Semantic Representation and Location Provenance of Cultural Heritage Information: the National Gallery Collection in London

Joseph Padfield [1,†,*], **Kalliopi Kontiza** [1,†] , **Antonis Bikakis** [2,†] and **Andreas Vlachidis** [2,†]

[1] Scientific Department, The National Gallery, Trafalgar Square, London WC2N 5DN, UK;
    Kalliopi.kontiza@ng-london.org.uk
[2] Department of Information Studies, University College London, London WC1E 6BT, UK;
    a.bikakis@ucl.ac.uk (A.B.); a.vlachidis@ucl.ac.uk (A.V.)
*   Correspondence: Joseph.Padfield@ng-london.org.uk
†   These authors contributed equally to this work.

Abstract: This paper describes a working example of semantically modelling cultural heritage information and data from the National Gallery collection in London. The paper discusses the process of semantically representing and enriching the available cultural heritage data, and reveals the challenges of semantically expressing interrelations and groupings among the physical items, the venue and the available digital resources. The paper also highlights the challenges in the creation of the conceptual model of the National Gallery as a Venue, which aims to i) describe and understand the correlation between the parts of a building and the whole; ii) to record and express the semantic relationships among the building components with the building as a whole; and iii) to be able to record the accurate location of objects within space and capture their provenance in terms of changes of location. The outcome of this research is the CrossCult venue ontology, a fully International Committee for Documentation Conceptual Reference Model (CIDOC-CRM) compliant structure developed in the context of the CrossCult project. The proposed ontology attempts to model the spatial arrangements of the different types of cultural heritage venues considered in the project: from small museums to open air archaeological sites and whole cities.

Keywords: Ontology-based representation; CIDOC-CRM; Venue data model; Semantic Web applications for Cultural Heritage

1. Introduction

CrossCult[1], Empowering reuse of digital cultural heritage in context-aware crosscuts of European history, is a three-year H2020 research project, which started in March 2016. It consists of 11 European institutions and 14 associated partners, from Computer Science, History and Cultural Heritage. The goal of the project is to spur a change in the way European citizens appraise History, fostering the re-interpretation of what they may have learnt in the light of cross-border interconnections among pieces of cultural heritage, other citizen's viewpoints and physical venues. Four distinct Pilots contribute data to the CrossCult project covering a unique range of cultural heritage venues across Europe; from the large venue of National Gallery in London (Pilot 1) to the considerably smaller venue of the Archaeological Museum in Tripolis, Greece (Pilot 3), from the archaeological sites of

---

[1]   https://www.crosscult.eu/ Accessed January 29, 2019.

the Roman healing spa of Lugo in Spain, of Chaves in Portugal, of Montegrotto Terme in Italy and the Ancient theatre of Epidaurus in Greece (Pilot 2) to the historical points of interest in the cities of Luxembourg in Luxembourg and Valletta in Malta (Pilot 4).

For over a decade the field of Cultural Heritage has received significant attention in application of Semantic Web technologies, aimed at facilitating a harmonised and interoperable access over heterogeneous resources [1]. A fundamental challenge in dealing with Cultural Heritage data is to make the content mutually interoperable, so that it can be searched, linked, and presented in a harmonised way across the boundaries of the datasets and data silos. The difficulty of finding and relating information in this kind of heterogeneous content provision and data format environment creates an obstacle for end-users of cultural contents, and a challenge to organisations and communities producing the contents. The CrossCult project ingests a wide range of diverse data associated to Cultural Heritage objects, events and subjects that span from antiquity to modern times. Such disparate data means there is a wide array of formats, technologies, management and classification approaches relevant to each data provider or source. Hence, modelling such data in a coherent way to enable interoperability among the Pilots requires addressing the diversity of content types, data formats, and levels of data detail between the four pilots. Semantic Web technologies ease access to Cultural Heritage content facilitating new ways of engaging with heritage by the general public and experts that go beyond a simple interactive engagement. They provide an intelligent integration of resources via machine readable and human interpretable representations of a domain of knowledge (i.e. ontology), enabling retrieval, reasoning, optimal data integration and knowledge reuse of disparate cultural heritage resources. The benefits of Semantic Web technologies to Cultural Heritage are evident in literature including; a harmonised view to disparate and distributed contents, intelligent content aggregation, semantic search-browsing and recommendation, content enrichment and reuse [2,3]. In this respect, the Conceptual Reference Model (CRM) of the International Council of Museums (ICOM)—International Committee for Documentation (CIDOC), CIDOC-CRM (ISO 21127:2014), provides an object-oriented schema based on real world concepts and events implementing data harmonisation based on the relationships between things rather than artificial generalisations and fixed field schemas.

The model has been gaining increased popularity and it is considered to be a major standard in knowledge representation of Cultural Heritage data. Adopted by numerous small and large scale projects, it offers rich semantic representation and rigour definitions, sympathetic to the data and the different and varied perspectives of the cultural heritage community [4]. The CLAROS (Oxford University) project [5] was one of the first cases (2010) to provide interoperability over a large collection of cultural heritage data (20 million records) using CIDOC-CRM as the underlined semantic layer. Since then prestigious CH institutions, such as the British Museum (BM) [6] and the American Numismatic Society (ANS) [7] have pursed projects that advance knowledge representation and content provision of their collections using CIDOC-CRM semantics. The BM ResearchSpace[2] is a Semantic Web platform that provides a collaborative research environment for uncovering relationships and connections between CIDOC-CRM harmonised datasets, whereas the ANS Kerameikos initiative proposes use of CIDOC-CRM semantics for normalising classical pottery databases to facilitate large scale data aggregation and subsequent analyses. In addition, the EU FP7 ARIADNE Infrastructure aimed at integrating existing archaeological distributed and disparate data across Europe has used CIDOC-CRM as the backbone of the ARIADNE Reference Model [8].

Our first step for achieving interoperability (at the semantic level) was to adopt the CIDOC Conceptual Reference Model[3] as the core conceptual component of the CrossCult Knowledge Base (CCKB), a semantic knowledge base that stores all Pilots' data. The employment of CIDOC-CRM

---

[2] https://www.researchspace.org/ Accessed January 29, 2019.
[3] http://www.cidoc-crm.org/ Accessed January 29, 2019.

enabled us to integrate the disparate datasets of the four project Pilots and their metadata under a common semantic layer driving cross-search and inference capabilities. On the other hand, CIDOC-CRM as a formal and generic structure of concepts and relationships is not tied to any particular vocabulary of types, terms and individuals. In order to address the vocabulary needs of the project and enable interoperability also at the syntactic level, we developed and connected to CIDOC-CRM an additional vocabulary structure, which integrates terms from standard external glossaries and thesauri. A detailed description of the CCKB and its main components can be found in Section 2.

One of the requirements for the CCKB is to store semantic descriptions not only of the collections of the four Pilots (objects and Points of Interest), but also of the Venues themselves, building from a generic venue description structure. Metadata standards for the documentation of the built heritage and the archaeological complexes attempt to record the semantics of the building's components but have in the past often failed in describing the completeness of information about the building and the relationships among the parts and the whole. The aim of the conceptual model of Venues in the CCKB is to i) describe and understand the correlation between the parts of a building and the whole; ii) to record and express the semantic relationships among the building components with the building as a whole; and iii) to be able to record the accurate location of objects within space and capture their provenance in terms of changes of location. A detailed description of the CCKB and the conceptual model of the proposed CrossCult Venue Ontology and its main components can be found in Section 2.

The remainder of the paper is focused on Pilot 1, whose aim is to demonstrate how the CrossCult platform can facilitate the discovery and exploration of connections between objects (paintings), subjects depicted, people (painters) and events (painting creation) across European history. In recent years the National Gallery London (NG) has contributed to a number of collaborative documentation research and development projects. From examining searching and the semantic web in EU projects like ARTISTE[4] and SCULPTEUR[5], to general information resource building with the Andrew W. Mellon Foundation funded Raphael Research Resource[6], and the current H2020[7] projects, developing the potential of cultural heritage digital documentation in IPERION CH[8] and CrossCult. This research has examined and developed a variety of processes and tools to facilitate the gathering, storage, use and presentation of cultural heritage related material and has led to the work presented in this paper; using international standards to interact with large numbers of images, combining separate sources of digital information and mapping the complex semantic relationships that connect them together.

During the first two years of the CrossCult project we focused on aggregating the NG data and developing the semantic definition of the NG collection information (an example of this mapping is available in Section 4). This detailed definition allowed us to describe how we can structure and store the varied complex relationships and connections between paintings, artists and materials and map these relationships to the agreed project ontology. We also tried to record the actual location of paintings to a specific point on a wall, moving beyond the simple room location data that has been available at the existing NG dataset. The process of how we expanded the existing NG data to cover more detailed painting location information and the mechanism to track how it changes over time can be found in Section 3. The paper concludes in Section 5 summarising the presented work.

---

[4] http://www.southampton.ac.uk/~{}km2/projs/artiste/ Accessed January 29, 2019.
[5] http://www.sculpteur.ecs.soton.ac.uk Accessed January 29, 2019.
[6] Raphael Research Resource - http://cima.ng-london.org.uk/documentation Accessed January 29, 2019.
[7] Horizon 2020 EU Research and Innovation programme - https://ec.europa.eu/programmes/horizon2020/en/ Accessed January 29, 2019.
[8] Integrated Platform for the European Research Infrastructure ON Cultural Heritage - http://www.iperionch.eu/ Accessed January 29, 2019.

2. CrossCult Ontology

2.1. The CrossCult Knowledge Base

The CrossCult Knowledge Base (CCKB) [9] is a multi-layered structure of semantics aimed at facilitating interoperable connections between cultural heritage data. Based on maximum reuse of well-established technologies, it incorporates a set of standard Semantic-Web technologies and formats to support the data modelling requirements and objectives of CrossCult project. The CCKB stack (see Figure 1) illustrates the architecture of the knowledge base, where each section carries different semantics: a) the bottom section carries the semantics of different standard ontological schemas adopted in the CCKB; b) the middle section accommodates the project-specific cultural heritage semantics; c) the side section refers to the complementary CrossCult Classification Scheme (CCCS) vocabulary; and d) the top section to the representation of venues and users.
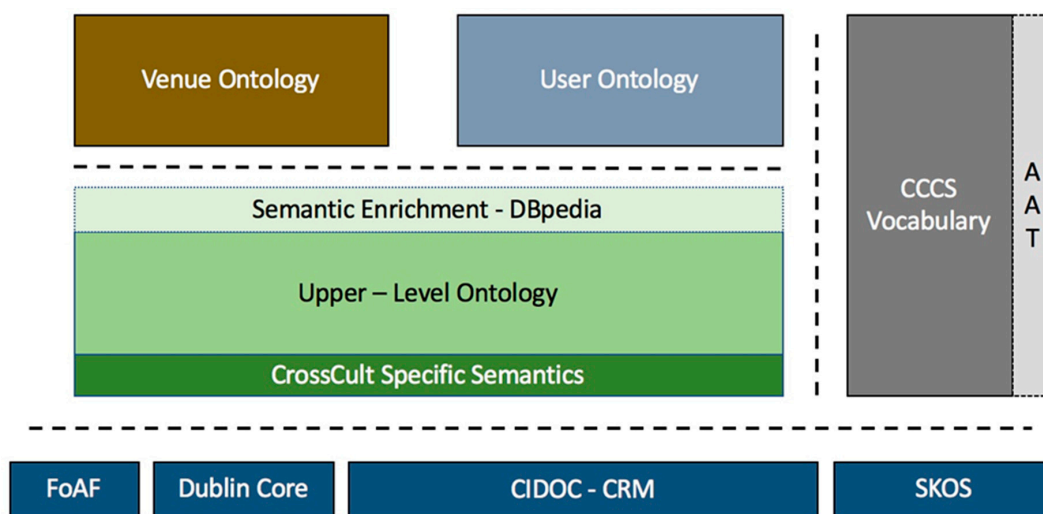
Figure 1. The architecture of the CrossCult Knowledge Base. CCCS: CrossCult Classification Scheme; CIDOC-CRM: International Committee for Documentation Conceptual Reference Model; AAT: Arts and Architecture Thesaurus of Getty; FOAF: FOAF (Friend-Of-A-Friend).

The four schemas of the bottom section constitute the foundation of the architecture with CIDOC-CRM being the most prominent. The framework is complemented by the semantics of the Simple Knowledge Organization System (SKOS)[9]; the Dublin Core Schema, a standard vocabulary for describing web resources; and the FOAF (Friend-Of-A-Friend)[10] ontology, which is used for mediating the semantics between the User Ontology layer and the Upper-Level Ontology layer in terms of describing user related entities and their interests. The middle layer accommodates the semantics of the Upper-level ontology, which is defined as a generic conceptual structure for accommodating common concepts and relationships across a diverse range of cultural heritage data. To this aim, CIDOC-CRM as the core model of the layer guarantees the use of well-defined and interoperable semantics, whilst allowing for project-specific specialisations which address the requirements of reflection, holistic understanding and reinterpretation of the European history.

On the other hand, CIDOC-CRM as a formal and generic structure of concepts and relationships is not tied to any particular vocabulary of types, terms and individuals. The particular need for an additional level of vocabulary-based semantics is addressed by the side section, which accommodates the faceted vocabulary structure CrossCult Classification Scheme (CCCS). The scheme provides

---

9 https://www.w3.org/2004/02/skos/ Accessed January 29, 2019.
10 http://xmlns.com/foaf/spec/ Accessed January 29, 2019.

skosified concepts to the middle and top layer of the architecture which are linked to ontology instances via the P67. refers to or P2. has type properties. The role of CCCS is not to classify objects according to their characteristics, which is handled by the ontology, but to provide a supplementary layer of terminology (subjects, types etc.) that can be useful during retrieval. Wherever possible, CCCS concepts are linked to external semantic definitions from standard thesauri resources such as, the Arts and Architecture Thesaurus of Getty (AAT)[11], the EUROVOC[12], the UNESCO Thesaurus and the Library of Congress Subject Authorities (LC) vocabulary[13]. The CCCS polyhierarchical structure also allows for concepts to be linked to multiple parents, thus one concept may appear in multiple hierarchical views. The CCCS was developed using the TemaTres[14], a web application for managing documentation languages, oriented to the development of hierarchical thesauri, on which several editors can be working at the same time. It allows both a systematic and an alphabetical list of terms, and offers different options to perform searches, such as simple search or expanded search through related or hierarchical terms.

The top section of the architecture contains the Venue and the User ontologies. The Venue ontology is a fully CIDOC-CRM compliant structure, which aims to model the spatial arrangements of the different venues that participate in the project. The User ontology is a CrossCult centric structure aimed at supporting the user modelling requirements of the project with respect to the user interests, visit experience, user background and other demographic information. The ontology combines elements from the Friend of a Friend (FOAF) and CIDOC-CRM models while it introduces project-specific classes and properties to address particular user modelling requirements, such as fatigue, prior knowledge and behaviour.

2.2. CrossCult Upper level Ontology

The CrossCult Upper-level ontology is a single and generic conceptual structure that acts as a semantic layer of common concepts and relationships across the four pilots of the project. It delivers formalisms and conceptual arrangements which enable augmentation, linking, semantic-based reasoning and retrieval across disparate data resources. The ontology builds on standard Semantic Web technologies and maintains full compatibility with CIDOC-CRM, containing the least minimum set of CRM concepts as described in the latest specification document version 6.2.3. Aimed at maximum reuse of established Semantic-Web definitions, the structure is written in OWL2[15], following the Erlangen CRM[16] (version 140220) implementation and complemented by SKOS, FOAF (Friend-Of-A-Friend) and Dublin Core[17] semantics. Project-specific entities which address the requirements of reflection, holistic understanding and reinterpretation of the European history are also incorporated in the ontology whilst a selected set of ontology instances is enriched with links to DBpedia concepts[18].

Figure 2 presents the core elements (classes and properties) of the Upper-level ontology and the modelling arrangements of the common semantics across the four project pilots for modelling cultural heritage objects. At the core of the model resides the CIDOC-CRM entity E18.Physical Item, which comprises all persistent physical items with a relatively stable form, man-made or natural. The entity enables the representation of a vast range of items of interest, such as museum exhibits, gallery paintings, artifacts, monuments and points of interest, whilst providing extensions to specialised

---

[11]   The Art & Architecture Thesaurus: a structured vocabulary of approximately 44,000 concepts of art, architecture and culture items http://www.getty.edu/research/tools/vocabularies/aat/ Accessed January 29, 2019.
[12]   EuroVoc: a multilingual, multidisciplinary thesaurus, aiming to support the information management and dissemination services of the EU and its members http://eurovoc.europa.eu Accessed January 29, 2019.
[13]   Library of Congress Subject Authority Records http://id.loc.gov/authorities/subjects.html Accessed January 29, 2019.
[14]   http://www.vocabularyserver.com/ Accessed January 29, 2019.
[15]   https://www.w3.org/TR/owl2-syntax/ Accessed January 29, 2019.
[16]   http://erlangen-crm.org/ Accessed January 29, 2019.
[17]   http://dublincore.org Accessed January 29, 2019.
[18]   DBpedia: a crowd-sourced generic dataset containing information created in various Wikimedia projects structured in RDF http://wiki.dbpedia.org Accessed January 29, 2019.

entity definitions of targeted semantics for man-made objects, physical objects and physical features. The arrangement benefits from a range of relationships between E18.Physical Item and a set of entities that describe the static parameters of an item, such as dimension, unique identifier, title and type. The model also allows the description of more complex objects through a composition of individual items (i.e., P46.is_composed_of). Moreover, the well-defined semantics enable rendering of rich relationships between the physical item and entities describing the item in terms of ownership, production, location, and other conceptual associations. The project-specific property reflects enables specific, direct connections between existing concepts and the CrossCult class Reflective Topic.
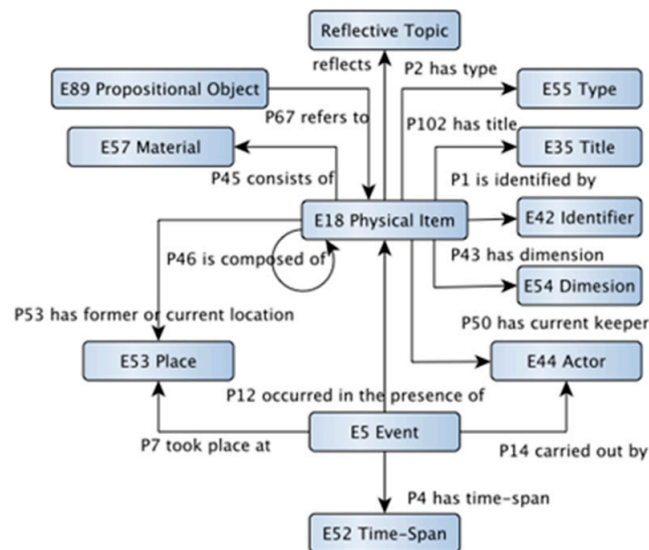
Figure 2. Core Elements of the Upper-level Ontology.

A fully-fledged example of the Upper-level ontology is shown in Figure 3, which presents a detailed modelling view of the National Gallery painting ID NG6576 (Eustache Le Sueur, Alexander and his doctor, about 1648-9). The painting is modelled as an instance of E22.Man-Made Object uniquely identified by a National Gallery (UK) reference and associated with a skosified type (e.g., Canvas painting). The modelling of typical information about the painting such as its size, material, medium and support, date of production and ownership is not different from the approach proposed by the CIDOC-CRM official tutorial and evident in well-known projects, such as the ResearchSpace of British Museum. A unique element of the CrossCult Upper-level ontology is the semantics of the Reflective Topic entity, which encompasses all those connections that can be made to create a network of points of view to aid reflection and prospective interpretation over a topic and to enable interconnection between physical or conceptual things of man-made or natural origin. A broader reflective topic can be composed by more specific (narrower) topics, in the same way as an E89. Propositional Object can be composed by other objects, using the P148.has_component property. The core CRM classes of the model are shown on blue and the skosified entities are in pink, whereas the ontology individuals are represented in boxes with DBpedia links shown in bright yellow.
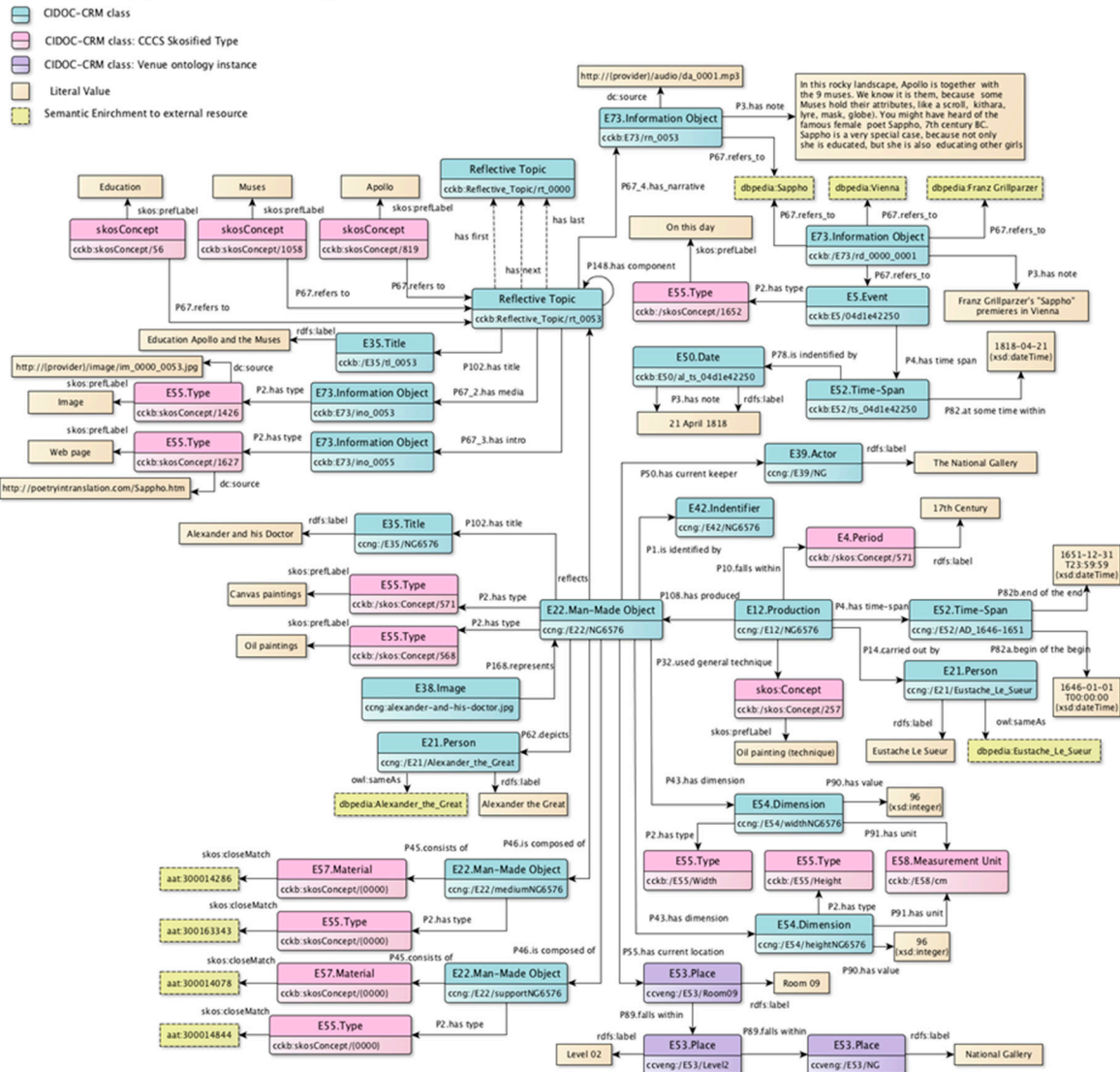
Figure 3. A detailed example of the CrossCult Upper-level Ontology and relationships of ontology individuals. CCCS: CrossCult Classification Scheme.

2.3. CrossCult venue Ontology

The CrossCult Venue ontology is a fully CIDOC-CRM compliant structure, which aims to provide a simple generic model of the spatial arrangements of the different venues that participate in the four project pilots which captures the provenance of POIs (Points of Interest) movement. The venues of the four pilots can be clustered broadly as indoor and outdoor "exhibitions" of POIs, with similar characteristics: (i) Pilot 1, an indoor gallery with a large multi-thematic collection spread over 66 rooms and 2 floors. (ii) Pilot 2, four open air archaeological sites with location and POIs alterations over the various historical periods starting from the classical period and the Roman times. (iii) Pilot 3, a small museum with dense displays of archaeological exhibits confined in a small number of rooms. (iv) Pilot 4, two whole cities with disperse POIs located on façades of buildings, near bridges, in crossroads, near statues, on top of columns etc. A POI in CrossCult is any physical thing (place or object), either immobile or portable, which is of historical, social or cultural interest, e.g., a painting at the National Gallery, the Asklepieion at Epidaurus or the statue of "The Tall Banker" in Luxembourg.

Although the purposes of the different venues are quite different, they are characterised by similarities that allow the construction of a common model that describes their spatial arrangements.

The semantic representation of the city's structure conceptualised as an outdoor exhibition has similar characteristics to the indoor gallery and the small museum. It is composed of sections filled with other elements; for example, buildings composed of walls, floors, ceilings—that have dimensions and materiality; windows and doorways—spaces that are completely void. In all venues the POIs, within a building or outdoors, are also characterised by events; POIs are moved from one location to another to serve for example the needs of exhibitions. They are also moved to receive treatment or for the needs of rehanging or changing the display of objects at a specific part of the building's structure. Finally, the POIs move as the city's structure changes or as the result of constant alterations throughout time. Historic buildings and archaeological venues are, in most cases, the result of a series of matter addition and removal due to construction and destruction activities that modified their appearance over the various historical periods. The identification of these processes, together with the analysis of the different building techniques and the materials utilised over its existence, provides historians with an understanding of the continuity and discontinuity of matter and activities on a built structure. All these strands of information can be used to produce a detailed understanding of the development of the historical provenance of any building, whether standing or in ruins, and to identify significant phases of the monument's appearance throughout the centuries.

The process of building the Venue ontology involved first developing the appropriate underlying conceptual model to support the requirements of the four venues and, second, populating the model with sufficient detail to realise its full potential. We kept the resulting model as generic as possible and we progressed with the task of populating the model with examples. The data for populating the ontology came from a variety of sources and differed in their underlying structures, accuracy and the level of detail in the representation of the places. Therefore, as more data was included in the process, the model was further specialised to meet the specific needs of each Venue.

The proposed CrossCult Venue ontology attempts to address these emerging data modelling requirements and has been inspired from the CIDOC-CRMba, an extension of CIDOC CRM that has been proposed for approval by CIDOC CRM-SIG to support buildings archaeology documentation[19]. We decided on CIDOC CRM as the integrating framework, as a sensible first step on the road to interoperability. From the modelling process outlined above, we concluded that the resulting Venue ontology does cover the basic needs and characteristics of the four pilot venues in terms of their spatial arrangements. Finally, if we need to scope the needs of all our indoor and outdoor venues in more detail and cater for additional functionalities (for example, model the spatial semantics related to the alterations of buildings that modified their appearance over the various historical periods), then the Venue ontology can be enhanced with additional classes and properties from the CIDOC-CRMba[20]. The CIDOC-CRMba incorporates parts of the CRMgeo, a detailed model of generic spatial-temporal topology and geometric description [10], parts of CRMsci, a model for scientific observation, measurements and processed data in descriptive and empirical sciences (such as biology, geology, geography and cultural heritage conservation) and CRMarcheo, a model developed for the documentation of archaeological excavations.

To address the data modelling requirements discussed in the paragraphs above, we defined the Venue ontology as a subset of CIDOC-CRM. Similar to the Upper- level ontology, the structure maintains full compatibility with CIDOC-CRM containing the least minimum set of CRM concepts as described in the latest specification document version 6.2.3. Figure 4 depicts its graphical representation; Major components of the Venue ontology arrangements are the subclasses of the E18. Physical thing, E19. Physical object, E26. Physical feature and E24. Physical man-made thing, which are used to model physical objects and features as well as man-made structures. Physical thing and Physical man-made thing Instances such as a 'Building', a 'Room', a 'Floor' or a 'Wall'. It can also be combined together to

---

[19] http://icom.museum/resources/publications-database/publication/definition-of-the-crmba-an-extension-of-cidoc-crm-to-support-buildings-archaeology-documentation/print/1/ Accessed January 29, 2019.

[20] http://www.cidoc-crm.org/crmba/sites/default/files/2016-12-3%23CRMba_v1.4.1_UR.pdf Accessed January 29, 2019.

form more complex structures. These classes are further related to other ontology classes to model the physical and man-made structures' dimensions, conditions or events. The class E.55 Type has also been employed to differentiate between the functionalities of a room in a museum as a 'Gallery', a 'Cafe', a 'Temporary exhibition room' etc. Complementary to the notion of the E19. Physical object and E24. Physical man-made thing classes is the E53. Place class, which is used to model the different types of the venue spaces. Place instances can be combined together to form complex spaces, whereas spatial coordinates and appellations are used to model the details of such spaces.
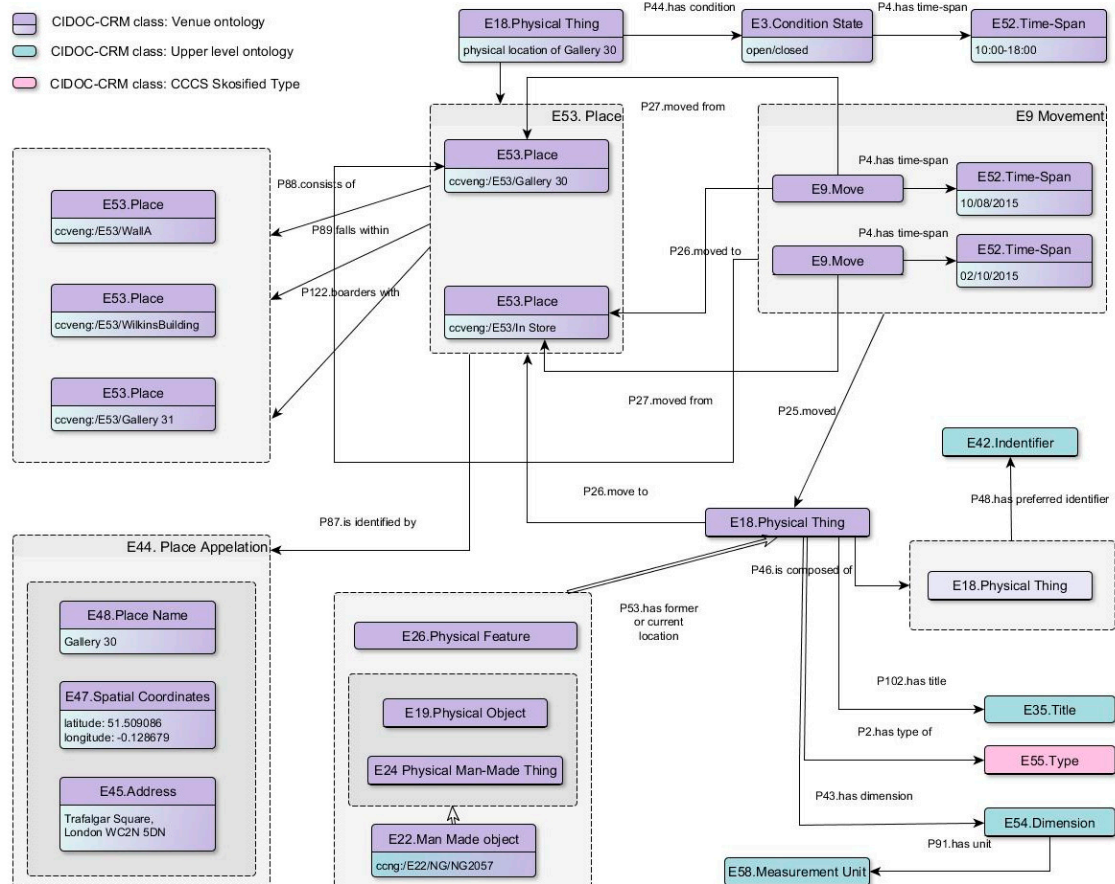
Figure 4. The conceptual model of the CrossCult Venue ontology, demonstrating the documentation of the movement of paintings and recording former or current locations.

We use the E9. Move class to describe changes of the physical location of the instances of E19. Physical object, for example the movement of a painting from one room to another. This class inherits the property P7.took_place_at (witnessed), which has range E53. Place. We use this property to describe the larger area within which a move takes place, whereas the properties P26.moved_to (was_destination_of) and P27.moved_from (was_origin_of) describe the start and end points only. For example, (E9) "Movement of the painting" moved the (E19) "Painting"; (E53) "East Wall location" is the origin of the (E9) "Movement of the painting" and (E53) "West wall location" is the destination of the movement; the (E9) "Movement of the painting" took place at (E53) "the location of Room 9". In some cases, we can also use the P8.took_place_on or within (witnessed) which has range E19. Physical Object. This property is in effect a special case of P7.took_place_at and we can use it to describe, for example, a movement that can be located with respect to the space defined by an E19. Physical Object such as a 'Building', a 'Room' or a 'Wall'.

3. Preparing National Gallery Data

3.1. Internal Data Aggregation

Memory institutions have been working to enrich their cultural resources either by converting them into digital objects or by collecting born digital ones. Various types of metadata, meaning data about data, are created for those resources such as bibliographic information, technology and structure features and preservation information. Characteristic features of metadata are that it "can be embedded in the body of the digital resource, may be a first-class object as well as a primary resource and may be linked to each other in order to produce a richer environment for users to access the resources over the internet" [11]. We consider metadata as a secondary resource created from a primary resource (a painting, a book, a music performance).

The National Gallery (NG) uses a range of systems to hold and manage information about its primary resources. Most forms of documentation within the NG make direct use of or reference these resources, particularly images and metadata. For the CrossCult project the NG needed to provide dynamic access to a full set of painting images and its core (Tombstone) data, retrieved from collection information held in the NG collection management system (CMS) TMS (The Museum System™).

The existing data consisted of the painting details dataset, the artist's details dataset and the images dataset:

- Painting inventory ID: Accession Number, unique painting ID.
- Painting date(s): relevant dates for the painting, including date of production, dates of exhibitions and modifications, etc.
- Artist(s): The name of the artist or artist involved in the production of the painting. This will also include details relating to unknown groups of related artists, such as "Workshop of . . . ", "Follower of . . . " etc.
- Group: Indicates if a given painting is part of a defined group of paintings. The paintings in these groups are normally directly physically related rather than of a similar type. For example paintings that used to be part of the same altarpiece, paintings that were all created as part of one installation, double sided paintings, etc.
- Painting title: Full title of a painting. Additional alternative titles may also be available; a shortened version of the title will also be available.
- Medium and support: Short terms used to describe the main key materials used to create a given painting, for example "Oil on Canvas".
- Painting dimensions: The physical height and widths of a given painting in centimetres.
- Credit line: Were available, details of specific acquisitions credits, including the name and date of a given bequest. This can include details of more than one event and date.
- Public locations: The name or number of the specific Gallery in which the painting is held. All paintings that are not on display are given the generic location of "Not on display".
- Inscription summary: Textural details describing the presence and locations of any specific marks, signatures, dates or more general inscriptions noted on a given painting.
- Classifications and keywords: General type and grouping classification terms, along with more general subject matter related keywords.
- Additional paintings details:
- Description: Short textural description of the painting and its history, drawn from the National Gallery public website content management system.

The Artist details dataset includes:

- Unique Artist ID.
- Artist name: Were possible including know variations and translations of these names.

- Artist date(s): Generally the date of birth and death on an artist, but possibly dates relating to when they were known to be alive, active or when their work was documented.
- Short artist's biography, where available.

The image details dataset provided to the CrossCult system included a full set of 800 pixel images of almost all of the NG acquisitioned paintings (~2300), drawn from an internal bespoke digital asset management system, presented via an IIIF[21] compliant IIP Image server[22].

In order to dynamically re-use all of these resources, an internal Application Programming Interface (API) was developed to present a single, aggregated view of all of the available data and allow direct access to structured linkable data describing the NG Collection (see Figure 5). As a second step and in order to interlink the NG digital information and to share its data unambiguously with external users, the NG established a unique persistent identifier (PID) for every entity referred to by its digital information. A persistent identifier (PI or PID) is a long-lasting generic reference to an image, document, file, web page, or digital description of any physical thing or concept that one might want to describe or discuss. Many things one might want to discuss or refer to already have IDs within existing local databases or catalogue systems. The purpose of a PID system is to provide unique generic identifiers that can be used and reused across multiple systems, particularly in relation to publishing information that can be accessed over the Internet. Finally, a subset of the NG available data has been provided in the form of a basic JSON[23] array (see Table 1), and shared externally using the PIDs through the NG public beta API[24]. The work that is currently underway aims to fully map the data to the CIDOC CRM and provide a standard semantic presentation of the data (an example of this mapping is available in Section 4).
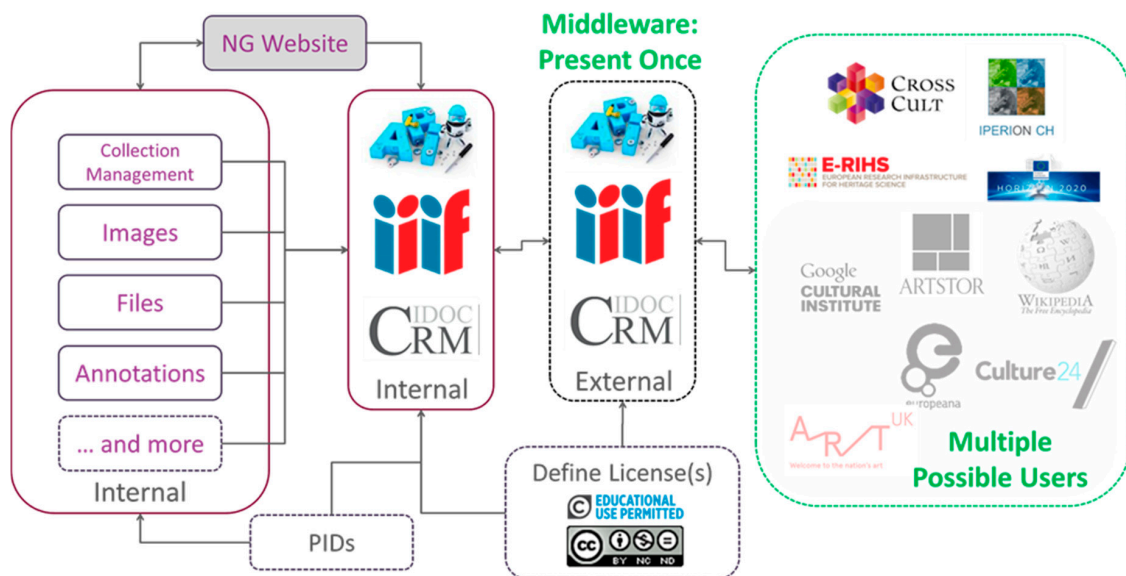
Figure 5. Simplified diagram of the major sources of digital information aggregated within the National Gallery to create the Application Programming Interface (API) used within CrossCult and available to other possible users. PID: persistent identifier.

---

21 International Image Interoperability Framework - https://iiif.io/ Accessed January 29, 2019.
22 http://iipimage.sourceforge.net/documentation/server/ Accessed January 29, 2019.
23 https://www.json.org/ Accessed January 29, 2019.
24 https://data.ng-london.org.uk/resource/examples Accessed January 29, 2019.

Table 1. Simplified example of the JSON data created for Room 30 - 006-001M-0000.

{"type":"location","pid":"006-001M-0000","name":"Room 30","title":"Spain","description":"<p>Spanish painting flourished during the 17th century principally in the service of God and King. ...","objects":{"000-00A8-0000":{"pid":"000-00A8-0000","no":"NG6566"}, ... , ,"example_object":"000-00A8-0000","artists":{"001-01WB-0000":"Italian, Neapolitan","001-03FC-0000":"Jusepe de Ribera", ...},"date_range":{"begin":"1618-01-01","end":"1684-12-31"},"contains":[],"keywords": {"00A-0001-0000":"Religion","00A-0002-0000":"Christianity", ...},"license":"https:\/\/creativecommons.org\/licenses\/by-nc-nd\/4.0\/","attribution": "This data is licensed ... "}

3.2. Extending the National Gallery Core Data—Keywords

Compliant with international IT standards the Getty vocabularies[25] were chosen by the CrossCult project as the main building block of CCCS, the CrossCult vocabulary, as they provide authoritative information for cataloguers, researchers and data providers; they contain structured terminology for art, architecture, decorative arts, archive materials, visual surrogates, conservation and bibliographic materials. These multilingual semantically structured thesauri can be powerful tools for enriching knowledge and providing meaningful links for cultural heritage information resources. As Linked Open Data, the Getty vocabularies are expressed as structured and openly reusable machine-readable data, that information systems can interpret and use to create semantically relevant relationships across other linked datasets [12].

For the needs of Pilot 1 we selected a flat list of approximately 500 keywords, which have been identified and aggregated from internal available NG datasets and a number of resources such as the keywords of the NG picture library[26]. The flat list of the NG keywords was initially cleansed, verified against and linked to the external semantic definitions from the Getty authority vocabularies such as, the Arts and Architecture Thesaurus of Getty (AAT), the Getty Thesaurus of Geographic Names (TNG), the Union List of Artist Names (ULAN), the Cultural Objects Name Authority (CONA) as well as the Conservation & Art Materials Encyclopaedia (CAMEO). On a second stage we incorporated the flat list in the CrossCult Classification Scheme (CCCS). The reuse of standardised resources ensured the validity of the CCCS structure and the consistency in the use of its terms. The connections between the CCCS and the CrossCult Upper-level ontology introduced an augmented view of cultural heritage information enabling further fluency to express interrelations and groupings among physical items, venues, digital resources and concepts.

3.3. Extending National Gallery Painting Location Data

At the NG, paintings move from one location to another, to serve the needs of exhibitions, to receive treatment or for the needs of rehanging and changing the objects displayed in a specific location within the building. Although the record of an object's location is part of its provenance, the time for recording detailed location-based information is limited. Our intention during the CrossCult project was to develop administrative tools that would also allow the basic room data to be augmented to include specific wall and position based information in order to record the changes in location and capture the movement provenance.

The existing location data we had to start with was available from the NG architectural drawings (see Figure 6); the process began by manually extracting the dimensions of the rooms (height and width), walls (height and width) and when available the dimensions of the room's door(s). A relational MySQL database introduced a series of tables that were populated with the existing location data and

would store the generated location data (see Figure 7). At the core of the structure is the ng-location painting_position database table. This table holds data that allows us to capture temporal information related to each movement of an object such as painting_position_date. Supplementary database tables are the ng_location painting which holds the painting dimensions and the ng-location wall_object which holds location data related to wall objects such as skirting boards, cornicing, doors etc.

During the project we developed a game application that allows the end user to move paintings and record their positions, on a given wall within a virtual space and can be used to quickly capture a more precise location for paintings. The "Moving Paintings" as a standalone administrative tool has been created to assist the museum staff to accurately record the positions of paintings when they are moved or re-positioned. In order for the application to use NG resources (images, room information, etc.) a specific data structure needed to be used. The system uses an XML[27] structure which is automatically generated from internal NG data, where metadata of rooms (and their contents), walls (and their dimensions), paintings (and their images), artists and categories are stored. In the current prototype [13], the application consumes the XML structure of a room and its paintings, and downloads images of these paintings to integrate into the application. At the same time the data from the game can be used to populate and update the XML structure and be added to the location database.
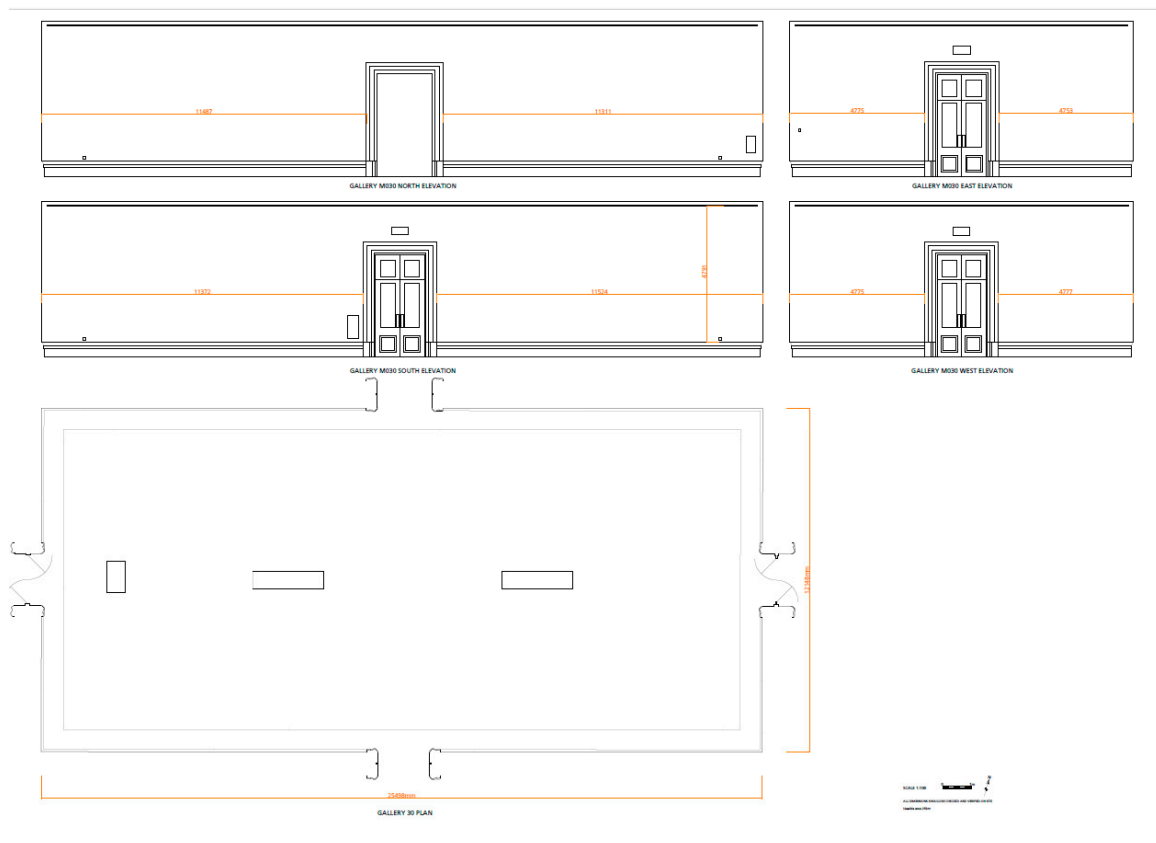
Figure 6. The architecturaldrawing with the available National Gallery London (NG) location data (wall and room dimensions) of Room 30.

---

27 XML stands for eXtensible Markup Language and is a software- and hardware-independent tool for storing and transporting data.

Figure 7. The database tables structure and the relationships formed for storing the NG location data. Simple width and height values are stored in centimetres and the various x, y and z dimensions can be defined as absolute values of latitude, longitude and altitude or as simple values relative to a local zero point for a given institution or location. A graphical representation of the possible absolute and relative positions relating to the location of a painting on a wall can be seen in Figure 8.
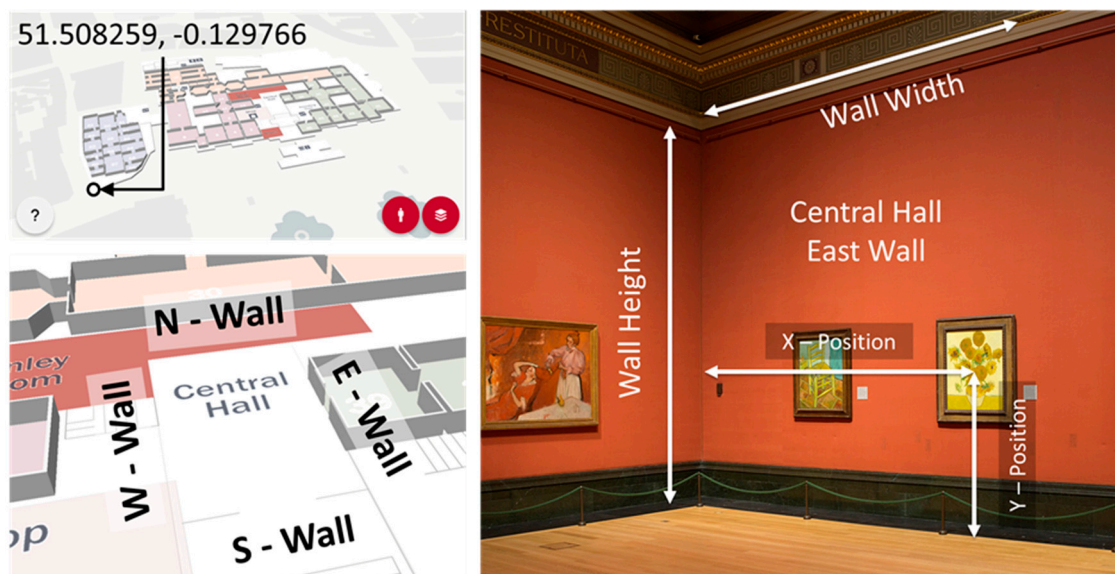
Figure 8. This is a graphical representation of the possible absolute and relative positions of a painting relating to its location on a wall.

4. Mapping NG data to the CIDOC-CRM

As noted in Section 3.1 the internal NG API aggregates information together from a number of sources and formats it as JSON data. The next step was to map this JSON data to the CIDOC CRM, as a series of RDF triples using the work on the upper-level CrossCult ontology, (see Figure 3), as a guide.

A series of specific PHP[28] functions were created to break the JSON down into its component parts and dynamically generate the relevant triples for each of the types of digital object, paintings, location, artist etc. In addition to the PIDs which have been generated for all of the referenced digital objects, blank node have been used to correctly relate the digital objects to specific literal values (see Table 2). This work which aims to fully map the NG data to the CIDOC-CRM and provide a standard semantic presentation is on-going and a number of complete examples, which will be continue to be updated as required, are provided as part of the NG external API[29]. A more complete current example of the triples produced for "Room 30" can be seen in Table 3 while Table 4 presents the triples produced for the object 000-01D6-0000.

Table 2. Simple example of JSON data, for Room 30-006-001M-0000, being mapped to the CIDOC CRM.

JSON	Triples		
{"type":"location","pid":"006-001M-0000","name":"Room 30"}	006-001M-0000	rdf:type crm:P102.has title	crm:E53.Place _:006-001M-0000title
	_:006-001M-0000	rdf:type rdf:label	crm:E35.Title Room 30@en

Table 3. Room Details of 006-001M-0000.

Subject	Predicate	Object			
ng:006-001M-0000	rdfs:comment	Spanish painting flourished during the 17th century principally in the service of God and King. The evolution of a Catholic Counter-Reformation religiosity is revealed in a variety of powerful, individual styles. Not long after El Greco had portrayed the divine with ethereal idealisations of figures, space and light, Diego Velázquez and Francisco de Zurbarán turned to realism to represent the mystical. . . . .@en			

Subject	Predicate	Object	Subject	Predicate	Object
ng:006-001M-0000	cc:licence	https://creativecommons.org/licenses/by-nc-nd/4.0/	ng:006-001M-0000	crm:P102.has title	_:006-001M-0000title
ng:006-001M-0000	rdf:type	crm:E53.Place	_:006-001M-0000	rdf:type	crm:E35.Title
ng:006-001M-0000	rdf:label	Room 30@en	_:006-001M-0000	rdf:label	Room 30@en
ng:006-001M-0000	P89.falls within	ng:006-0033-0000	ng:006-001M-0000	crm:P102.has title	_:006-001M-0000subtitle
ng:006-001M-0000	P89.falls within	ng:006-003S-0000	_:006-001M-0000subtitle	rdf:type	crm:E35.Title
ng:000-00A8-0000	crm:P55.has current location	ng:006-001M-0000	_:006-001M-0000subtitle	rdf:label	Spain@en
ng:000-0196-0000	crm:P55.has current location	ng:006-001M-0000	_:006-001M-0000subtitle	crm:P2.has type	ng:00A-00DU-0000
_:006-001M-0000	crm:P2.has type	ng:00A-00DP-0000	ng:00A-0002-0000	crm:P67.refers to	ng:006-001M-0000

As an example of how this data can be used let us imagine that a visitor is located at the East Wing (006-0033-0000) of the 2nd level (006-003S-0000) at The National Gallery (001-02VZ-0000), in Room 30 (006-001M-0000), where she looks at the painting (000-01D6-0000) 'The Toilet of Venus' or 'The Rokeby Venus' (00C-01C7-0000) made by Velázquez, Diego (1599–1660) as part of her visiting route. The date of artwork is between 1647 and 1951 and the painting provides a notable example of colour change and material degradation [14]. In order to retrieve more information about the painting the user reads a textual description and she views the painting's image[30]. The textual description also contains the following keywords representative of the subject depicted in the painting: People (00A-0003-0000),

---

28   http://php.net/ Accessed January 29, 2019
29   https://data.ng-london.org.uk/resource/examples/rdf Accessed January 29, 2019.
30   https://media.ng-london.org.uk/iiif/examples/009-01S9-0000 Accessed January 29, 2019.

Christianity (00A-0002-0000) and Female (00A-000H-0000). The Room where the painting is located in is also connected to the concept of Christianity (00A-0002-0000). In the same Room the user can view other paintings such as 'A Cup of Water and a Rose' (000-00A8-0000) made by Francisco de Zurbarán and 'The Heavenly and Earthly Trinities'(000-0196-0000) made by Bartolomé Esteban Murillo.

Table 4. Object details of 000-01D6-0000.

Subject	Predicate	Object			
ng:000-01D6-0000	rdfs:comment	This is the only surviving example of a female nude by Velázquez. The subject was rare in Spain because it met with the disapproval of the Church.Venus, the goddess of Love, was the most beautiful of the goddesses, and was regarded as a personification of female beauty [. . .] The painting is known as 'The Rokeby Venus' because it was in the Morritt Collection at Rokeby Park, now in County Durham, before its acquisition by the Gallery. @en			

Subject	Predicate	Object	Subject	Predicate	Object
ng:000-01D6-0000	cc:licence	https:// creativecommons.org/ licenses/by-nc-nd/4.0/	ng:000-01D6-0000	crm:P50. has current keeper	ng:001-02VZ-0000
ng:000-01D6-0000	rdf:type	crm:E22.Man-Made Object	ng:001-02VZ-0000	rdf:type	crm:E39.Actor
ng:000-01D6-0000	rdf:label	The Toilet of Venus ('The Rokeby Venus')@en	ng:001-02VZ-0000	rdf:label	The National Gallery (London)@en
ng:000-01D6-0000	crm:P48. has preferred identifier	_:000-01D6-0000	ng:001-02VZ-0000	crm:P2. has type	ng:00A-00DO-0000
_:000-01D6-0000	rdf:type	crm:E42.Identifier	ng:000-01D6-0000	crm:P55. has current location	ng:006-001M-0000
_:000-01D6-0000	rdf:label	NG2057@en	ng:000-01D6-0000	crm:P102. has title	ng:00C-01C7-0000
_:000-01D6-0000	crm:P2. has type	ng:00A-00DL-0000	ng:00C-01C7-0000	rdf:type	crm:E35.Title
ng:000-01D6-0000	crm:P43. has dimension	_:000-01D6-0000width	ng:00C-01C7-0000	rdf:label	The Toilet of Venus ('The Rokeby Venus')@en
_:000-01D6-0000width	rdf:type	crm:E54.Dimension	ng:00C-01C7-0000	crm:P2.has type	ng:00A-00DQ-0000
_:000-01D6-0000width	crm:P90. has value	177.000 # xsd:decimal	ng:009-01S9-0000	crm:P138. represents	ng:000-01D6-0000
_:000-01D6-0000width	crm:P91. has unit	ng:00A-00DK-0000	ng:002-01D6-0000	crm:P108. has produced	ng:000-01D6-0000
_:000-01D6-0000width	crm:P2.has type	ng:00A-00DM-0000	ng:002-01D6-0000	rdf:type	crm:E12.Production
ng:000-01D6-0000	crm:P43. has dimension	_:000-01D6-0000height	ng:002-01D6-0000	crm:P81.ongoing throughout	_:productionTimeSpan000-01D6-0000
_:000-01D6-0000 height	rdf:type	crm:E54.Dimension	ng:002-01D6-0000	rdf:type	crm:E12.Production
_:000-01D6-0000 height	crm:P90. has value	122.500 # xsd:decimal	_:productionTimeSpan000-01D6-0000	rdf:type	crm:E61.Time Primitive
_:000-01D6-0000 height	crm:P91. has unit	ng:00A-00DK-0000	_:productionTimeSpan000-01D6-0000	rdf:label	1647-51@en
_:000-01D6-0000 height	crm:P2.has type	ng:00A-00DN-0000	ng:00A-0002-0000	crm:P67. refers to	ng:000-01D6-0000
ng:00A-0003-0000	crm:P67. refers to	ng:000-01D6-0000	ng:00A-000H-0000	crm:P67. refers to	ng:000-01D6-0000
ng:000-01D6-0000	crm:P46.is_ composed_of	_:000-01D6-0000medium0	ng:000-01D6-0000	crm:P46.is_ composed_of	_:000-01D6-0000support0
_::000-01D6-0000medium0	rdf:type	crm:E57.Material	_:000-01D6-0000support0	rdf:type	crm:E57.Material
_:000-01D6-0000medium0	rdfs:label	NG2057 Medium@en	_:000-01D6-0000support0	rdfs:label	NG2057 Support@en
_:000-01D6-0000medium0	crm:P2.has_type	ng:00A-00DI-0000	_:000-01D6-0000support0	crm:P2.has_type	ng:00A-00DJ-0000
_:000-01D6-0000medium0	crm:P45.consists_of	ng:00A-00B6-0000	_:000-01D6-0000support0	crm:P45.consists_of	ng:00A-00BU-0000

5. Conclusions

This paper has presented a working example of semantically representing and using cultural heritage information and location provenance data of the National Gallery of London; we have described how we aggregated the available NG data from a large number of internal systems and resources, how we developed the semantic definition of the NG collection information and the tools needed to expose this data to external stakeholders. This semantic definition allowed us to map the inherent relationships and connections between paintings, artists and materials to the CrossCult

Upper-level ontology, using international standards for cultural heritage documentation such as the CIDOC-CRM and the Getty vocabularies. We introduce how the CrossCult project employed CIDOC CRM as the core conceptual component of its semantic knowledge base and we discussed the conceptual model of the CrossCult Venue ontology which demonstrated how the National Gallery can document the movement of the paintings, recording former or current locations and thus providing a location provenance of its collection information. This work has allowed the National Gallery data to be used within the CrossCult project, but has also created the foundations on which this work can continue. Future work will develop the existing processes to fully describe the connection between National Gallery terms and the Getty Vocabularies and go on to consider further sources of less defined National Gallery data, such as geographical locations related to artists and the production events, along with many more keywords which can allow richer connections to external sources of data.

References

1. Hyvönen, E. *Publishing and Using Cultural Heritage Linked Data on the Semantic Web*; Morgan & Claypool: Palo Alto, CA, USA, 2012; pp. 1–159.
2. Oldman, D.; Doerr, M.; de Jong, G.; Norton, B. Realizing lessons of the last 20 years: A manifesto for data provisioning & aggregation services for the digital humanities (a position paper). *D-lib Magazine* **2014**, *20*, 7–8. [CrossRef]
3. Vlachidis, A.; Bikakis, A.; Kyriaki-Manessi, D.; Triantafyllou, I.; Padfield, J.; Kontiza, K. Semantic representation and enrichment of cultural heritage information for fostering reinterpretation and reflection on the European history. In *Digital Cultural Heritage*; Springer: Cham, Switzerland, 2018; pp. 91–103.
4. Doerr, M. The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata. *AI Mag.* **2003**, *24*, 75–92.
5. Kurtz, D.; Parker, G.; Shotton, D.; Klyne, G.; Schroff, F.; Zisserman, A.; Wilks, Y. CLAROS—Bringing Classical Art to a Global Public. *Fifth IEEE Int. Conf. e-Sci.* **2009**, *11*, 20–27. [CrossRef]
6. Oldman, D.; Tanase, D. Reshaping the Knowledge Graph by Connecting Researchers, Data and Practices in ResearchSpace. In Proceedings of the International Semantic Web Conference, Monterey, CA, USA, 8–12 October 2018.
7. Gruber, E.; Bransbourg, G.; Heath, S.; Meadows, A. Linking Roman Coins: Current Work at the American Numismatic Society. In Proceedings of the 40th Conference in Computer Applications and Quantitative Methods in Archaeology, CAA2012, Southampton, UK, 26–29 March 2012.
8. Meghini, C.; Niccolucci, F.; Felicetti, A.; Ronzino, P.; Nurra, F.; Papatheodorou, C.; Gavrilis, D.; Aloia, N.; Binding, C.; Cuy, S.; et al. ARIADNE: A Research Infrastructure for Archaeology. *J. Comput. Cult. Herit.* **2017**, *10*, 1–27. [CrossRef]
9. Vlachidis, A.; Bikakis, A.; Kyriaki-Manessi, D.; Triantafyllou, I.; Antoniou, A. The CrossCult Knowledge Base: A Co-Inhabitant of Cultural Heritage Ontology and Vocabulary Classification. In *New Trends in Databases and Information Systems. ADBIS 2017. Communications in Computer and Information Science*; Kirikova, M., Ed.; Springer: Cham, Switzerland, 2017; pp. 353–362.
10. Doerr, M.; Hiebel, G. Where did the Varus battle take place?—A spatial refinement for the CIDOC CRM ontology. In Proceedings of the Seventh World Archaeological Congress, The Dead Sea, Jordan, 13–18 January 2013.
11. Ruthven, I.; Chowdhury, G.G. *Cultural Heritage Information: Access and Management*; Neal-Schuman, an imprint of the American Library Association: Chicago, IL, USA, 2015.

12. Baca, M.; Gill, M. Encoding Multilingual Knowledge Systems in the Digital Age: The Getty Vocabularies. *Knowl. Organ.* **2015**, *42*, 232–243. [CrossRef]

13. Kontiza, K.; Liapis, A.; Padfield, J. Capturing the Virtual Movement of Paintings: A Game and A Tool. In Proceedings of the Digital Heritage Conference New Realities: Authenticity & Automation in the Digital Age, 3rd International Congress, San Francisco, CA, USA, 26–30 October 2018.

14. Keith, L. Colour change and restoration. In *Colour Change in Paintings*; Gent, A., Rhiannon, C., Dowding, H., Eds.; Archetype Publications Ltd.: London, UK, 2016.