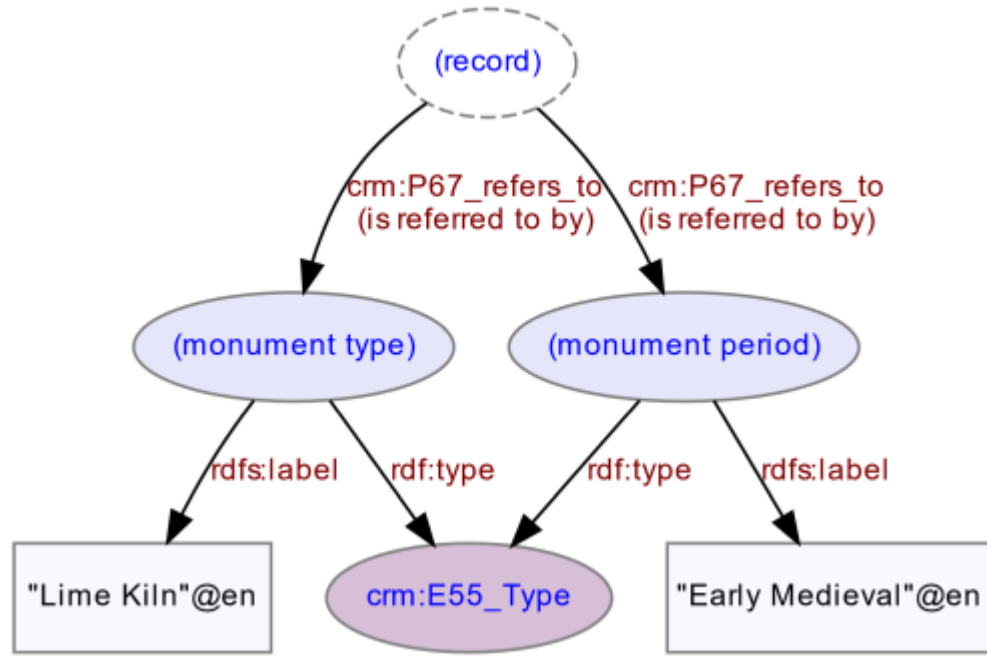
Figure 3: Monument type and period associated with a site record

For details of the other elements of the model, see the GitHub Repository.

4. Data Conversion

Having designed the mapping to the CIDOC CRM, the next stage was to convert the OASIS export data to RDF representation, according to the conceptual mapping. A template-based data conversion method was employed using the STELETO tool (Binding *et al*. 2019). STELETO is a refinement of methods developed in a previous collaboration between HRG and ADS in the STELLAR project that enabled the ADS to create Linked Data versions of archaeological datasets (for details, see Binding *et al*. 2015). Templates define a conversion between the relevant elements of source datasets and the output data model underlying any particular template. In general, STELETO can produce output in any format (e.g. XML or plain text) if a suitable template is

provided. For LD4HE, a STELETO template was designed that transformed the data format of the OASIS export to RDF, conforming to the CIDOC CRM ontology mapping described in [Section 3](#).

Data conversion proceeded in stages - creating the conversion template, running the conversion on test datasets exported from OASIS by ADS, checking and validating the outputs, then undertaking iterative refinements by updating the template as necessary and re-running the conversion on successively larger datasets. Thus, a STELETO template was produced and tested against an initial test JSON dataset from ADS, producing NTriples RDF as output. RDF inverse properties were generated by the template to allow querying properties in either direction without requiring the employment of more advanced reasoning mechanisms. The resulting RDF was imported to a triple store, both to validate the output and to formulate some example SPARQL queries based on the data model produced. A larger test JSON dataset exported from OASIS contained 5000 records, including some test records encountered from the first dataset. The data conversion template was applied to this new dataset, producing 735,003 RDF triples in NTriples format output (with a few duplicated statements). The output was imported to an RDF triple store for further analysis. Following review of the outcomes of the conversion process, a final test JSON dataset exported by ADS included DOI numbers for reports. The template was adjusted to account for the additional field and the conversion was re-run, producing 756,543 RDF triples.

4.1 URI scheme

The creation of URI unique identifiers is an important element of the design of the data conversion process. Wherever possible, any existing URIs present in the input dataset were used as identifiers in the output. In addition, a consistent URI scheme was required to uniquely identify all entities in the data model. Since the project was investigating (but not publishing) Linked Data, a temporary project-specific dataset URI prefix was used in all generated entity URIs ("http://tempuri/ld4he/oasis/") - this can be revised later by simple replacement if required. The dataset prefix had (singular) entity types and identifiers appended as appropriate to create unique URIs for each entity modelled to simulate a suitable REST URI scheme. In cases where the generated URIs were required to incorporate data values, these were trimmed of any extraneous white space and converted to lower case for consistency, then URI-encoded to ensure a valid URI was produced:

```
Records: {dataset}/record/{record id}
Record identifiers: {dataset}/id/{record id}
Sites: {dataset}/site/{site id}
Investigations: {dataset}/investigation/{record id}
Investigation titles: {dataset}/investigation/{record id}/title
```

```
Places: {dataset}/place/{country}/{county}/{district}/{parish}
People: {dataset}/person/{name}
Organisations: {dataset}/organisation/{name}
Reports: {dataset}/report/{record id}/{report title}
Timespans: {dataset}/timespan/{year}
```

To avoid possible ambiguity where place names were involved, a hierarchical URI scheme was adopted by appending the country, county, district, and parish name values to produce URIs that, although readable, could potentially be lengthy, e.g.

```
http://tempuri/ld4he/oasis/place/england/greater+london/enfield+london+boro/enfield%2C+unparished+area
```

In future when place name data contains Linked Data URIs then direct references to [Ordnance Survey Boundary Line](#) Linked Open Data resources can be substituted for place names.

4.2 OASIS modelling

The link between people and organisations is not made explicitly in the source data (though they are present in the same list). It is also not possible to distinguish two people having the same name as there is no identifier or additional qualifying metadata present, for example:

```
"oasisProjPeopleList": [
{
"forename": "Fred",
"surname": "Bloggs"
},
{
"organisation": "English Heritage Architectural Survey"
},
{
"forename": "Fred",
"surname": "Bloggs"
}
]
```

They have therefore been listed as being associated with an investigation, without making a specific direct link between the person and the organisation. There remains no (simple) way to determine whether the two people in the example are necessarily the same person.

The original data mapping included DOI and URL fields for unique identification of reports. However, these fields were not present in the JSON data received. This necessitated a change to the original model to cater for the possibility that these identifiers may not be present. The report title seemed the most consistently present element, so although it produced a long URI it was used in combination with the investigation URI to create a unique identifier for each report.

The original modelling regarded sites as places that fell within parishes/counties/countries. While this approach was logically correct, it did not allow any subsequent distinction between sites and parishes when searching for 'places' falling within particular counties. Therefore an extra triple was added to distinguish the sites, by declaring crm:P2_has_type aat:300000809 ("sites (locations)")

Report publication dates in the input dataset were numeric years, so could be simply represented as xsd:gYear values by the template. However investigation dates had a different string format "dd-mmm-yyyy hh:mm" e.g. "01-Sep-2006 12:00". These values need to be converted to xsd:dateTime values (format "yyyy-mm-ddThh:mmZ") for date comparisons to work correctly within the RDF/SPARQL environment. Template functionality in this regard is very limited, so this necessitated a change to the STELETO application to include an additional custom 'text filter' that would do the required conversion and formatting.

The template approach results in occasional duplication of triples. These duplicates can be removed by importing the resultant RDF output to a triple store and then re-exporting. This also assists with validation as the triple store import process will flag up any potential issues.

Site location coordinates are present as "Well Known Text" (WKT) point or polygon strings in the source data using two fields representing two coordinate systems, OSGB36 and WGS84. For example:

```
"geomNgrOut" : "POINT(496000.000865923 201599.999520326)"
"geomLlOut" : "POINT(-0.612137804625569 51.704937810468)"
```

Owing to uncertainty concerning how to represent and distinguish these two coordinate systems in the RDF triples, only the "geomNgrOut" field is currently used by the conversion template to output coordinate data, rather than having a mixture of

coordinate types in the output. This decision can be revisited, and the template adjusted if the desired output format is further clarified.

5. SPARQL Queries on RDF Resulting from Data Conversion

Once the OASIS data had been converted to RDF, a series of SPARQL queries were created, in order to test the data conversion, give an overview of the integrated OASIS data and illustrate potential search strategies over the resulting integrated linked dataset. The following queries were run on the RDF conversion of the final JSON dataset from ADS of 5000 OASIS records.

5.1 Overall entity counts

Useful for testing purposes but also gives an overview of the range of entities in the dataset.

```
SELECT DISTINCT ?entityType (count(?entityType) AS ?counter)
WHERE {
?s a ?entityType .
}
GROUP BY ?entityType
ORDER BY DESC(?counter)
```

entityType	counter
crm:E21_Person	15,920
crm:E42_Identifier	15,536

crm:E53_Place	14,896
crm:E41_Appellation	11,087
crm:E7_Activity	10,035
crm:E35_Title	10,035
crm:E74_Group	7,419
crm:E31_Document	5,035
crm:E12_Production	5,035
crm:E73_Information Object	5,001
crm:E55_Type	2,812
Total	**102,811**

5.2 Overall property counts

Useful for testing purposes and gives an overview of the range of properties from the modelling.

```
SELECT DISTINCT ?property (count(?property) AS ?counter)
WHERE {
```

```
?subject ?property ?entity .
}
GROUP BY ?property
ORDER BY DESC(?counter)
```

property	counter
rdf:type	102,857
crm:P1i_identifies	48,889
crm:P1_is_identified_by	48,889
rdfs:label	45,211
crm:P14_carried_out_by	26,432
crm:P14i_performed	26,432
crm:P89_falls within	24,093
crm:P89i_contains	24,093
crm:P67_refers_to	17,406
crm:P67i_is_referred_to_by	17,406

crm:P2_has_type	15,864
crm:P9_consists_of	10,070
crm:P9i_forms_part_of	10,070
crm:P7i_witnessed	10,054
crm:P7_took_place_at	10,054
crm:P4i_is_time-span_of	10,038
crm:P4_has_time-span	10,038
crm:P81b_begin_of_the_end	5,606
crm:P81a_end_of_the_begin	5,606
crm:P82b_end_of_the_end	5,606
crm:P82a_begin_of_the_begin	5,606
crm:P168_place_is_defined_by	5,220
crm:P87_is_identified_by	5,037

crm:P87i_identifies	5,037
crm:P70i_is_documented_by	5,035
crm:P108_has_produced	5,035
crm:P16_used_specific_object	5,035
crm:P108i_was_produced_by	5,035
crm:P102i_is_title_of	5,035
crm:P16i_was_used_for	5,035
crm:P102_has_title	5,035
crm:P70_documents	5,035
crm:P148i_is_component_of	5,000
crm:P148_has_component	5,000
crm:P3_has_note	4,907
crm:P2i_is_type_of	3,643

Total	554,444

Inverse properties generated by the template are evident in this table e.g. crm:P87_is_identified_by / crm:P87i_identifies

5.3 List 10 sites and their locations

Useful for testing site data conversion and an example low-level spatial query.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX crm: <http://www.cidoc-crm.org/cidoc-crm/>
PREFIX aat: <http://vocab.getty.edu/aat/>
PREFIX os: <http://data.ordnancesurvey.co.uk/ontology/admingeo/>

SELECT DISTINCT ?siteName ?parishName ?districtName ?countyName ?countryName
WHERE {
?site crm:P2_has_type aat:300000809;
crm:P1_is_identified_by [rdfs:label ?siteName] .
OPTIONAL {
?site crm:P89_falls_within [
crm:P2_has_type os:CivilParish ; rdfs:label ?parishName] .
}
OPTIONAL {
?site crm:P89_falls_within [
crm:P2_has_type os:District; rdfs:label ?districtName ] .
}
OPTIONAL {
?site crm:P89_falls_within [
crm:P2_has_type os:County; rdfs:label ?countyName] .
}
OPTIONAL {
?site crm:P89_falls_within [
crm:P2_has_type os:EuropeanRegion; rdfs:label ?countryName] .
}
```

```
}
LIMIT 10
```

siteName	parishName	districtName	countyName	countryName
Land north of St Leonard's Church	Leverington	Fenland	Cambridgeshire	England
Land south of Wragby Road, Lincoln, Lincolnshire	Lincoln, unparished area	Lincoln	Lincolnshire	England
south of Skirbeck Road	Boston, unparished area	Boston	Lincolnshire	England
Little Moreton Hall (overflow car park)	Odd Rode	Cheshire East	Cheshire	England
Kirby Bellars Uplands	Kirby Bellars	Melton	Leicestershire	England
Abbey Mill House, Abbey Mill Lane	St Albans, unparished area	St Albans	Hertfordshire	England
Land off Maltby Lane, North Lincolnshire	Barton-upon-Humber	Lincolnshire	Lincolnshire	England
Sheepdyke Lane	Bonby	Lincolnshire	Lincolnshire	England
Land at Elmdene, Cotesbach, Leiestershire	Cotesbach	Harborough	Leicestershire	England

5.4 Counties containing more than 100 sites

Example of a simple numerical query at county level.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX crm: <http://www.cidoc-crm.org/cidoc-crm/>
PREFIX aat: <http://vocab.getty.edu/aat/>
PREFIX os: <http://data.ordnancesurvey.co.uk/ontology/admingeo/>

SELECT DISTINCT ?county (count(?site) AS ?counter) where {
?county crm:P2_has_type os:County; rdfs:label ?countyName .
?site crm:P2_has_type aat:300000809; crm:P89_falls_within ?county .
}
GROUP BY ?county
HAVING (?counter > 100)
ORDER BY DESC(?counter)
```

county	counter
Greater London	1181
Kent	480
Lincolnshire	420
Warwickshire	380

Essex	276
Devon	244
West Midlands	218
Derbyshire	180
East Sussex	172
City and County of the City of London	167
Hampshire	133

5.5 Sites where magnetometer surveys took place

Example query on site and type of archaeological investigation.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX crm: <http://www.cidoc-crm.org/cidoc-crm/>
PREFIX aat: <http://vocab.getty.edu/aat/>
PREFIX os: <http://data.ordnancesurvey.co.uk/ontology/admingeo/>

SELECT DISTINCT ?siteName ?countyName WHERE {
?record crm:P67_refers_to
<http://purl.org/heritagedata/schemes/agl_et/concepts/145144> ;
crm:P67_refers_to [crm:P7_took_place_at ?site] .
?site crm:P1_is_identified_by [rdfs:label ?siteName] ;
```

```
crm:P89_falls_within [crm:P2_has_type os:County; rdfs:label ?countyName] .
}
```

siteName	countyName
Warwick Road, Coventry, West Midlands, England	West Midlands
Castor	Cambridgeshire
Skeeby Solar Site	North Yorkshire
The Dairy	Derbyshire
Land south of Desborough	Northamptonshire
Clun Castle	Shropshire
Cawood, Cawood, North Yorkshire, England	North Yorkshire

5.6 Reports published in 1997, documenting investigations in England

Example query on location and year of publication.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX crm: <http://www.cidoc-crm.org/cidoc-crm/>
PREFIX aat: <http://vocab.getty.edu/aat/>
PREFIX os: <http://data.ordnancesurvey.co.uk/ontology/admingeo/>
```

```
SELECT DISTINCT ?doi ?reportTitle WHERE {
?publication crm:P4_has_time-span [crm:P82b_end_of_the_end "1997"^^<http://www.w3.org/2001/XMLSchema#gYear>];
crm:P16_used_specific_object ?report .
?report crm:P102_has_title [rdfs:label ?reportTitle] ;
crm:P70_documents ?investigation .
?investigation crm:P7_took_place_at [crm:P89_falls_within ?country] .
?country crm:P2_has_type os:EuropeanRegion; rdfs:label "England"@en .
OPTIONAL { ?report crm:P1_is_identified_by [rdfs:label ?doi] }
}
```

doi	reportTitle
10.5284/1038992	An Archaeological Watching Brief at Hartwell (Smithfield) Garage site, Digbeth, Birmingham
10.5284/1038992	The Churchyard of St Philip's Cathedral, Birmingham: An Archaeological Desk-Based Assessment
10.5284/1038992	An Archaeological Desk-Based Assessment of the Proposed Martineau Galleries Development, Birmingham City Centre

5.7 Names of organisations who carried out investigations in Wales

Example query relating to archaeological units active in a particular country.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX crm: <http://www.cidoc-crm.org/cidoc-crm/>
PREFIX os: <http://data.ordnancesurvey.co.uk/ontology/admingeo/>

SELECT DISTINCT (UCASE(?organisationName) AS ?name) WHERE {
?country crm:P2_has_type os:EuropeanRegion; rdfs:label "Wales"@en .
?investigation crm:P7_took_place_at [crm:P89_falls_within ?country] ;
crm:P14_carried_out_by [a crm:E74_Group; crm:P1_is_identified_by ?org] .
```

```
?org rdfs:label ?organisationName .
}
```

---

organisationName

EXETER ARCHAEOLOGY

BIRMINGHAM ARCHAEOLOGY

JEN'S DIGGERS

ANTLER HOMES

5.8 Investigations where the record refers to a hearth

Example query referring to a particular type of context, as part of a potential archaeological research question.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX crm: <http://www.cidoc-crm.org/cidoc-crm/>

SELECT DISTINCT ?title WHERE {
?record crm:P67_refers_to <http://purl.org/heritagedata/schemes/eh_tmt2/concepts/70374> ;
crm:P67_refers_to [a crm:E7_Activity; crm:P1_is_identified_by ?inv] .
?inv a crm:E35_Title; rdfs:label ?title
}
```

---

investigationTitle

ARCHAEOLOGICAL EXCAVATIONS AT LAND AT DURRANTS LANE, BERKHAMSTED, HERTFORDSHIRE,

Geophysical Survey at Clun Castle

120 Cheapside

Land at Elmstead Hall, Elmstead Market, Essex

Royal Institute of Chartered Surveyors, 12 Great George Street, SW1P

Lincoln College

Bansons Yard Excavation

Monitoring at the Recreation Ground, School Lane, Watton-at-Stone

Holbury Infant School, Holbury, Southampton, Hampshire

An Archaeological Evaluation Along the Route of the Proposed Isle of Grain Gas Transmission Pipeline

Plot 9, Cabot Park, Avonmouth, Bristol

Land to the Rear of 106 High Street, Maldon, Essex

Brook House, Henbrook Lane, Upper Brailes, Warwickshire

5.9 Investigations carried out in the West Midlands between 1996 and 1998

More elaborate query on location and publication dates.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX crm: <http://www.cidoc-crm.org/cidoc-crm/>
PREFIX aat: <http://vocab.getty.edu/aat/>
PREFIX os: <http://data.ordnancesurvey.co.uk/ontology/admingeo/>

SELECT DISTINCT ?title WHERE {
?investigation crm:P1_is_identified_by [a crm:E35_Title; rdfs:label ?title];
crm:P7_took_place_at [crm:P89_falls_within [crm:P2_has_type os:County; rdfs:label "West Midlands"@en]];
crm:P4_has_time-span [crm:P82a_begin_of_the_begin ?minDate; crm:P82b_end_of_the_end ?maxDate] .
FILTER(year(?minDate) >= 1996 && year(?maxDate) <= 1998) .
}
```

title
An Archaeological Desk-Based Assessment of the Proposed Martineau Galleries Development
Hartwell (Smithfield) Garage Site, Digbeth, Birmingham
An archaeological watching brief at Hartwell (Smithfield) Garage site, Digbeth, Birmingham

Early Gasworks, Gas Street, Birmingham, Architectural Recording and Analysis: An Interim Report

The Church of St Philip's Cathedral, Birmingham: An Archaeological Desk-Based Assessment

An archaeological watching brief at The Old Crown, Deritend, Birmingham

5.10 Locations of sites in Cornwall

Example spatial query of sites within a county.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX crm: <http://www.cidoc-crm.org/cidoc-crm/>
PREFIX aat: <http://vocab.getty.edu/aat/>
PREFIX os: <http://data.ordnancesurvey.co.uk/ontology/admingeo/>

SELECT DISTINCT ?coordinates WHERE {
?site crm:P2_has_type aat:300000809;
crm:P168_place_is_defined_by ?coordinates;
crm:P89_falls_within [crm:P2_has_type os:County; rdfs:label "Cornwall"@en ].
}
```

coordinates

POINT(179180.000615553 61459.9993489836)

POINT(200050.00063613 80939.9993715717)

POINT(213280.00064826 88079.9993809834)

POINT(213310.000648253 87859.9993808218)

POINT(224550.000660276 106731.999398628)

POINT(224348.000660102 106680.999398539)

POINT(171950.000605711 36349.9993267181)

POINT(175750.00060882 35299.9993275672)

POINT(207380.000640189 66449.9993624105)

POINT(203560.000637877 72509.9993660299)

POINT(199630.000635425 78479.9993695164)

POINT(235510.000661279 53849.9993597119)

POINT(224930.000661506 113219.999403928)

POINT(224850.000661432 113169.999403876)

POINT(224830.000661028 110409.999401651)

POINT(225460.000661548 110409.999401788)

POINT(224600.00066085 110499.999401664)

POINT(194290.000627184 52799.9993478575)

POINT(207061.034860021 87006.5378345016)

POINT(183477.031453526 62694.3024857912)

POINT(221080.000658287 113039.999402884)

POINT(168797.081565743 37353.8656281533)

5.11 Visualisations of the Linked Data

Of course, the tabular results from SPARQL queries can be displayed in other ways. As an example the 5.10 result coordinates could be batch converted to Lat/Long values using https://www.doogal.co.uk/BatchReverseGeocoding.php and output in KML format, then input to https://geojson.io/ to generate a display of the locations of sites in Cornwall (Figure 4).

Figure 4: Example visualisation of 5.10 Query (Location of sites in Cornwall). Extract from a larger map via http://geojson.io/ © Mapbox © OpenStreetMap

Tools such as the [AllegroGraph Gruff](#) utility can be used to illustrate typical entities, properties and connectivity or the semantic context of specific cases within the linked data. Figures 5-8 are indicative of the possibilities and illustrate the range of semantic links within the data model.
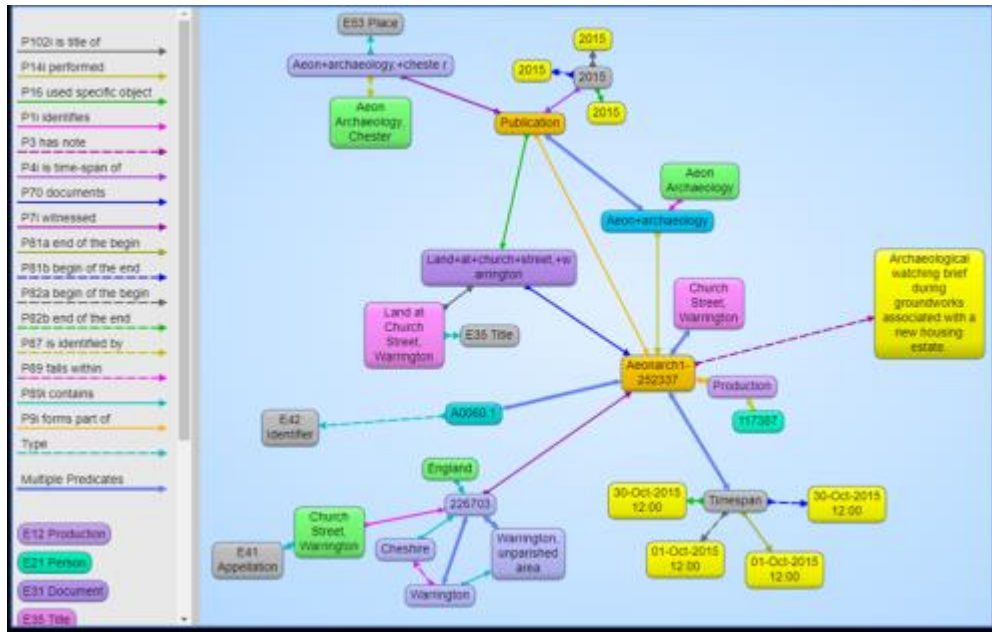

Figure 5: Example visualisation showing activity on a site with location and timespan
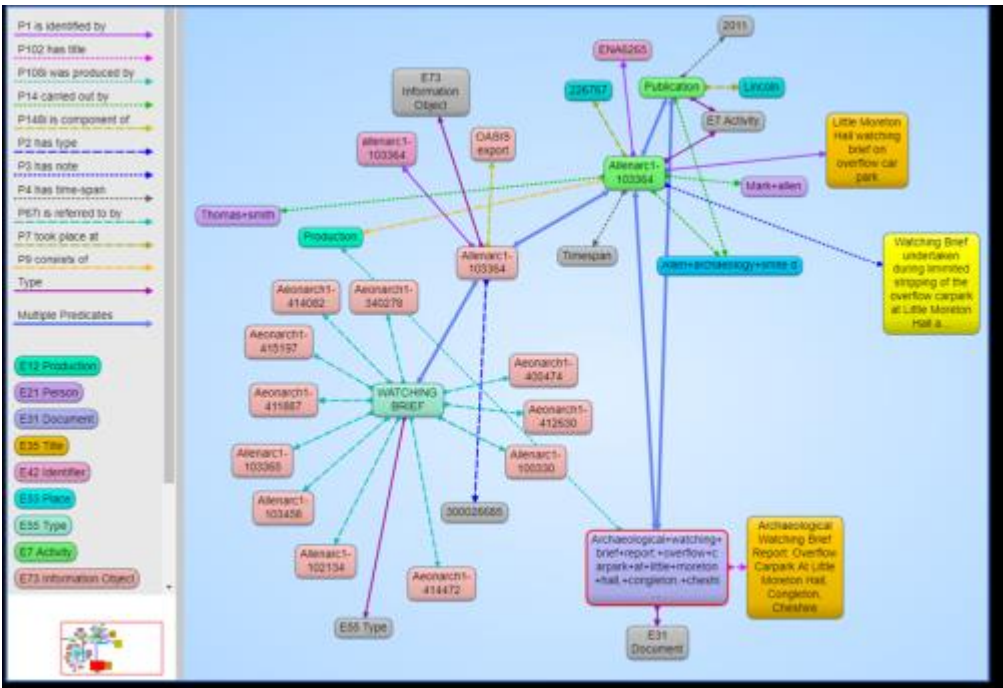
Figure 6: Example visualisation showing a report publication resulting from a particular site activity

Figure 7: Example visualisation showing report and other records referring to 'watching brief' activities

Figure 8: Example visualisation showing activity on one site in Southampton, as well as other sites in Southampton

6. Further Work and Discussion

The LD4HE outcomes enable the production of RDF from OASIS exports. This case study has demonstrated the feasibility and provided open source tools to achieve the conversion, a first step in the production of Linked Data. Various issues arise from the exercise, together with potential avenues for further work. We start with some practical points and then move on to more general reflections on Linked Data arising from this case study and other work on wider considerations for heritage interoperability and linked data.

6.1 Further modelling in OASIS

If use cases could be identified then further work could consider whether there is a need for additional detailed modelling of OASIS metadata, for example explicitly modelling HERs and their regions of responsibility or distinguishing the DOIs of single reports from

DOIs of series (collections) of reports. Further links could be added to other relevant Linked Data collections. The potential for automatic linking of closely associated OASIS reports (e.g. different stages of work on the same archaeological site) could also be explored.

Further work might also examine the potential for making additional links to external datasets, facilitated by the conversion to RDF. This could, for example, include the possibility of automatically creating links with other HER-derived data, such as the Heritage Gateway. This could extend to Linked Data from other UK institutions (non-cultural heritage), such as the Ordnance Survey and British Geological Survey (BGS). At the time of writing OASIS does interact with the BGS Web Mapping Service (WMS) for 1:50 000-scale geological maps for England, Wales and Scotland. A spatial query to the WMS returns drift and solid geology, storing the geological term and URI in the OASIS database. This only applies to geophysical projects - as geology is a factor that determines methodology and impacts upon interpretation - and thus is not part of the core OASIS metadata used by the LD4HE project. A follow up to the work described here could add this into the model and enhance the re-use of OASIS metadata.

6.2 OASIS data as a research tool

There is potential for a wider re-use of OASIS data. The example queries in Section 5 begin to go beyond simple *When, What, Where* queries by combining elements of the data model in more elaborate queries. These could be further elaborated to allow the investigation of archaeological research questions that perhaps might search across sites or HERs or archaeological units (for example future data deriving from the HS2 work). A longer-term aim of OASIS is to better incorporate the classifications being developed by the new generation of Research Frameworks in England, Scotland and Wales. These add a greater level of archaeological understanding and context, such as the process of 'Romanisation' or the transition from hunter-gathering to agriculture.

To encourage any programmatic access, it could be useful to provide a menu with a wide-ranging set of queries (and explanations) that would facilitate the tailoring of queries for particular purposes. For example, see Zeng and Mayr's (2019) discussion of the Getty's provision of a comprehensive set of well-documented query templates to allow programmatic users of the Getty Art and Architecture Thesaurus to locate and tailor the example queries.

6.3 UK vocabulary interoperability

Currently, different SPARQL queries are necessary to make the same search over English, Scottish and Welsh derived OASIS data. This is because of the different national vocabularies employed (see for example the Monuments example in Section 3). If vocabulary mappings were created between the corresponding UK-national vocabularies (English, Scottish, Welsh) using the standard SKOS mapping relationships then semantic search functionality could search across the different national OASIS data. For instance, this could be achieved by SPARQL queries using the new mappings.

6.4 Archaeological challenges for data interoperability

There are various issues specific to archaeological datasets that can pose challenges for interoperability and re-use of data. One underlying issue derives from the process of archaeological investigation and publication – there can be various stages of data production, as observed in the STAR project (Tudhope *et al*. 2011). Not all projects necessarily complete every stage. May *et al*. (2015) identified four possible stages in the data workflow:

- Immediate excavation data archived after work on site completed
- Excavation data with analysis outcomes after interpretation completed
- Final publication of synthesised results
- Semantically integrated data archive.

This workflow is a generalisation; some projects may omit/combine stages and add other elements to the workflow. The point remains that given datasets may result from different stages in this process, which poses problems for semantic interoperability. In the absence of standard metadata, it may not even be obvious what workflow has been applied to a particular dataset. For example, interpretation analysis notes provided (only) in text fields may override earlier, provisional category assignments in data fields. In some cases, final publication of results may be found in a journal article or monograph rather than any dataset.

May *et al*. (2015) describe different recording methodologies commonly used in different countries; not everyone uses the single context recording system common in the UK. Cross-analysis at a general level may be possible but detailed comparison of excavation data from different recording systems can be challenging if that is required. This may be partly addressed by conversion to a semantic framework with fine granularity (as discussed in Section 6.5). In addition, semantic integration, in our experience, requires a high degree of data cleaning, which necessarily involves changes to the source data and sometimes implicit judgements of intended meaning. Metadata describing the workflow associated with production of a particular dataset should include some

characterisation of any data cleaning applied. There may be a case for archiving different versions (stages) of a dataset, each with its metadata. For systematic re-use of data, a wider set of contextual information (or paradata) is desirable, including the archaeological methodologies applied, coverage of data, etc. See the review and discussion by Huggett ([2018](#)) of wider issues in data re-use and associated literature.

6.5 Routes to semantic interoperability

There are various avenues open to heritage organisations for enhancing the semantic interoperability and re-use of their resources. Discussion around Linked Data may tend to conflate distinct issues – these include the mechanism for making the data available, the extent or selection of the data to be converted, the output data model and the anticipated use cases for the new expression of the data.

There are several options available to an organisation for making data available for re-use by third parties and these mechanisms can also be combined in tandem. These include making exports of the data available for download, making the dataset available for harvesting, making elements of the dataset available for programmatic access via an API, making the dataset available following Linked Data principles and technical standards via a Linked Data server and/or SPARQL endpoint. If the data are periodically updated then that should be taken into consideration in the selection of mechanisms. Good practice general [guidelines](#) are available from bodies such as [data.gov.uk](#) for harvesting public data, including persistent identifiers and common metadata standards for datasets, such as [DCAT](#). The choice of mechanism partly depends on the anticipated external use cases of the data made available. A use case may, for example, involve routine processing of the data according to its source data model or depend on specialist libraries provided by a third-party platform. On the other hand, [Linked Data](#) generally aims to facilitate connections to other datasets and encourage links back. To this end, [FAIR Principles](#) and [W3C Linked Data guidelines](#) emphasise the assignment of persistent identifiers (PIDs) as web URIs, the use of standard vocabularies (with PIDs) and, where appropriate, conversion from a source data model to common data models or semantic frameworks for the field in question. Linked Data also involves the conversion to semantic web representation formats such as RDF and/or JSON.

Ambitions for the re-use of OASIS data include archaeological research questions involving cross-search and meta research, both internally and via third-party data repositories of relevant material. This use case supports the choice of mechanisms, such as Linked Data, that involve the conversion of the source data model to a common framework, with the benefit of semantic interoperability over disparate terminologies and data schema and possibilities of enhanced user services and cross-search. A detailed Linked Data

survey by OCLC ([2018](#)) considered opportunities and challenges and distinguished the concerns of publishers from those of consumers as well as general benefits, such as increased kudos and staff development. For example, publishers of Linked Data could potentially expose their data to a larger audience on the web, increase data reuse and interoperability, linking information across different institutions. Projects/services consuming that linked data might enhance local data with Linked Data from other sources and provide their users with a richer experience. In an extensive review of archaeological Linked Data, Geser ([2016](#)) considers the major benefits as arising from the integration of heterogeneous datasets and enhanced services. However, these are anticipated future benefits and the conversion process (as described in earlier sections of this article) can be technically challenging for many organisations. Some organisations may have the advantage of participation in infrastructure initiatives, such as ARIADNE (Aloia *et al*. [2017](#)) and [ARIADNEplus](#), [Europeana](#), [Linked.Art](#), where there may be some support for conversion to a semantic framework. For example, ARIADNEplus has developed an e-infrastructure for European archaeological research that allows data providers to provide access to a variety of data resources, enabling a diverse range of impact possibilities (Niccolucci and Richards [2019](#)). More widely, the challenges for widespread adoption of Linked Data across different types of organisation are related to possible judgements of an unfavourable cost/benefit ratio for its production (Geser [2016](#)). Can this ratio be improved? The wider availability of practical guidelines, tools and exemplars is probably the most important step - the [LOUD](#) (Linked Open *Usable* Data) principles are an example of this approach, with an emphasis on minimal barriers for getting started and practical documentation with working examples. We discuss various other issues below, drawing on experience from the LD4HE case study and other reflections.

Firstly, it should be noted that some of the Linked Data principles, such as using PIDs for data elements and concepts in standard vocabularies, can be applied to source datasets and data models, without necessarily requiring the creation of a full Linked Data server. Datasets could be made available for export as JSON-LD, for example, with links to standard vocabularies and external datasets (such as [Ordnance Survey Linked Data](#)). It would not facilitate external links back to elements of the source dataset but a looser linking data method could be easier to achieve. Reflecting on experience with [Open Context](#), Kansa ([2014](#)) makes the case for vocabulary alignment as cost-effective, low-hanging fruit in Linked Data. Binding and Tudhope ([2016](#)) review work in this area and discuss the use of the [Getty AAT Linked Data](#) as a vocabulary mapping hub in [ARIADNE](#).

There are cost/benefit considerations in the level of granularity adopted in the target data model or framework for the conversion to Linked Data. The choice of the standard CIDOC CRM ontology as the model for LD4HE's Linked Data, in combination with the standard HeritageData vocabularies, facilitates wider connections with national and international datasets using the same conceptual model. Previous work in the STAR project investigated the potential for [highly specific queries](#) (e.g., *hearths containing coins* or *contexts containing coins that are stratigraphically below contexts of type floor*) on diverse archaeological datasets and reports, using a more detailed archaeological extension of the CIDOC CRM ontology (Tudhope *et al*. [2011](#)). In contrast, LD4HE selected a

core set of CIDOC CRM elements to form the basis of the data model, rather than more specialised CIDOC CRM extensions, in light of the OASIS focus on higher level metadata (and research questions) in order to facilitate the widest opportunity for potential connection with other datasets. This approach, combining a core ontology with specific vocabularies, was also followed in a semantic integration case study on the theme of wooden objects and dendrochronological dating combining data and reports in different languages (Binding *et al*. 2019). Consideration should be given to the granularity of the anticipated Linked Data use cases and whether the aim is mainly the discovery of datasets for download and subsequent local processing, fairly high-level research questions, or rather the investigation of detailed research questions via the integrated Linked Data platform.

Mapping a local data model to a target data model or ontology and converting the data can be resource intensive and requires detailed knowledge of the ontology. Mapping patterns that express the mapping for a local data model can reduce the effort of repeated conversions of similar datasets. Similarly, it may be cost effective to convert heterogeneous local datasets to an intermediate data model and then apply a standard mapping pattern to a more complex ontology (or multiple ontologies) – detailed knowledge of the ontology is only required when designing the final mapping stage. This general approach can be followed with different conversion tools. The Linked Art Data Model is an example of the mapping pattern approach orientated to artwork and museum collections. LD4HE used STELETO and mapping templates - once created a template can be used repeatedly to convert similar datasets to the target model or ontology. (Sometimes a data cleaning stage is necessary depending on the consistency of the source dataset.) Given the complexity of ontologies, such as the CIDOC CRM, it is possible for different users to make different valid mappings that can hinder practical interoperability. Once a mapping pattern has been agreed, using an appropriate conversion tool, then data conversion based on that pattern can produce consistent output (Binding *et al*. 2015).

All of the mechanisms for making data available can be applied to the whole dataset or a selection from the dataset, including choosing to make only (a selection of) the metadata available. Since the LD4HE case study discussed here involves a repository of fieldwork reports, we have been concerned with the conversion of OASIS *metadata*. On the other hand, archaeological fieldwork produces datasets as well as reports. OASIS can include both but there may be occasions when only a dataset is available. A strategy that converted full datasets to Linked Data at a fine granularity might be resource intensive and risk the conversion of low-level data elements that never saw third-party use (e.g. administrative data, individual cuts or deposits). Costs might be reduced by the selection of a subset of key data elements, as with Open Context, or a focus on a particular dimension, as in contributors to Pelagios (spatial) and PeriodO (temporal) Linked Data initiatives. Another choice is to select the metadata only to be made available as Linked Data, as with LD4HE. For purposes of dataset discovery for archaeological research, the metadata could be enriched to include significant findings including data elements (monuments, finds, contexts, etc.) from the intervention, essentially applying good practice in subject indexing.

7. Conclusions

Various outcomes have been achieved. New specialised vocabularies required by OASIS have been published on the HeritageData platform. A mapping from mandatory OASIS fields based upon the CIDOC CRM data model has been designed and published. A STELETO template has been developed to produce the data conversion that can be re-used to allow periodic conversions of OASIS exports to be converted to RDF, a major step in the production of Linked Data. The template has been refined and tested on various OASIS exports. A set of SPARQL queries on the OASIS exports demonstrates the outcomes of the data conversion and illustrates a range of possible queries and their potential for more elaborate archaeological research investigations.

In general, Linked Data affords potential benefits for cross-search, synthetic research and investigation of patterns that are not apparent in simple inventories. Through being expressed as RDF conforming to a standard conceptual framework, the OASIS data is made automatically readable and understandable by humans and machines. This facilitates inquiry and meta research across HER boundaries and over different types of archaeological intervention (as seen in the example queries). Examples of enquiries are given in the previous section for illustrative purposes but many more are possible – a richer set of OASIS metadata is made available for programmatic search than is possible with the ADS Library's user interface. Reflections on the case study and cost/benefit considerations for Linked Data conversion have been discussed, together with possible strategies for reducing the costs of producing Linked Data.

Acknowledgements

Bibliography

Aloia, N., Binding, C., Cuy, S., Doerr, M., Fanini, B., Felicetti, A., Fihn, J., Gavrilis, D., Geser, G., Hollander, H., Meghini, C., Niccolucci, F., Nurra, F., Papatheodorou, C., Richards, J., Ronzino, P., Scopigno, R., Theodoridou, M., Tudhope, D., Vlachidis, A. and Wright, H. 2017 'Enabling European archaeological research: the ARIADNE E-Infrastructure', *Internet Archaeology* **43**. https://doi.org/10.11141/ia.43.11

Binding, C. and Tudhope, D. 2016 'Improving interoperability using vocabulary linked data', *International Journal on Digital Libraries* **17**(1), 5-21. Springer. https://doi.org/10.1007/s00799-015-0166-y

Binding, C., Charno, M., Jeffrey, S., May, K. and Tudhope, D. 2015 'Template based semantic integration: from legacy archaeological datasets to linked data', *International Journal on Semantic Web and Information Systems* **11**(1), 1-29. https://doi.org/10.4018/IJSWIS.2015010101

Binding, C., Tudhope, D. and Vlachidis, A. 2019 'A study of semantic integration across archaeological data and reports in different languages', *Journal of Information Science* **45**(3), 364-86. https://doi.org/10.1177/0165551518789874

Bizer, C., Heath, T. and Berners-Lee, T. 2009 'Linked Data - the story so far', *International Journal on Semantic Web and Information Systems* **5**(3), 1–22. https://doi.org/10.4018/jswis.2009081901

Charles, V., Freire, N. and Isaac, A. 2014 'Links, languages and semantics: linked data approaches in the European Library and Europeana', *Proc. IFLA World Library and Information Congress*, Lyon. http://ifla2014-satdata.bnf.fr/pdf/iflalld2014_submission_Charles_Freire_Isaac.pdf

Evans, T. 2015 'A reassessment of archaeological grey literature: semantics and paradoxes', *Internet Archaeology* **40**. https://doi.org/10.11141/ia.40.6

Geser, G. 2016 *ARIADNE WP15 Study: Towards a Web of Archaeological Linked Open Data*. http://legacy.ariadne-infrastructure.eu/wp-content/uploads/2019/01/ARIADNE_archaeological_LOD_study_10-2016-1.pdf

Gruber, E. 2016 'Linked open data for numismatic library, archive and museum integration' in M. Matsumoto and E. Uleberg (eds) *Proc. 44th Conference on Computer Applications and Quantitative Methods in Archaeology* (CAA2016). 33-40. https://zenodo.org/record/1304270#.YijZVC2cbVo

Huggett, J. 2018 'Reuse remix recycle: repurposing archaeological digital data', *Advances in Archaeological Practice* **6**(2), 93-104. https://doi.org/10.1017/aap.2018.1

Isaac, A., Waites, W., Young, W. abd Zeng, M. 2011 *W3C Library Linked Data Incubator Group Report 25: Datasets, Value Vocabularies, and Metadata Element Sets*. https://www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset-20111025/

Isaksen, L., Simon, R., Barker, E. and de Soto, C. 2014 'Pelagios and the emerging graph of ancient world data', *Proc. ACM Conference on Web Science (WebSci'14)*, 197–201. ACM. https://doi.org/10.1145/2615569.2615693

Kansa, E. 2014 'Open Context and Linked Data' in T. Elliott, S. Heath and J. Muccigrosso (eds) *Current Practice in Linked Open Data for the Ancient World*, ISAW Papers 7(10) http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/kansa/

Kansa, E. and Kansa, S. 2013 'Open Archaeology: We all know that a 14 is a sheep: data publication and professionalism in archaeological communication', *Journal of Eastern Mediterranean Archaeology and Heritage Studies* **1**(1), 88-97. https://doi.org/10.5325/jeasmedarcherstu.1.1.0088

May, K., Binding, C. and Tudhope, D. 2015 'Barriers and opportunities for Linked Open Data use in archaeology and cultural heritage', *Archäologische Informationen* **38**, 173-84. https://doi.org/10.11588/ai.2015.1.26162

Myers, D., Dalgity, A. and Avramides, I. 2016 'The Arches heritage inventory and management system: a platform for the heritage field', *Journal of Cultural Heritage Management and Sustainable Development* **6**(2), 213–24. https://doi.org/10.1108/JCHMSD-02-2016-0010

Niccolucci, F. and Richards, J. (eds) 2019 *The ARIADNE Impact*, Archaeolingua Foundation, Hungary. https://doi.org/10.5281/zenodo.4319058

OCLC 2018 *International Linked Data Survey*. https://www.oclc.org/research/areas/data-science/linkeddata/linked-data-survey.html

Shaw, R., Rabinowitz, A., Golden, P. and Kansa, E. 2016 'A sharing-oriented design strategy for Networked Knowledge Organization Systems', *International Journal on Digital Libraries* **17**(1), 49-61. https://doi.org/10.1007/s00799-015-0164-0

Tudhope, D., May, K., Binding, C. and Vlachidis, A. 2011 'Connecting archaeological data and grey literature via semantic cross search', *Internet Archaeology* **30**. https://doi.org/10.11141/ia.30.5

Wilkinson, M. *et al*. 2016 'The FAIR Guiding Principles for scientific data management and stewardship', *Scientific Data* **3**(1), 160018. https://doi.org/10.1038/sdata.2016.18

Zeng, M.L. and Mayr, P. 2019 ' Knowledge Organization Systems (KOS) in the Semantic Web: a multi-dimensional review', *International Journal on Digital Libraries* **20**, 209–30. https://doi.org/10.1007/s00799-018-0241-2