The CIDOC Conceptual Reference Model (CIDOC-CRM): PRIMER
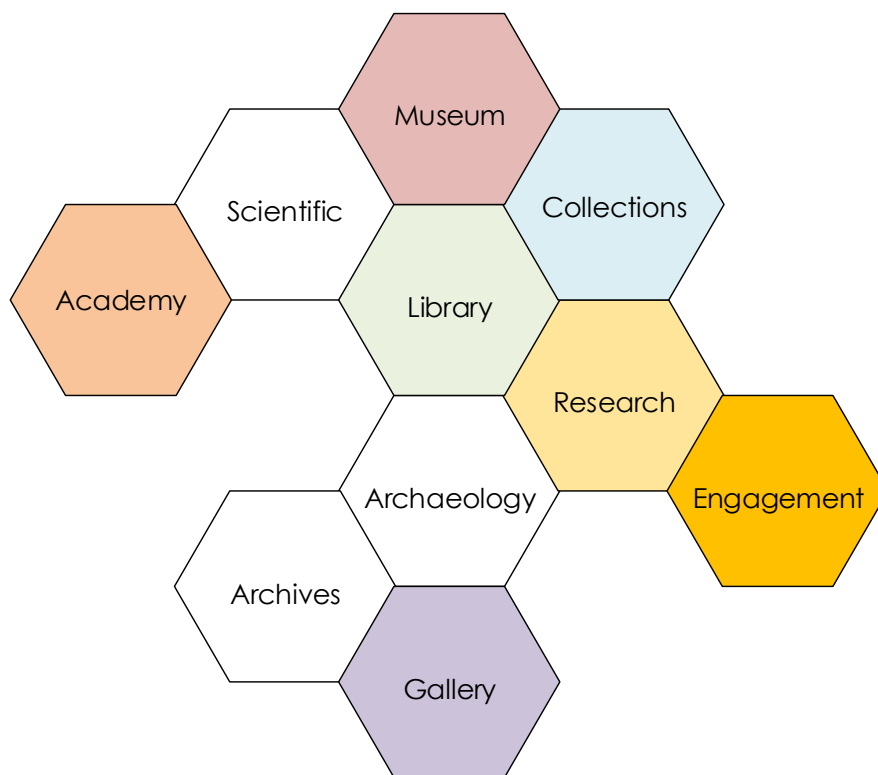
Dominic Oldman[1] and CRM Labs[2]

Edited by Professor Donna Kurtz[3]

In association with CLAROS[4] (Oxford University), CultureBroker[5], CultureCloud, Delving BV[6], Institute of Computer Science, F.O.R.T.H (Heraklion)[7], and ResearchSpace (British Museum[8])

Version 1

July 2014

"The primary role of the CRM is to serve as a basis for mediation of cultural heritage information and thereby provide the semantic 'glue' needed to transform todays disparate, localised information sources into a coherent and valuable global resource."

Nick Crofts

---

[1] Dominic Oldman is Principal Investigator of ResearchSpace at the British Museum

[2] CRM Labs is an initiative of the CIDOC CRM Special Interest Group

[3] Professor Donna Kurtz (Professor of Classical Art and Senior Research Fellow of Oxford e-Research Centre)

[4] www.clarosnet.org

[5] http://culturebroker.eu/

[6] http://www.delving.eu/

[7] Foundation for Research and Technology – Hellas, http://www.ics.forth.gr/

[8] http://www.researchspace.org

Contents

1 Introduction

This document is for cultural heritage managers, professionals, researchers and scholars who need short and concise introductions to new techniques, methods and technologies. Knowledge representation (a way of representing real world things in ways that can be interpreted by computers) is an increasingly important methodology for expressing the richness and variability of cultural data. The CIDOC CRM ontology provides a real world, empirically based representation aimed at harmonizing heterogeneous data. However, the CIDOC CRM method of harmonisation retains the individual nature of the data proving a semantic framework or context that supports the full variability and richness of the information and brings to life the concealed and implicit relationships that exist between things.

It is based on the documentation models and practices of real organisations and provides a full semantic and scientific representation of cultural information. It is independent of any particular technology but is commonly implemented with linked data solutions or as an intellectual guide for designing local information systems and related submission formats to linked data services. Linked Open Data is a method for publishing structured data on the Web (rather than Web pages of information) with the aim of linking it.

This document provides a general understanding of the basic concepts of the CRM and how it is applied. It is not a full explanation of the CIDOC CRM which is referenced in other more comprehensive documentation[9]. While the CIDOC CRM covers a wide range of use cases this guide restricts itself to examples designed to illustrate the most important concepts of the model.

The CRM provides a core ontology that can harmonise between museum, archive, library, and other specialised cultural datasets. More specialist extensions integrated with the core model are also available. These include;

> **FRBRoo[10]** – "is a formal ontology intended to capture and represent the underlying semantics of bibliographic information and to facilitate the integration, mediation, and interchange of bibliographic and museum information. The FRBR model was originally designed as an entity-relationship model by a study group appointed by the International Federation of Library Associations and Institutions (IFLA).

> **CRMSci[11]** – "is a formal ontology intended to be used as a global schema for integrating metadata about scientific observation, measurements and processed data in descriptive and empirical sciences such as biodiversity, geology, geography, archaeology, cultural heritage conservation and others in research IT environments and research data libraries."

> **CRMarchaeo** is an extension of CIDOC CRM aiming to encode metadata about the archaeological excavation process. It is being developed within the framework of the ARIADNE European Research Infrastructure for Archaeology. The goal of this model is to provide the means to document excavations to maximize the interpretation capability, make comparisons between sites, justify the continuation of excavations (find new research questions) and facilitate a range of statistical studies.

---

[9] See http://www.cidoc-crm.org/official_release_cidoc.html

[10] http://www.cidoc-crm.org/frbr_inro.html

[11] http://www.ics.forth.gr/isl/index_main.php?l=e&c=663

2 Background

Until 1998 the CIDOC[12] organisation (the documentation wing of the International Council of Museums, ICOM)[13] had maintained a traditional Entity Relationship model (E-R model) - a modelling system used in the design of relational database systems) of the cultural heritage domain largely derived from work by the Smithsonian Institute[14]. However, the E-R model exposed some major flaws. Its lack of flexibility and semantic capability meant that the model continually expanded to reflect new information requirements and variations, but consequently became too complex; as a result additional areas of practice were increasingly difficult to represent properly and the model became unmaintainable. The CIDOC committee decided to move away from the E–R model and adopt an object-oriented approach. This resulted in an initiative in 1996 to create the CIDOC CRM (Conceptual Reference Model).

The object-oriented model supports a semantically richer (more meaningful) form of representation that is easier to extend sustainably and support a wider range of use cases. It allowed the removal of redundant representations that had accumulated over time in the E-R model, and provided the ability to represent a range of generalisation and specialisation. Although the model itself is object-oriented it can be implemented in any database management system regardless of the underlying model the database system uses. The new CRM model can support and import constructs from any E-R model and improve its semantic characteristics. Conversely, crucial information and meaning are lost translating from the CRM model back to an E-R model and would need additional software in order to simulate the missing semantics[15]. *The primary objective of the CRM initiative was to allow exchange and sharing of information*. The CIDOC CRM is an international standard (ISO 21127:2006) and is maintained by the CRM Special Interest Group (SIG)[16]. The CRM Special Interest Group (SIG) now meet on a regular basis to maintain the standard, resolve issues and incorporate new practice into the model. It is an international and democratic committee open to new proposals from the user community.

3 The CIDOC-CRM Rationale – Significance and Relevance

3.1 A Practical Strategy

The CIDOC CRM creates a framework for data harmonisation. If heterogeneous data sources from different types of cultural heritage organisations can be integrated using a consistent knowledge representation framework, then large scale automated reasoning (the ability to formally manipulate the data using logical rules in order to generate new information) can be applied, creating a highly significant research resource. This type of reasoning has only been achieved to date with small discrete datasets, specially curated and usually in the context of analysing literature (for example the analysis of vocabulary, style, characters, authorship, etc.).

Effectively, the CIDOC CRM transforms cultural heritage data from internal institutional inventories or catalogues into a highly valuable community resource because data accrues greater relevance and significance when harmonised to create densities of information, and also because the process of mapping data (the translation of source model to a target model) to the CRM returns both the meaning and context to the things represented in the data, essential for understanding. *In contributing to this resource of information institutions become important members of a revolutionary digital research community.* Since research is foundational for other cultural heritage activities institutions can increase their research profiles but also transform educational services and produce more interesting ways of

---

[12] International Documentation Committee of ICOM - http://icom.museum/the-committees/international-committees/international-committee/international-committee-for-documentation/
[13] http://icom.museum/
[14] http://www.si.edu /
[15] A formal definition of compatibility with the CRM can be found on page ii of the reference document.
[16] http://www.cidoc-crm.org/who_we_are.html

engaging existing and new audiences at a higher, but accessible, intellectual level. This is a different strategy to that currently being pursued by many cultural heritage institutions that are grappling with resourcing issues, but it provides a more positive response and approach that safeguards the educational and 'memory' role of cultural heritage institutions in society.

4   New Opportunities

In his digital publication, 'Museums, Libraries, and Archives in a Digital Age', G. Wayne Clough, the Secretary of the Smithsonian Institute quotes from a book by Robert Janes[17].

"I will argue that the majority of museums, as social institutions, have largely eschewed on both moral and practical grounds, a broader commitment to the world in which they operate. Instead, they have allowed themselves to be held increasingly captive by the economic imperatives of the marketplace and their own internally driven agendas."

When Janes talks about relevance and resilience he equates this with "innovation" and "progressive museum practice" design to preserve core values in difficult times. The concerns of both Clough and Janes resonate with large numbers of cultural heritage professionals who understand the potential relevance of their institutions in society but have seen this potential gradually eroded as a casualty of unbalanced and short-sighted responses to funding concerns.

The CIDOC CRM is about relevance and resilience. The value and relevance of data increases when it is communicated with its full meaning and context. This relevance is magnified when the knowledge of different institutions is combined to enable different perspectives (shaped by history, location and by different disciplinary concerns) to be preserved. It is further enhanced when these initiatives are built on sustainable infrastructures. Unlike other models used for aggregation CIDOC CRM does not attempt to squeeze cultural information into artificially fixed models that would inevitably misrepresent it. CIDOC CRM provides a semantically richer version of the data compared to its source because employing it involves collaborating with the experts who have produced the data. It provides a basis for the semantic interoperability between different data sources regardless of the subject matter and the classifications that have been applied. *It produces a platform for a powerful harmonisation of archives, libraries and museums (and other specialist research datasets) benefiting both the institutions, scholars and society in general.*

5   What CIDOC CRM is, and what it is not

- The CIDOC CRM is an ontology - a form of knowledge representation. An ontology represents the categorical knowledge within a domain, in this case the cultural heritage domain. The function of a domain ontology is to mediate the variability within a domain and provide a framework under which we can collaborate despite having different datasets – by modelling the constants used in the expert discourse rather than the hypotheses which are produced by experts and are expressed via these constants It is a language, not a statement of current scholarly convictions.
- It is independent of any technical implementation framework. It is commonly employed using Resource Description Framework (RDF) databases, the lingua franca of linked data (see below), but could also be used with other meta-models. Different technologies create a different set of constraints. The design of a knowledge representation system should not be based, or dependent upon, a particular technology. It should represent knowledge in a more generic form.

---

[17] Editor-in-Chief, Museum Management and Curatorship Adjunct Professor, Department of Archaeology, The University of Calgary, Canada

Its only logical restriction is the kind of positive statements information systems can support so far.

- It does not mandate any fields or values. Unlike other standards that work by using an agreed set of fields and/or values the CRM supports variability. The reason why there are so many field/value based standards is because different cultural groups will naturally have different requirements. The CRM provides a semantic framework that describes more general entities (including events) and the relationships between them. *It provides homogeneous access, but does not homogenise data* with respect to the kind of represented content.

- It is an empirically based ontology. Rather than being defined by a committee (top down), the CRM is based on empirical analysis of real practice and local knowledge (bottom up). The CRM develops as a result of understanding existing models of practice that have themselves developed over a considerable period of time; it represents nearly twenty years of international research. It is unlikely that a similar exercise would come up with a significantly different result. *It is scientifically constituted and not influenced by the strength of opinion of a particular group or expert.*

- It is poly-hierarchical (not a flat linear structure) providing an optimal range of generalisation/specialisation above the point of individual institutional terminological descriptions. In such a framework context and semantics become important.

- It does not concern itself with differences in terminology between institutions, it supports the ability to "plugin" local terminologies and provides an ontological framework under which these vocabularies (conceptual terminology) can be compared and linked.

- It provides a framework for matching instances of people, places, things, events and periods using the information and context around these entities. It does not need to rely on primitive string matching techniques.

- It has the ability to support rich computer-based reasoning. The ontology is based on the concept of object-oriented classes with carefully designed relationships that conform to rules of logic. The *CRM provides the opportunity for a computer to infer new information by putting together fragments of information (semantically harmonised) from different sources and creating the conditions in which logical propositions can be concluded.*

- The most important kinds of computer-based reasoning the CRM can support are generalisations of relationships and deductions from highly indirect relations such as what parts have in common with their wholes, what wholes inherit from their parts and what is transferred across meetings and processes of derivation. These are not meant to replace scholarly conclusions but to comprehensively detect facts relevant to answer research questions. Besides others this ensures that highly specialized knowledge stays accessible to generic questions regardless the specificity of representation.

6   Unlocking the CIDOC CRM – Concepts

Concept 1 – Entity Types and Relationships

The CIDOC CRM consists of set of entity types (*real world things*) that can be connected through the use of relationships (also known as *properties*). These relationships have been designed to support computerised reasoning but this ability is dependent on using relationships correctly and with the correct entity types. *Therefore understanding the initial mapping process is very important*. For example, the CRM relationship, **"carried out by"** can only be used between an "**Activity**" entity type and an "**Actor**" (Person or Group) entity type. The short labels used for relationships and entity types can inevitably be misinterpreted and therefore full and precise definitions are given in the CIDOC CRM reference[18].

---

[18] http://www.cidoc-crm.org/official_release_cidoc.html

<u>Concept 2 – CRM is already embedded in cultural data</u>

CIDOC CRM entity types and relationships have been abstracted as generalisations from large numbers of cultural heritage data models and by talking to significant numbers of cultural heritage experts through meetings and workshops, and also directly by visiting different institutions and exchanging knowledge. The entity types and relationships described by the CRM are already embedded within internal collection/information systems and in the knowledge of local experts. Not all of them will apply to a particular institution and dataset. *Using the CRM requires the institution to identify and map the entity types and relationships that apply to their data.* This can be done manually but gradually new tools are being designed to help with this process and share practices.

Concept 3 – Persistent Things and Temporary Things

Consider this representation at the top of the CRM hierarchy;
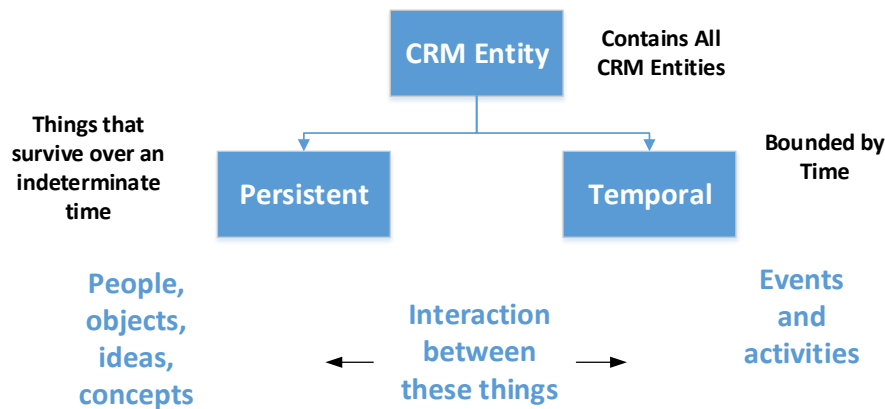
There are the two key (disjoint) branches of the CIDOC CRM. There are things that have a persistent identity that can by their nature survive one or more events (physical things or ideas and concepts), and there are the temporal concepts (or phenomena) that have a nature of happening rather than being, over a limited time frame (an event or activity, like creation, or a historical period). Persistent entity types define instances that are initiators, recipients or witnesses of events and activities. They may originate, survive or terminate in events. This is the essence of the CIDOC CRM's event based model.
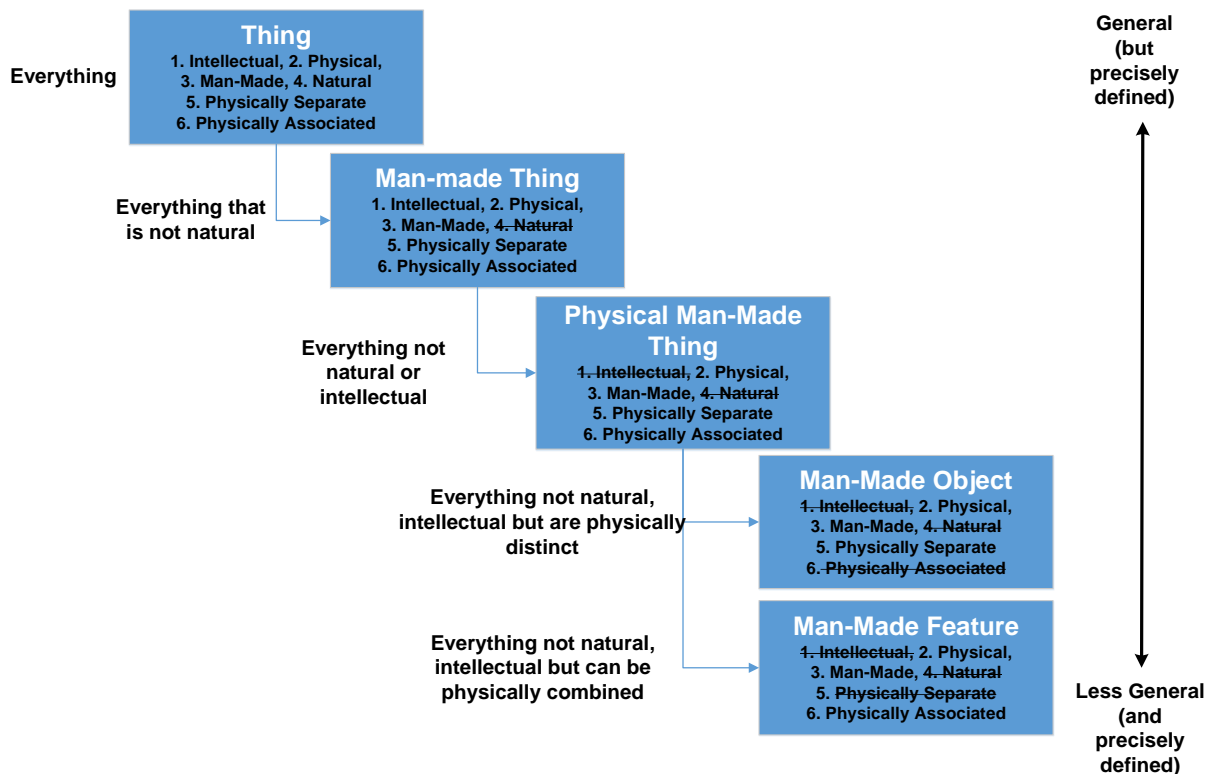
For example, a person's identity endures regardless of his/her death. Death is a temporal concept just as events that occurred during a person's physical life are also examples of things bounded by some period of time. Leonardo da Vinci no longer exists physically but his identity survives. The sinking of the Titanic was a temporal event; although the ship sank to the bottom of the Atlantic Ocean, the identity of the Titanic lives on.

For example, we have sources that say that Pope Leo I met Attila 452AD. Since Pope Leo I is a person, he is a persistent item, and his beginning of existence is a birth. Since he met someone 452, his beginning of existence is his birth and must have happened well before 452AD, as well as Attila's birth, if the source is correct. Since Pope Leo met Attila, they must have been at the same place at least once in their life. Even if the source is wrong about the date, we have other sources that claim that Attila died 453AD. Since being dead he cannot have been negotiating, the meeting cannot have taken place after 453AD. Since Pope Leo I met Attila, the interpretation is likely that Attila's subsequent retreat was influenced by information the Pope gave to Attila. Pope, Attila and the exchanged information exist with a continued structure beyond the meeting. Therefore they must be persistent items. Meetings, birth and death "happen", they are processes we recognize by result rather than structure, therefor they must be temporal.

Concept 4 – The CRM is a Hierarchy of Entities (In fact it is poly-hierarchical

Entity types and relationships both exist in a hierarchy of meanings that provide different levels of generalisation (or specialisation depending on the way you look at it). This is important because we cannot always be precise about everything we want to describe, but when we can, we should. However, there is a point at which specialisations cease to become useful for harmonising data and where institutions might disagree. The diagram below shows how this hierarchy works. Entity types have sub types that are increasingly more specialised. Computer scientists call these classes and sub classes. Take a general entity like **Thing.** A 'CRM' **Thing** refers to things that have a stability of form, it could be man-made or natural, physical or intellectual, a feature of something else or a distinct object. If we know little about a **Thing** then we might use just this broader definition.

Thing
1. Intellectual, 2. Physical,
3. Man-Made, 4. Natural
5. Physically Separate
6. Physically Associated

Everything

Everything that is not natural →

Man-made Thing
1. Intellectual, 2. Physical,
3. Man-Made, 4. Natural
5. Physically Separate
6. Physically Associated

Everything not natural or intellectual →

Physical Man-Made Thing
1. Intellectual, 2. Physical,
3. Man-Made, 4. Natural
5. Physically Separate
6. Physically Associated

Everything not natural, intellectual but are physically distinct →

Man-Made Object
1. Intellectual, 2. Physical,
3. Man-Made, 4. Natural
5. Physically Separate
6. Physically Associated

Everything not natural, intellectual but can be physically combined →

Man-Made Feature
1. Intellectual, 2. Physical,
3. Man-Made, 4. Natural
5. Physically Separate
6. Physically Associated

General (but precisely defined)

Less General (and precisely defined)

When talking about a **Thing** we are talking about the entity type '**Thing**' and all the sub-entities (sub-classes) underneath it in the hierarchy. Mapping data requires choosing the level in the hierarchy that reflects the information available about the **Thing** being described. If there is uncertainty the process can begin broadly and, as more information becomes available, become more precise. Note that other specialisations exist, like **Biological Object or Information Carrier**, for example.

The key to understanding the CRM is understanding its structure - an understanding of the entity types and properties (and their descriptions through "scope notes"), the framework of relationships and how those relationships can be applied to the entity types, and applying these to an organisation's understanding of their own data. This is not a technological undertaking.

Concept 5 – Use as little as you want – extend if you need to

A frequent misunderstanding is that people think they have to implement the whole of the CRM - you do not. You may implement only a few entities and relationships and that is fine. Using CRM encoding never means that a property not used does not apply to the thing you describe. A property not used is either not applicable or unknown, which makes no difference. A CRM compatible description may consist of one relation only. The question is, if this is a correct statement, not if it is complete. At the level of information integration, we have to assume that scholars provide what they know or regard as relevant. It is the task of a research methodology, and not of a data integration standard, to require fields. In general, cultural-historical information has no intrinsic completeness. Individual collections have completely different requirements of completeness according the individual availability of evidence.

It is worth noting however, that if you mapped the entire E-R model of a medium or large sized museum to the CRM, the resulting CRM implementation would have fewer entities (and be a smaller model compared to the source) – but be semantically richer and convey more information. Conversely, because the CIDOC CRM is an object-oriented model if you want to introduce specialisation for your organisations, you can extend the CRM by creating new sub-entities and sub-properties.

Concept 6 – Standard Compatible Knowledge Patterns

Not only is cultural data variable in terms of its institutional documentation, but it also differs in the levels of knowledge it represents. The CIDOC CRM allows standard patterns to be defined, for ease of use, that reflect different levels of knowledge and these patterns themselves make use of the same framework to ensure harmonisation between them. In other words, the same semantic and contextual framework that allows data to be harmonised by the CRM, also ensures that different mapping patterns, that represent the same types of knowledge, are also compatible.

For example, a museum records the historic acquisition activity (purchases, donations, loans, etc.) for the objects it now curates. However, sometimes the collection system fails to provide information about how these different acquisitions relate to each other. There is no documented chain of these acquisitions in the database – just isolated facts. In this case CRM patterns exist that allow each acquisition to have its own individual acquisition event, reflecting the level of knowledge available. Another organisation may have all the information about the chain of acquisitions and can use a pattern with one acquisition event and use the appropriate CRM relationships to reflect the chain of acquisitions under it. Both patterns semantically reflect the knowledge correctly. The different knowledge patterns use the same entities and relationships but in a slightly different standard pattern (a single acquisition event pattern and a multiple acquisition event pattern). However, a single query is able to return acquisition information of across both datasets regardless of the CRM knowledge pattern that is used. The ability to define these different standard patterns for different levels of knowledge but to retain the ability to query over all of them is a concept that is difficult to understand if you are used to fixed field models, but simply adds to the CIDOC CRM's effectiveness because it supports both the stability of mapping (standard patterns) as well as flexibility for different knowledge levels. This ability is often misinterpreted as a lack of determinism and standardization and as an obstacle to information integration. It is just the opposite.

7 The CRM Top Level

This is the upper level of the CIDOC CRM.
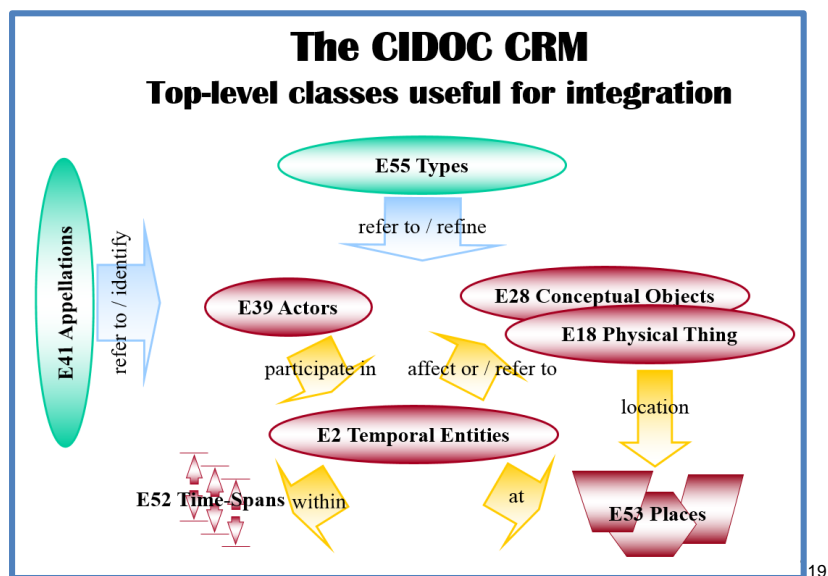
If you understand this then you understand the CRM because everything else is a specialisation of this top level.

---

[19] Diagram from training video by Steve Stead (**Paveprime** Ltd)

- The CIDOC CRM is event based. At the core of this event model are **Temporal Entities** (E2) - things that have happened in the past.

- Only **Temporal Entities** can be linked to time and have **Time Spans** (E52). **Objects** (**Conceptual** (E28) and **Physical** (E18), **Actors/People** (E39), and **Places** (E53) cannot be directly linked to time. Therefore they must be linked to an event – a **Temporal Entity** (E2).

- A **Place** (E53) could be a geographical location on earth, but equally it could be a location defined as the front of a ship or the inside of a ring. These are places that are geometrically defined.

- **Actors** (E39) are entities with legal responsibility and an actor could be an individual or a group, for example, a school of artists or a company, and so on. Actors interact with things – both **Physical Things** (E18) and **Conceptual Things** (E28).

- **Physical Things** (E18) are destroyed when they cease to be functional in the sense of our domain of documentation and therefore destruction is not necessarily linked to physically disappearing. A thing could be physically destroyed and transformed (created) into something else preserving parts of it. That new thing then becomes part of our domain of interest.

- **Conceptual Objects** (E28) cannot be destroyed unless all carriers of it are destroyed. A carrier could be a book, a computer disk, a painting, etc., but it could also be the human mind. So destroying a conceptual object requires destroying all of its carriers, including people.

- It is very common to apply names to things and this is an **Appellation** (E41). An object title and inventory number are forms of appellation. The CRM allows us to name anything. Things can have multiple names and these names can change over time as a result of an event. This means that the use and application of names can be studied over periods of time. A thing and its name are separate entities.

- Different organisations have different classification systems. In the CRM classifications are called **Types** (E55). Again, any number of types can be applied as many times as we like. Because a type is also a **Conceptual Object** (E28) we can also discuss the classification of things over time and the history of definition and redefinition of types.

This is essentially the CIDOC CRM.

8   Why is CIDOC CRM often used with Linked Data?

The W3C's[20] Resource Description Framework (RDF) is the standard for Linked Data and provides a natural fit for CRM described graphs. Whenever two entities are related there is a direction to that relationship and a property used to describe it. For example, an object "**is identified by**" an identifier and not vice versa. *RDF specifies each relationship and ultimately captures an entire CRM based model as triples (statements consisting of a subject, predicate, and object)*. There are a number of standardised representations (file formats) that are used to transmit RDF over a network that facilitate the exchange of CRM based models. Linked Data provides the technical conventions for CIDOC CRM data to be exposed as conveniently as Web pages.

RDF comes with a type of database management system that supports these statements natively and additionally provides a standardised (SPARQL[21]) means for querying CRM data. RDF is the basis for a stack of other standard technologies supported by many of these databases that facilitate the reasoning capabilities (the ability to infer additional knowledge by asserting different logical rules) inherent in CRM models.

---

[20]   World Wide Web Consortium - http://www.w3.org/

[21]   SPARQL – **S**PARQL **P**rotocol **a**nd **R**DF **Q**uery **L**anguage - http://en.wikipedia.org/wiki/SPARQL

9 CIDOC CRM and Resource Description Format - Implementation

9.1 Entities and Relationships

The CIDOC CRM defines entity types and gives these short labels using the prefix of '**E'**. For example,

E22_Man-made_Object

Entity labels capitalise the first letter of each word. Relationships (or Properties) use the prefix '**P'**, again with a label, for example,

P1_is_identified_by

Property labels are lowercase. Labels are convenient placeholders but the label does not (and cannot) convey the full meaning. They may be translated to other languages. Therefore the identity of the concept lies in the short label. To understand what any entity or relationship (property) actually means you need to refer to the scope note contained within the CIDOC CRM reference manual [22] and understand the context in which they can be used. Along with the scope notes, the 'domain and range' (rules for what relationships can be used with certain entities) of the relationships, provide the information you need to map your data.

The scope note for **E22_Man-made_Object** is straight forward and says,

"This class comprises physical objects purposely created by human activity."

The scope note for **P1_is_identified_by** says.

"*This property describes the naming or identification of any real world item by a name or any other identifier…*"

P1 is valid for any entity type, from **E1_CRM_Entity** at the top of the hierarchy all the way down the hierarchy (because any entity can have an identifier). However, the target must be **E41_Appellation** (or sub entities of **E41**, e.g. **E35_Title**, **E42_Identifier**, **E44_Place Appellation**, **E49_Time Appellation** etc.). Appellation is a more general entity and "Title" or "Place Appellation" are more specialised, reflecting real world practice.

9.2 Example of using Entities with Properties with RDF

A CRM statement consists of entities that are related using properties. For example,

Entity	Relationship / Property	Entity
E22_Man-Made_Object	P1_is_identified_by	E42_Identifier

As mentioned above, actual implementations of the CRM will be different depending upon the metadata model used. CRM can be implemented in many different database formats. Commonly, it is implemented in RDF (Resource Description Format) the language of linked data. In RDF, resource information is given a unique URI (Uniform Resource Identifier) including CRM entities and properties. An object will be represented by a URI and statements about it take the form of what technical experts call a triple statement. An object with the URI http://collection.amuseum.org/id/object/1234 can be identified as a particular type of CRM entity using the RDF statement '**rdf:type**'.

---

[22] http://cidoc-crm.org/official_release_cidoc.html

| Subject (Entity)	Predicate (Property/Relationship)	Object (Entity)
http://collection.amuseum.org/id/object/1234	rdf:type	E22_Man-Made_Object

Once the type of the resource is established then CRM relationships can be applied.

| Subject (Entity)	Predicate (Property/Relationship)	Object (Entity)
http://collection.amuseum.org/id/object/1234	P1_is_identified_by	http://collection.amuseum.org/id/object/1234/accession number

The target a relationships also needs to be identified as a CRM entity type.

Subject (Entity)	Predicate (Property/Relationship)	Object (Entity)
http://collection.amuseum.org/id/object/1234/accessionnumber	rdf:type	E42_Identifier

Here is a graphical representation using the CRM view of a BM object.
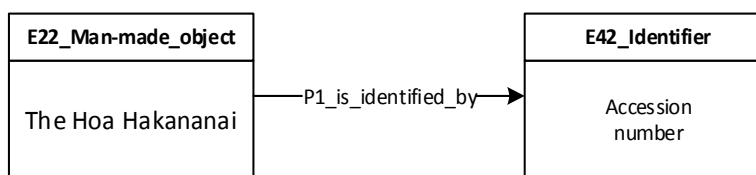
Here is the RDF view.
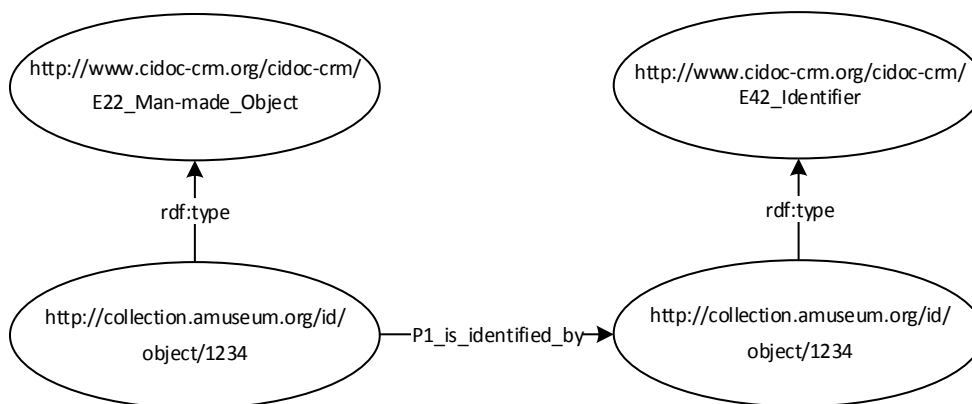
An object, the Hoa Hakananai (an Easter Island statue), represented by a URI, has a type '**Man-made object'**, also represented by a URI. The relationship **rdf:type** is also a URI that has been shortened by substituting a prefix for the main and consistent part of the address. 'rdf:type' is actually;

http://www.w3.org/1999/02/22-rdf-syntax-ns#type

 Prefixes are used by replacing http://www.w3.org/1999/02/22-rdf-syntax-ns# with "rdf:" creating "rdf:type"

9.3   URI Schema

Resources (like an object or an identifier) are assigned a URI providing a logical structure when implementing RDF. URI schemata are created to reflect the resources in question. Here are some examples:

http://collection.[domain]/id/object/[idenitifier],
http://collection.[domain]/id/object/[idenitifier]/title,
http://collection.[domain]/id/object/[idenitifier]/production,
http://collection.[domain]/id/object/[idenitifier]/acquisition,
http://collection.[domain]/id/object/[idenitifier]/material,
http://collection.[domain]/id/thesauri/, and so on.

Domain, for example, could be, "britishmuseum.org".

Therefore, when defining the acquisition of an object:

Subject (Entity)	Predicate (Property/Relationship)	Object (Entity)
http://collection.amuseum.org/id/object/1234	P24i_changed_ownership_through	http://collection.amuseum.org/id/object/1234/acquistion

Your new URI to represent the acquisition of object 1 needs to be identified as an acquisition.

Subject (Entity)	Predicate (Property/Relationship)	Object (Entity)
http://collection.amuseum.org/id/object/1234/acquistion	rdf:type	E8_Acquistion

With an acquisition resource created other information can be attached to it and further acquisition information can be added. For example,

Subject (Entity)	Predicate (Property/Relationship)	Object (Entity)
http://collection.amuseum.org/id/object/1234/acquistion	P14_carried_out_by	http://collection.amuseum.org/id/person/3456

The person resource can then be defined using your person authority.

Subject (Entity)	Predicate (Property/Relationship)	Object (Entity)
http://collection.amuseum.org/id/person/3456	rdf:type	E21_Person

9.4 Comprehensive Digital Representation

Rather than describing a limited number of common fields, as in many digital representations, the CRM describes objects more fully (reflecting full institutional documentation) including the semantic meaning of the data. An aggregator may define a field called 'description' as a target for some object text, but organisations have different types of object description that are created for different purposes, by different people and, therefore, may be interpreted and represented differently. These different perspectives are understood by organisations internally and may or may not be clear in digital collection management systems. Either way the CIDOC CRM provides the means to transfer that internal meaning and provide a more meaningful way of integrating data.

General software developers (people who create applications and web sites) are rarely museum documentation experts. It is far easier to define a set of fields based on a more superficial understanding and ask organisations to approximate to them. This creates overly generalised data integrations that have limited use. *Essentially the representation of data has been left to groups that do not have the necessary understanding and without engagement from those who do.* This is why true digital representation of cultural heritage data requires a positive collaboration with those with local and expert knowledge of the data. We wouldn't produce a physical exhibition without providing the intellectual context for it, but we seem content to publish data without any concerns about its interpretation and context. This essentially undermines and erodes the purpose of cultural heritage organisations as centres of knowledge.

10 Terminological concepts

10.1 Harmonisation and Concepts

The primary purpose of the CRM is semantic data harmonisation that retains source perspectives and heterogeneity. The variations in terminology and its use in different organisations mean that terminological ontologies are not useful for cultural heritage data integration. By freeing itself from the need to support 'every' individual terminological concept contained in 'every' authority of 'all' organisations, the CRM can use its semantic framework to provide ***better*** support for different terminologies. Recognising these differences in perspective between organisations creates a far richer resource than enforcing a homogenisation of terminology. Note that we are talking here about concepts (descriptive terminology). Instances of things, people and places are not terminological concepts but are often treated as such in digital representations. In CRM, conceptual terminologies are related in the model using the entity type **E55_Type** and this very often points to a SKOS[23] construct, used for representing terminological authorities and thesauri[24] (mechanisms for controlling values) in linked data.

10.2 Representing Perspective

As indicated in the examples above institutional terminologies can be used to enhance the mapping of data without affecting the harmonisation framework, and they add a layer of richness. The CRM relationship that achieves this, "**P2_has_type",** is most commonly used to describe or 'type' an event. For example, the CRM provides the entity type **E12_Production.** This provides a point of harmonisation with other datasets with production events. The output of a search across different datasets may return the production events for objects that conform to particular search criteria (type of object, people involved, etc.). This output can be delivered with the additional local vocabulary information encoded

---

[23] Simple Knowledge Organization System is a W3C recommendation designed for representation of thesauri, classification schemes, taxonomies, subject-heading systems, or any other type of structured controlled vocabulary. Wikipedia (http://en.wikipedia.org/wiki/Simple_Knowledge_Organization_System)
[24] A thesaurus is a hierarchical terminology system with broader and narrower alternatives).

against those production events using **P2_has_type**. This provides a range of interesting subtleties and nuances that are the hallmark cultural heritage data, provided by the individual characteristics of different institutional perspectives (e.g. terminology). For production events these are often internal association codes designed to record significant internal information. In this example, different local terms (stored in a SKOS[25] model) can be used to type the production event as a whole.
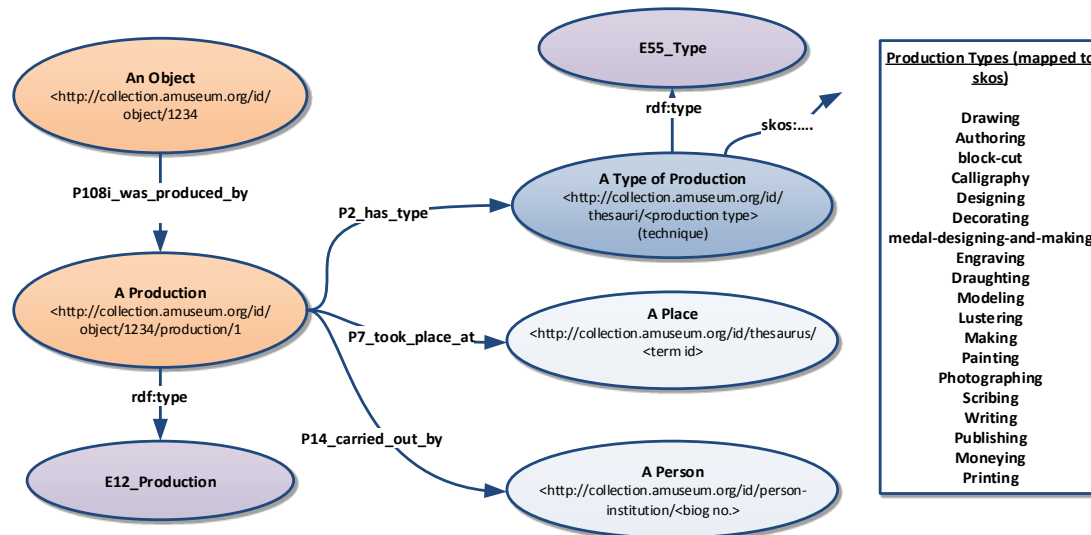
Terminology can also be used is to reify (provide additional definition) to CRM properties themselves. Taking production as an example again, the general relationship, **P14_carried_out_by** is used to provide a point of harmonisation with other CRM datasets. Additionally the CRM provides an alternative property **P17_was_motivated_by.** Different organisations may have different qualifications for these relationships. For example, the British Museum association code, "Made For" clarifies that the motivation was to produce something specifically for a particular person or group. Whereas "Authorised by" is a different type of motivation. In these cases where the qualification is specific to a relationship, local vocabularies can be used directly with a CRM relationship. In the example below, internal vocabularies, "Governor", "Issuer", etc. are used to reify the relationship **P17_was_motivated_by**. A standard pattern is used to *associate* local terms with the CRM relationship and this is used in all examples of reification.

---

[25] A model designed to support classification systems. http://www.w3.org/2004/02/skos/
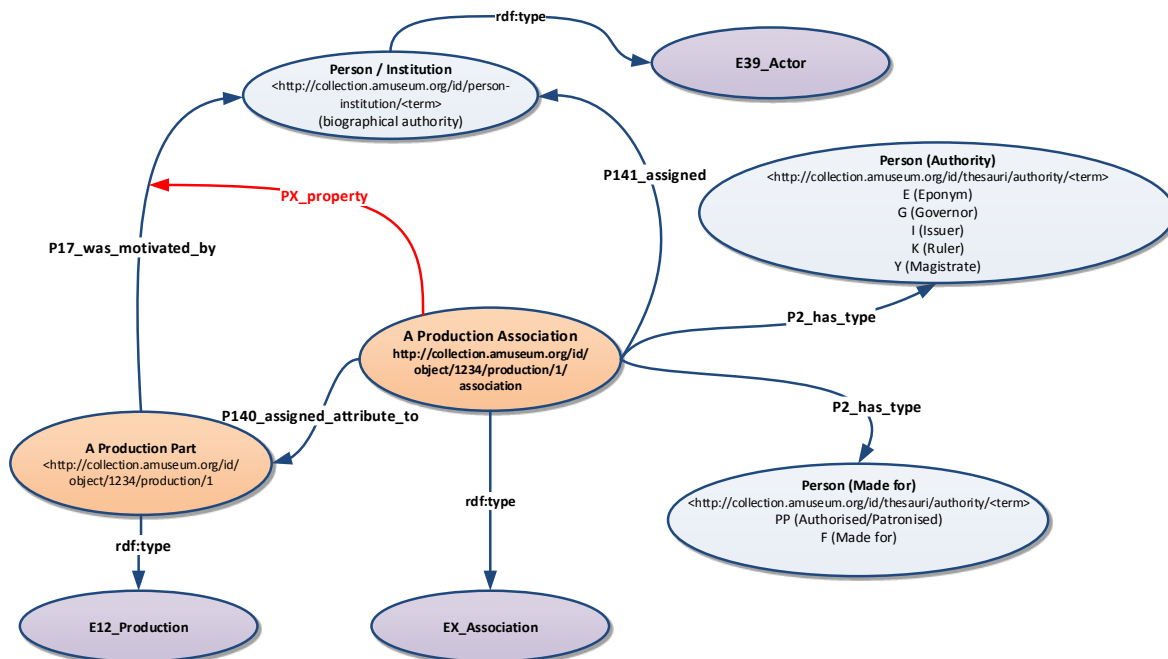
10.3   The Power of Big Data (large, complex datasets)

The ability to harmonise data creates resources of information that are far larger than those humanities researchers been used to ('big data'). The semantic framework provided by the CIDOC CRM is crucial for creating an environment for traversing these large datasets, providing the ability to reason and model (analyse computationally) them and additionally transform the data into a form suitable for a wide range of different kinds of reuse and services. Data modelling and reasoning have many uses but to illustrate why it is so important that an ontology does not impose a fixed set of values consider how the British Museum, for example, describes similar things compared with other institutions using different terminological descriptions. Sometimes one organisation may have more information than others about a particular thing, for example a person. When data are combined from different sources a density of information is constructed about particular things. The CIDOC CRM framework can use this density to infer that different organisations are talking about the same thing and then build up a picture using pieces of information derived from different sources. These connections might be made using many intermediate semantic connections using the CRM framework. This means that you can identity data about similar things even if the terminology used in the different record is different. These differences (which are preserved) are crucial in our full understanding of culture because history is made up of different and subjective perspectives.

11 Next Steps

The CIDOC CRM community is growing and the more people and organisations that join, the richer, more important and significant cultural data becomes. If you would like to find out more about the CIDOC CRM and how to implement it, please go to the CRM web site at http://www.cidoc-crm.org. An introduction to the CRM with links to resources is available at, http://www.cidoc-crm.org/comprehensive_intro.html.

For training enquiries, please email training@cidoc-crm.org (training AT cidoc-crm DOT org). An online training video is available at http://www.cidoc-crm.org/cidoc_tutorial/index.html.

For a high level statement on the CRM, data aggregation and reuse, please refer to the article,

12 **Further reading**

1. Doerr, M. & Crofts, N (1998). Electronic Esperanto—The Role of the oo CIDOC Reference Model (http://www.cidoc-crm.org/docs/doerr_crofts_ichim99_new.pdf)

2. Doerr, M., Dolores, I. (2008). The Dream of a Global Knowledge network—A New Approach." *Journal on Computing and Cultural Heritage* 1 (June 1, 2008): 1–23. doi:10.1145/1367080.1367085.

3. Crofts, N., Doerr, M., Gill, T., Stead, S., and Stiff, M., eds. (2011). Definition of the CIDOC Conceptual Reference Model - cidoc_crm_version_5.0.4. (http://www.cidoc-crm.org/docs/cidoc_crm_version_5.0.4.pdf).

4. Oldman, D., Doerr, M., de Jong, Gerald, Norton, B., Wikman, T., (2014). Realizing Lessons of the Last 20 Years: A Manifesto for Data Provisioning & Aggregation Services for the Digital Humanities (A Position Paper), D-LIB July/August 2014. (see http://www.dlib.org/ - July/Aug 2014 edition).

Annex 1 – Selection of other Examples

Introduction

The patterns below are simple examples to illustrate some of the principles. They are patterns that can be additional enriched with other entities and relationships that reflect other aspects of the events shown. For example production events can host a whole range of different information and relationships.

Acquisition Example

This pattern is based on the relationship P23_transferred_from. For example, transferred from, purchased from, donated by, etc. Different acquisition patterns revolve around other CRM relationships like P14_carried_out_by (for acquisitions through an agent), or P11_participated_in (for acquisitions financed indirectly), and so on. Acquisition involve custody and ownership. When an object is acquired both custody and title are transferred (for a loan only custody is transferred). It has a new owner and a keeper. The acquisition event transfers title of the object from another person or group. The event is typed by local concepts.
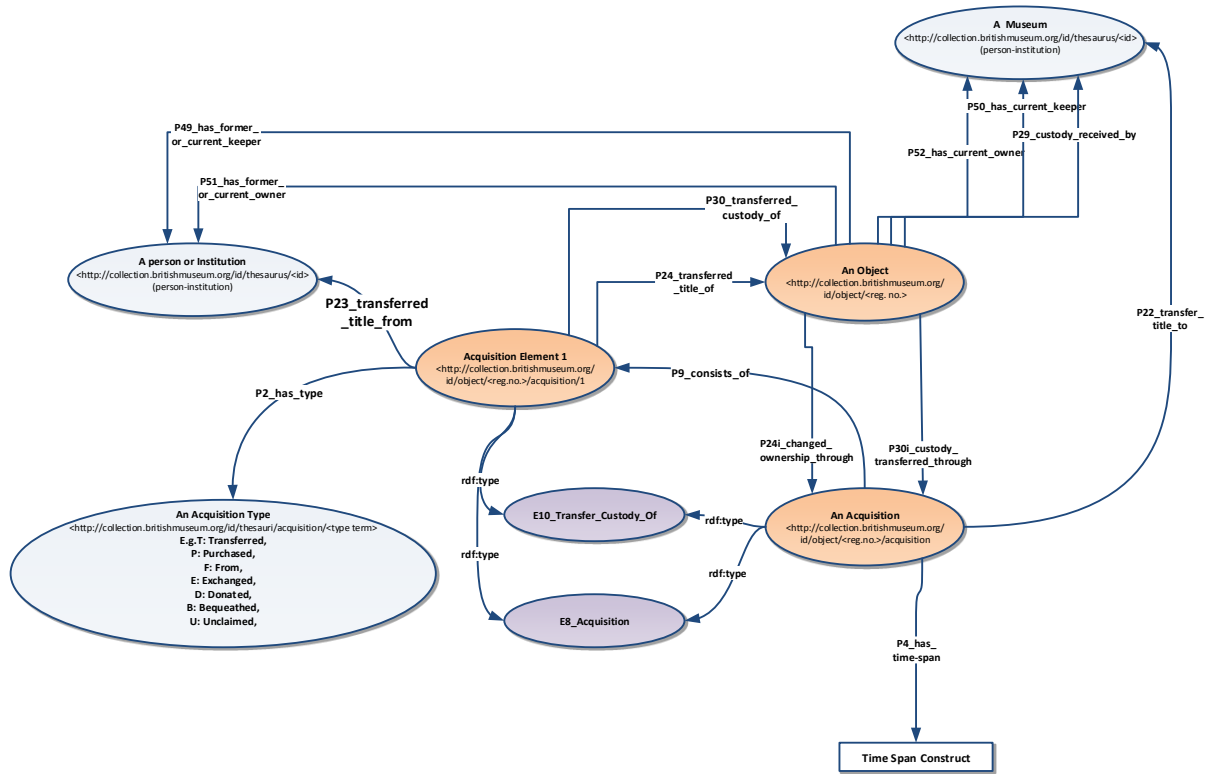
Production Example

An object was produced by a production event. These are divided into sub events for recording different production information. In this case the production fell within a named period e.g. "Medieval" (a thesaurus term). Another sub event records the production technique, again making use of a local concept thesaurus.
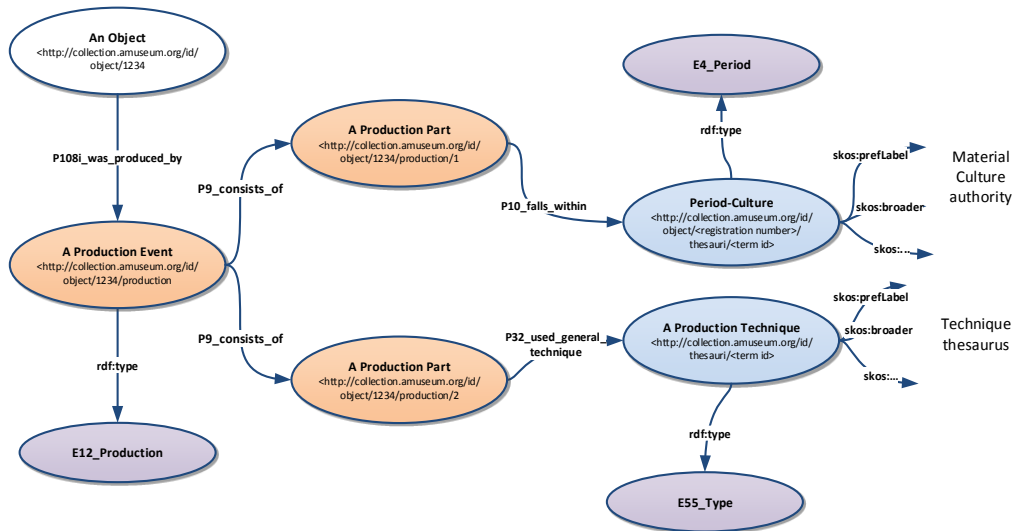

Inscription

This example is part of a construct for an inscription. An object shows a visual item that is typed as a CRM inscription. It has a language that is part of a language authority. It has a translation and the creation was carried out by a person using a method described by local concepts, e.g. engraving.
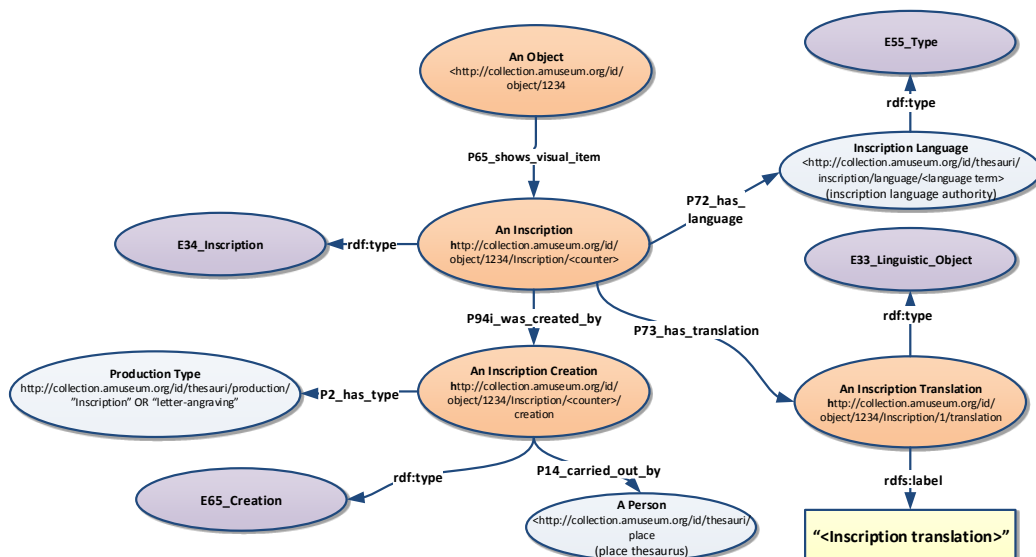
Depiction of Visual Items

An object may have illustrations that depict, people, places or groups, etc. These representation also have an internal association code to provide more information about the type or reason for the visual item, e.g. emblem. There is also a shortcut relationship called, depicts. A place may be a modern name or an archaic name.
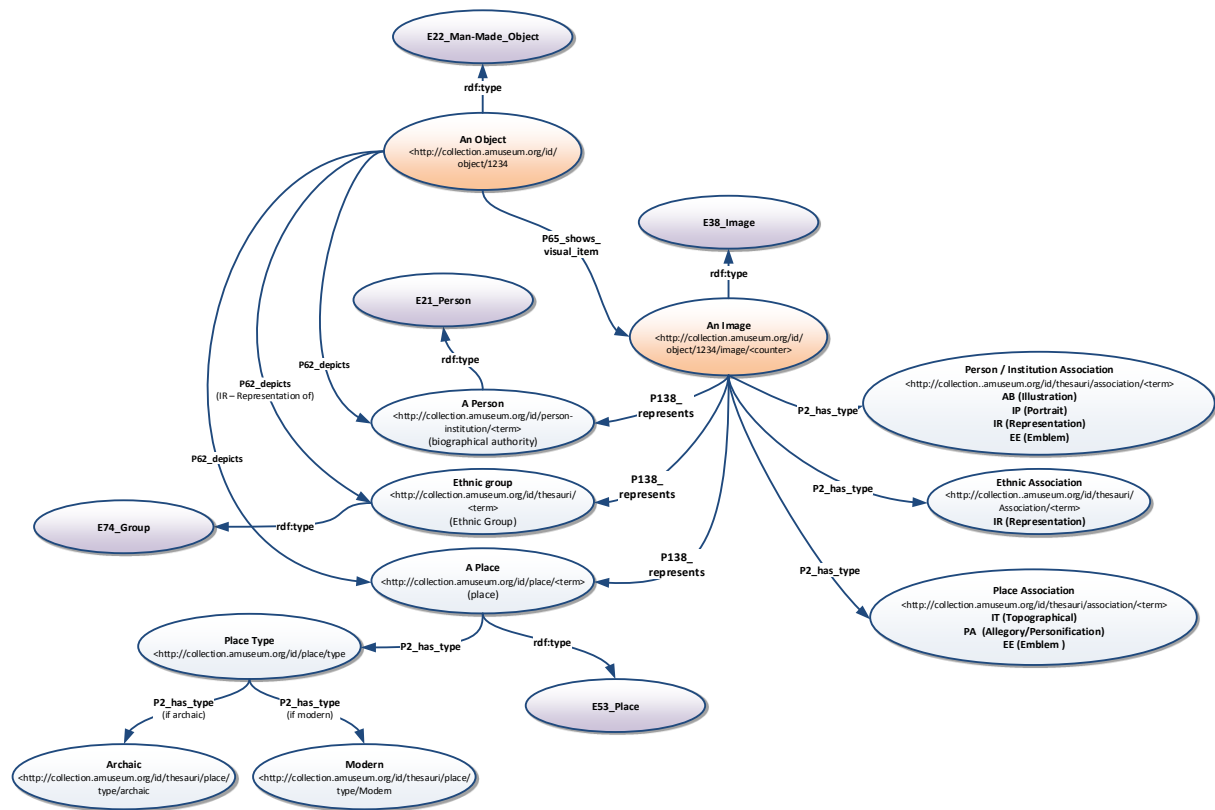
Bibliographic Item (core)

An object may be cited in a publication, and the events of 'publication' and 'authoring' will have their own patterns, associations and time-spans. This pattern is now superseded by FRBRoo but still provides a useful indication of how CRM is applied.