CIDOC CRM Art and Architectural Argumentation Conceptual Model Description

Document Type: Report
Author: George Bruseker (Takin.solutions)
Version: 1.0
Filename: CRM_AAA_Report_v.0.9.pdf
Report Date:1/2/21
Project Reference: Sustainable Art and Architectural Data Reuse in Teaching and Research

Table of Contents

Introduction

The CIDOC CRM is a semantic data standard for cultural heritage which, while originally developed within the framework of museum data integration, has grown in scope and audience and is now being widely and successfully applied in the humanities and social sciences in order to provide a logical framework for the integration, query and presentation of historical data. The event centric modelling technique of the CIDOC CRM makes the standard highly interesting and appropriate for adoption to researchers aiming to create sustainable and highly connected data structures for areas of analytic research that aim to understand the historical evolution and interaction of things, ideas and people. As a core data standard, the CIDOC CRM has a series of formally defined classes and relations which enable the re-representation of analytic data from existing structured data sources (databases, spreadsheets, structured documents) into a common information graph with a formalized vocabulary enabling the common representation of the same kinds of things under classes and the common representation of the same kinds of real world relations under properties. As such the CIDOC CRM forms a lingua franca for a project adopting the semantic technique allowing a common way of translating data into one format as well as a common way of querying across that data once the modelling/mapping of extant data has taken place. Given the increasing take up and enthusiasm for semantics and CIDOC CRM in general, however, and in particular in historical research contexts, the standard is necessarily going to be challenged with issues of accuracy and completeness as more demands are placed on its expressive system and new research questions and paradigms create new requirements for representing different kinds of historical circumstance and information, of interest to the researcher to track in an robust and well defined analytic manner.

The adoption of CIDOC CRM in the research programme of the Chair of the History and Theory of Architecture of Prof. Delbeke on art and architectural history brings a particular focus to the question of the efficacy of the standard in its present form to represent the kinds of data and kinds of argumentation over this data that researchers want generate and query. One of the main offers of semantic data lies in its ability to accurately represent researchers' data in a common format, creating an even balance between expressivity and usability. Finding this balance is crucial in the general push to manage the digital transformation in research by generating forms of digital representation of knowledge which are adequate to the task of representing complex information in a manner which is interpretable and interrogable by the researcher while processable by automated algorithms and general computer processes. The building out of semantic data models and patterns for art and architectural history is part in parcel of an effort to create a new digital literacy as a component and evolution of traditional scholarship: humanities research, in its proper sense, by other means. If this effort is to come close to reaching such a high goal, the foundations of this work must be thoroughly tested to make sure that it will bear the load that we consider posing upon it.

When we address the question of the adequacy of CIDOC CRM for art and architectural research, we immediately run into a fundamental problem. The event centric modelling strategy adopted to represent the historical facts of cultural heritage curation both renders the standard a useful semantic modelling language to adopt while limiting its applicability to many topics of

concern to historians of art and architecture. This paradox arises from the nature of the original scoping of the CIDOC CRM standard and the argumentation behind the adoption of an event oriented modelling strategy. As described in the literature, the development of formal ontologies in the computer science sense is relative to a scope of practical use and a domain of knowledge / community of users. An ontology is meant to represent a formalized set of terms that encapsulate a significant enough range of the common objects and properties of discourse and investigation of a knowledge community in order to be able to serve as an interlingua at a digital level to re-transcribe data held in heterogenous formats into a common frame for information query and exchange across technical limits. The ontology is meant to represent a 'commonly shared conceptualization' towards some end. The CIDOC CRM was originally conceived to integrate museum data, replacing the strategy of creating large, proscriptive database schemata in favour of deriving common a inter-lingua which would allow their mutual integration. In practice this means that the scope of the original efforts were given by museum datasets and museum priorities. The positive development out of this focus was to come to a generalization that placed events at the centre of the modeling strategy. Specifically the ontological analysis of discourse indicated that while museums document objects as the support and trace of a historical understanding and identity, as well as an object of care in its own right, the discourse around the object and the effort for understanding came through an understanding of its participation in history. The museum object is a witness to history and the museum record is a partial document of this function of witness. Therefore in contradistinction to the view presented by a standard recording mechanism for a museum object, the central information of the object lies not in its being qua object by its being in relation to other things through time. In short it is the event as the separator of states of reality that the object speaks to and allows us to query by tracing its relation to such events and their participants. This epistemic and ontologic analysis of the field of study and discourse drives the decision to create event oriented modelling, where the centre of information is not the object but the events in which it has participated (in being create, destroyed, modified, or having participated in any of these actions with regards to other objects).The resulting event oriented ontology's ontology quickly ranges beyond the its application in museum information integration but the ability to model huge swathes of data that is documented and described at the human (meso scopic) level of tables, chairs, cats and dogs and their histories is made possible.

But the articulation of an ontology means that the properties that it allows you to represent are functionally restricted by the space of questioning opened by the knowledge community for which the ontology is derived. In the case of museum documentation information, the knowledge community consists of the maintainers and operators of databases dedicated to the description and management of the cultural heritage objects within the care of that institution. The aim of the integration of knowledge is to find more connections with other objects over time. Museum databases are essentially inventory systems for unique items. The aim of the museum is to preserve the object and communicate its history. The museum must also represent the history and identity of the object according to the present best state of knowledge. Moreover, ontologies are meant to describe objectively retrieval states of affairs. That is to say, that ontologies meant to support scientific or practical applications are meant to derive terms and relations for representation that relate to classes of real world phenomena and their interactions.

The articulation of an ontology originally for museum information therefore both favours an aim which supports the representation of the present state of knowledge while equally being weighted towards the representation of what are perceived as ‘objective’ facts. This latter perspective, I will call presentism, while the former I will call objectivism.

The orientation of the conceptual modelling strategy of the CIDOC CRM development team to support a presentist knowledge paradigm combined with the pursuit of an objectivist orientation from the computer science perspective leads to limitations in what the core CIDOC CRM standard can represent in terms of historical information of interest to the art and architectural historian. The core CIDOC CRM purposefully limits itself to representing ‘just the facts’ with regards to the objects described within its scope. These facts described the historicity of things, but always already from the perspective of now. The description of the history of an art and architectural objects historical trajectory considered as a function of its objective movement through space and time and its contact with objects, person and ideas as a derivative knowledge from documents is more or less presentable.

But the goals and objects of historical research are not those of a theoretical physics and the interest in the objects which bear the traces of history is not usually in the reconstruction of the movement of a collection of atoms through time. Rather art and architectural historical research, while not uninterested in these topics, is as interested in the evolution of the knowledge and perspective of those who have interacted with objects through time. The historian works from archives and aims first not to represent the present state of knowledge, but past states of knowledge, regardless the present analysis relation to this knowledge. The interest is not only the event of the object but the horizon of understanding in which this object was captured. In the context of the integration of presentist, objectivist databases such horizons of understanding are eliminated purposefully and with reason from scope. The goal of such databases is to enable a knowledge of the present state of affairs and a consolidated up to date, consensus knowledge.

The historian working from archives aims to arrive at such conclusions, but must first of all be able to represent past representations of states of affairs, on their own terms. It was this vague ‘horizon of understanding’ that was purposefully expigated from the original CIDOC CRM modelling efforts and subsequently kept out which must be rendered clearer to analysis and therefore representable within a semantic model. What is shared as positive knowledge amongst historians must be representable by a proposed ontology if it is to be fit to purpose for supporting the representation of their knowledge. This calls in principle either for a new ontology or for an extension of the existing one to manage the expanded scope from the original functional limits of the project.

The CIDOC CRM aims to push aside the problem of subjectivity by being a representation only of what is objectively available for research. In its projection of what is objectively available it attempts to strictly limit what can count as a fact. What is allowed as strictly factual are things that are or are potentially observable. If we are to have a knowledge graph which represents potential points of disagreement, we must first of all make sure that the referents of our agreement are the same. By limiting the objective to the potentially empirically observable, the

CRM attempts to steer the ontology free of subjectively dependent facts. Whether we classify an individual as a hero or villain, we will all classify them as a person according to a very common definition. In the history of art and architecture the historian is interested not simply in raw physical facts. While it is the case that objective, observable events must found historical research, this only accounts for a fraction of the information that the historian is interested in collecting, preserving and interpreting.

In fact the historian is equally if not more interested in facts which, though grounded in a physical reality, have only a social ontological character.

Social Facts vs. Physical Facts

There is a large set of facts of interest to historians that are not empirically available in the sense of a physical object or event but which are, nevertheless, facts. We speak here of facts which are the result of social agreement. Such facts include things like something having a name, being considered to belong to a category, something being the property of someone. Such facts have no physical ontological character. The fact of something having a name changes nothing of the material composition of the universe. Judging from the physical evidence of a thing one cannot directly adduce either what it is/was called, what it is/was used for. Yet the fact of a name may have an enormous historical importance; the fact of a characterization can change history. In addition to the physical evidentiary chain of objects and individuals acting in events that the CIDOC CRM allows to represented semantically, the historian is interested in understanding and reconstructing the lived understanding and meaning of texts and acts for the actors that carried them out, to analyze, describe and understand how their understanding of this ‘same’ physical world may have had entirely different valences than our own.

Facts having a social character are those which entail the ascription of meanings and values to objects and events. Such meanings and values can be argued to be of a subjective character. By extension the argument can be made that such facts are not the appropriate target of ontological modelling for scholarship and research since they are fundamentally inaccessible to research. Such would be a false argument, however. While meanings and values ascribed to objects and events in a social setting are subjective, they are so in the sense hat they require and are grounded in a subjectivity in order for them to exist. Their ontologically subjective status, however, does not imply their epistemic subjectivity. There are interesting sets of ontologically subjective states which can be epistemically verifiable, that is epistemically objective. While subjective states belonging to individuals such as pain, emotion, my views on the quality of the work of Renoir are only directly accessible to the individual agent experiencing them, and their interest for empirical research, at least at a historical level of likely little value, other ontologically subjective states belonging to collectivities and which have empirical verifiability are of great historical interest: X person is president of the United States for such a time, the 5 dollar bill is worth so many RMB in this time frame, the Mona Lisa is known as La

Jaconde in France from this time. Each of the previous facts have no physical ontological ground which would allow us to verify their truth, yet each is a fact and the kind of fact whose truth or falsity and whose historical and social range are precisely of the kind that are of particular interest to the historian.

An essential element of the historian's work is to establish who believed what regarding which things for what time and under what circumstances. To follow such a research agenda is to follow a different functional practice regarding data to the paradigm of an inventory directed, presentist database. The question is not simply to have the latest state of knowledge and opinion of what ‘is the case’ but to capture the different states of understanding of different actors and understand how these actors and their understandings interacted and evolved. In other words, the historian needs to represent the past states of knowledge/belief of what ‘was the case’ before they can assay an argument on what is the case.

The attractiveness of the CIDOC CRM for historical research is its event oriented model alongside its wide acceptance as a standard for cultural heritage data meaning a large potential pool of compatible primary evidence from CH institutions. The meso-level picture of people and things meeting in time is a structure which allows the conveyance of the physical history of objects. The CIDOC CRM performing its function as an inter-lingua provides a simplification of human level conceptualizations in order to facilitate information exchange. What we will argue here is that the simplification performed hides a more complex process which, when the standard is adopted by historians, needs to be surfaced, properly understood and represented in order to enable the historian to trace the evolution of the social meaning, value and understanding through time. If we wish to extend the range of CIDOC CRM to support historical research and reasoning on the evolution of these phenomena over time, then we must examine again the basic simplification and determine at what joint to establish a more explicit modelling of the social aspect of facts.

We can find this juncture, I argue, in looking over the standard again and distinguishing between truly ontologically objective facts, facts true regardless of observers, from ontological subjective but epistemically objective (verifiable) facts that are represented in the ontology. CRM’s presentist simplification elides these two classes of facts, making all the statements of fact true, to the best of their knowledge, relative to the observer, now. So the CRM enables the encoding of the statement that such and such an object X has a name Y at the same level as such and such an object X is at place Q; it enables the encoding of the statement such and such an object X has function P and such and such an object X forms part of this object S. Primae faciae these are all simple statement types providing primary evidence, facts, regarding some object. Such mundane facts strung together by a historian or other researcher enable an empirically based argumentation regarding some matter. But while seeming to be equally simple statements, there are radical differences between these pairs of statements. While the latter statement in each of these pairs of statements truly matches the objectivity criterion set out in the design principles of the CIDOC CRM, such that it is possible to check whether object X is at place Y or is truly part of object S by a direct empirical observation of the state of affairs, the former statements do not have the same character. In the normal case, there is nothing on the

object that declares that it's name is Y, there is nothing necessarily inherent to the object to tell us that it has function P. There is something more going on these facts, something of inherent interest to the historian as a point of knowledge and argumentation.

I would point here then to a distinction between physical and social facts represented indifferently as the same kinds of facts in the CRM standard but, in fact, forming two distinct kinds of facts. The former are both ontologically and epistemically objective, observer independent components of reality that are subject to verification by an empirical circumspection using appropriate techniques. The latter are ontologically subjective and epistemically objective. That is to say that they are features not found in the world as physical facts, but are super-imposed on the world by social collectivities in a manner which is nonetheless objectively recoverable, provided an appropriate knowledge and orientation of the observer in the social system and language in which they are conveyed. It is this latter class of facts that is of interest to develop further with regards to its representation in the ontology in order to support historical knowledge representation and reasoning.

The aim of creating such a distinction is not to challenge the basic principles or modelling of CIDOC CRM, but rather to recognize a distinction point from which an extension might be built which would provide a systematic way of representing the kinds of social facts that are of inherent interest to the historian. By identifying this fundamental distinction we have a means to carry out a meta-analysis on the existing standard which allows for the identification of properties specified in the standard to represent facts which in a fuller representation extending the CIDOC CRM, could capture the social nature of these facts as not raw, ontologically objective truths, but in their actual context as ontologically subjective but epistemically objectively available social facts.

Collective Intentionality Oriented Semantic Modelling

To aid in this analysis and provide an initial orientation, I will rely largely on the work of John Searle(John R Searle; John R. Searle, *The Construction of Social Reality*; John R. Searle, *Intentionality*) to aid in the analysis and representation of such facts. The toolkit of Searle's social ontology is useful in this analysis since it shares to a large degree the realism of the CIDOC CRM approach. It takes as an ontological ground one, objectively available world on which all other properties are built. In his work on social ontology, Searle's effort is precisely to elaborate means by which a world of meaning for agents can be superimposed onto a world which objectively is and remains a physical system in the sense represented by the sciences.

Searle's philosophical picture fits well with the project of extending and elaborating the CIDOC CRM in order to account for the nature of social facts of interest to historians precisely because he investigates and provides a systematic account for the coming to be and passing out of existence of such facts. In particular, Searle is interested in how institutional facts are built up over the brute facts of a physical reality. The synopsis of Searle's work is a larger project than

can be properly done here, but we can rally certain basic elements from it in order to support an initial approach to a consistent modelling and extension of CIDOC CRM to account for social facts.

In order to describe and explain social facts, Searle puts forth the key concepts of the assignment of function and of collective intentionality. Combined these two tools are used to explicate his principle that social facts are built iteratively through the application of a logic 'x counts as y in c'. The notion of assignment of function is that as subjective, social beings, reality is always laden with meaning for the individual. An object is never given as-is in the brute scientifically objective sense of an object to human beings. Rather it is always already taken up in a system of symbolic interpretation which takes this thing as something else. The collection of matter consisting of four elongated elements topped by a horizontal, square element with a vertical square attachment attached on one edge is taken as 'a chair'. The collection of matter with four limbs, a billy club and a badge is taken as 'the man' and so on. The assignment of functions to things is ubiquitous. The evolution of the assignment of functions is part of the proper domain of history. In addition to the assignment of functions, Searle identifies the notion of collective intentionality and defends it against some typical rejections of it. For Searle it is important to recognize the existence of collective intentionality in order to ground the possibility of an empirical investigation and understanding of social facts. The notion is that two or more actors can share the same intention, that is take some object/act/event in the world as something else in precisely the same way. Such a claim is controversial in that traditional arguments would object that intentionality is only had by individuals. Introducing collective intentionality, so the argument goes, opens the door to a super entity, a group or zeitgeist that is the subject of this willing. Searle argues against this interpretation while endeavouring to outline the real nature of collective intentionality. According to his argument, collective intentionality is willed always and only in the individual, no group minds, but the contents of the intentionality need not express only a first person willing but may indicate in fact a third person perspective; not I intend x as y but we intend x as y. Searle argues that such collective intentionality are in fact ubiquitous in everyday life and calls them social facts. These are facts like 'the 5 dollar bill can be exchanged for goods', 'the state is the owner of this land', 'Sally is John's mother' and so on. The kinds of things that are social facts then are ones that are collectively willed and make an assignment of function of one thing to another.

Having outlined the concept of social fact, Searle also provides a mechanism for describing the means by which such facts come to be and pass away. To do so he has recourse to the work of Austin (Austin) and the idea of speech acts. Austin's aim in How to Use Worlds to Accomplish Things, was to point out the large sphere or linguistic production that fell outside of then current interest of philosophers of language. Language was treated then in philosophy of language as it is largely still now in computer and information science as a truth functional affair. The function of creating sentences is to generate propositions that do or do not mirror states of affair in the world. Austin wanted to point out a class of propositions that while commonly used in utterances had nothing to do with representation or truth function but were, instead, ways of generating new situations or states of affairs. He called this class of utterances performatives and speech acts and went on to carry out an analysis of the basic elements of a speech act. Austin's speech

acts provide a key engine for the production and analytic for the understanding of the origin of Searlean institutional facts. The notion of speech acts as ritualistic events in which society has invested meaning for the generation of yet more meaning is the appropriate mechanism for explaining the origin of at least a subset of Searle's institutional facts. Austin's basic analytic provides a useful toolset to understand the formal elements of a speech act. A speech act uses propositions in a non truth functional way (they bind physical reality to a social formulation of its meaning), they follow a rule (allowing us to identify when such an act has been performed or not) and they bring about a new state of affairs (factually we recognize as a real world phenomenon for human beings the real effects of the performance of a speech act and the subsequent imposition of the consequences of its generated institutional fact).

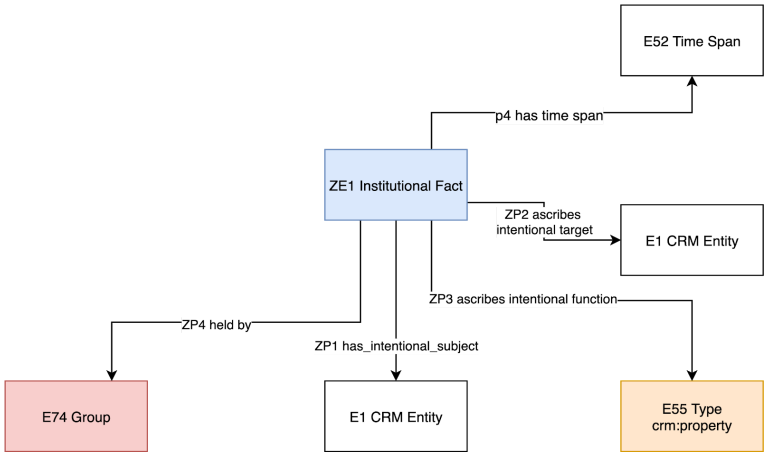
The reflections on the needs of historians for documenting positive information of past states of knowledge and belief, including contradictory ones, as well as the need for separately provenancing the present representation of this past state of affairs in information systems alongside the Searlean-Austinian analysis of the positive epistemologically objective nature of subjectively grounded social facts provides a rich conceptual toolset by which to provide a new approach to modelling historical information that is at once adapted to the CIDOC CRM as international standard for cultural heritage data but also to the data management needs of practicing historians wishing to accurately document their knowledge.

In the extension we propose then out of this research, a new basic modelling paradigm is adopted in adequation to the historical scenario. The CIDOC CRM is classed as an event oriented or event centric modelling design. It furthermore plants as its chief functionality criterion, the facilitation of data integration. For CRM AAA we propose underneath this larger umbrella to make the addition of an collective intentionality oriented modelling whose chief functionality criterion is the capture of the existence of institutional facts, their historical provenance qua historical fact and their present provenance qua represented information object. The aim of this extension is to provide an initial, philosophically grounded and consistent means of extending the CIDOC CRM model in order to allow historians to describe analytic information of importance in grounding their arguments in an adequate way.

The Institutional Fact Modelling Pattern

Under the above objectives, the first task in providing the extension was to articulate a basic collective intentionality oriented model for describing historical facts and to connect this to the basic event oriented system of CIDOC CRM.

The basic collective intentionality oriented modelling that we propose has the following structure:



This structure forms the nucleus of the modelling effort of CRM AAA. Ontologically it is meant to provide a hook to represent the basic elements of social facts documented with regards to their historical context. Documenting such a fact is to capture its historicity primarily. We wish to be able to support the documentation that such and such a collective intentionality existed for some group, in relation to which it object it was held and which social meanings it ascribed to that object. To do this we declare a high level class ‘institutional fact’. An institutional fact is modelled as a subclass of temporal entity the CIDOC CRM system. Ontologically the subjective intentionality of a collectivity is a passive temporal process. Once initiated it is a time bound phenomenon. As a mental phenomenon it has no geographic extent. It does not occur ‘anywhere’ but in the ideal world of the subject. Therefore it is modelled directly under temporal entity and bears ontological similarity to the notion of ‘state’. In future modelling, the The substance of institutional fact is the attribution by the collectivity for the time of validity of that social fact of a certain meaning to a certain object. This is represented through three essential properties. The first property, has intentional subject, indicates the subject of a collective intentionality, the object about which intentionality projects a certain meaning and a certain extra social property. At this highest level of abstraction any entity can form the thematic subject of an intentional relation, we therefore declare the range of this property to be E1 CRM Entity. The second property introduced, has intentional target, indicating the object of the intentional projection. In the Searlean formula, social facts are taking X as Y. In this formula ‘Y’ is the object under which the initial subject is considered. Here the work of intentionality is at play in generating social facts. There is nothing ‘in’ nature that connects X and Y expect the intentionality of the social collective. It invests the world with this meaning and in turn relates to the invested world ‘as if’ this projection were the case. Again, any entity may become the target of the investment of meaning in relation to another entity, therefore the range of the property has intentional target is also given as E1 CRM Entity. A third relation is required to adequately represent the basic case of collective intentionality. In the act of intention, in addition to the subject about which the intention is undertaken and the target which modified the subject according to a new aspect, there is also the question of the manner of this modification. We must be able to represent the way that the social collective projects this relation. For this function we introduce the ‘ascribes intentional relation property’. This property is given a range of E55 Type in the CRM system. This property allows the documentation of how X is connected

to Y by the social collective intention. Here we had to do with the most diverse of possible phenomena, the projection of meaning through the particular relation of physical separate entities through a projective imaginative act. From an information modelling point of view this poses a significant challenge as a too open data structure will result in analytically useless data results. For this reason we bind this property to the relationship properties of the CIDOC CRM itself in the first place with the notion capability of its expansion to further ontologies. The notion here is to borrow from the existing properties that have been of expressed to be of interest in the context of a presentist information system and modelling strategy, borrowing their consistent formulation and ontological modelling pattern, but placing them within a context where the assertion of these properties can be expressed with regards to its historical dimension. Two more properties are needed to support the documentation of the institutional fact in itself. Basic to the institutional fact is that it is held by some group. For this reason we introduce the 'held by' property with a range of E74 Group. We limit ourselves at this moment to the notion of group intentionalities and do not engage with the question of private mental beliefs. While this is of interest it is not in the scope of this project. The class E74 Group as defined in CIDOC CRM is restricted to groups with a common identity. In the historical context this stricture is probably too strict. Within the context of this project we have decided to more or less ignore this convention of the CIDOC CRM, but a more concentrated modelling effort in this direction might come up with a good representation for a super class fo E74 Group which better represents the kind of ad hoc, externally defined groups that can be the subject of historical analysis. Finally the new class of institutional fact being identified as a subclass of temporal entity allows it to express the temporal limits of the the collective intentionality through the use of the p4 has time span property and typical CIDOC CRM modelling.

Before looking at the other major structure of CRM AAA, the speech act, it is useful to dwell some further time on the institutional fact structure and its functionality. In particular it is important to note in what ways it resembles other modelling structures, how it differs and what justifies its introduction in further detail. Topologically, the institutional fact structure resembles the attribute assignment class of the CIDOC CRM as well as the general reification structures of RDF. It would be reasonable to ask if adopting one of these structures would not be sufficient. In fact the proposed representation differs in important semantic ways from both of these structures. The general reification mechanism in fact is insufficient semantically for what we aim to document here both because it doesn't fall within a formal semantics modelling structure and strategy like the CIDOC CRM. It would not offer the ecosystem of further relations necessary to document historical facts. Moreover, it is topologically as a data structure since it only connects two classes via a mediating class, a subject and object, but does not account for the relation stipulated between them. In this regard, the later versions of attribute assignment of the CIDOC CRM standard are much closer to purpose. Here, however, again we have an important semantic difference to take into account. The E13 Attribute Assignment class is a generic class for documenting the acts of attribution undertaken by scholars and scientists in a cultural heritage context. The E13 Attribute Assignment class is about documenting events of attribution, not about documenting their outcomes. In the historical context, however, the positive knowledge of the historian tends not to be about the event. The historian re-representing the historical record in an information graph, often may not know enough to

document events of attribution (the CIDOC CRM recommended way to document these things in order to ensure information integration) but instead only knows from a mention within a text that such and such a fact was the case: so and so was named X, such and such was used for Y. What the historian needs to be able to express is this knowledge of an institutional fact in its temporal and social context without having to express the event of its generation.

The last case founds the basic argumentation for introducing the institutional fact / collective intentionality centric modelling paradigm. While the CIDOC CRM aims to support museums in information integration of diverse datasets, CRM AAA as an extension aims to support historians in accurately recording the fragments of knowledge we have of past social states by providing robust information structures for capturing who projected of what object, what relation. Materializing institutional facts will lead to a proliferation of facts in relation to the objects of discourse in the CH knowledge graph. In the context of a museum presentation this may not feed the functional goal of having the up-to-date knowledge of the present scientific consensus. With regards to the historical method representing the cacophonic and contradictory projections of different groups over the same objects through time is a *condition-sine-qua-non* for performing historical analysis. Being true to the evidence available also requires being able to represent the partial knowledge made possible through the historical record, rather than passing through the event centric strategy recommended by the CIDOC CRM SIG for information integration which, at a historical level, is insufficient for the granularity of knowledge and commits historians to creating entities in the knowledge graph for which they can provide no historical ground.That it to say, for maintaining an objective representation of the facts of interest to historians, another overriding functional principle to integration must be projected, and that is the ‘fit to evidence’ principle and the ‘contextualization of social fact’ principle.

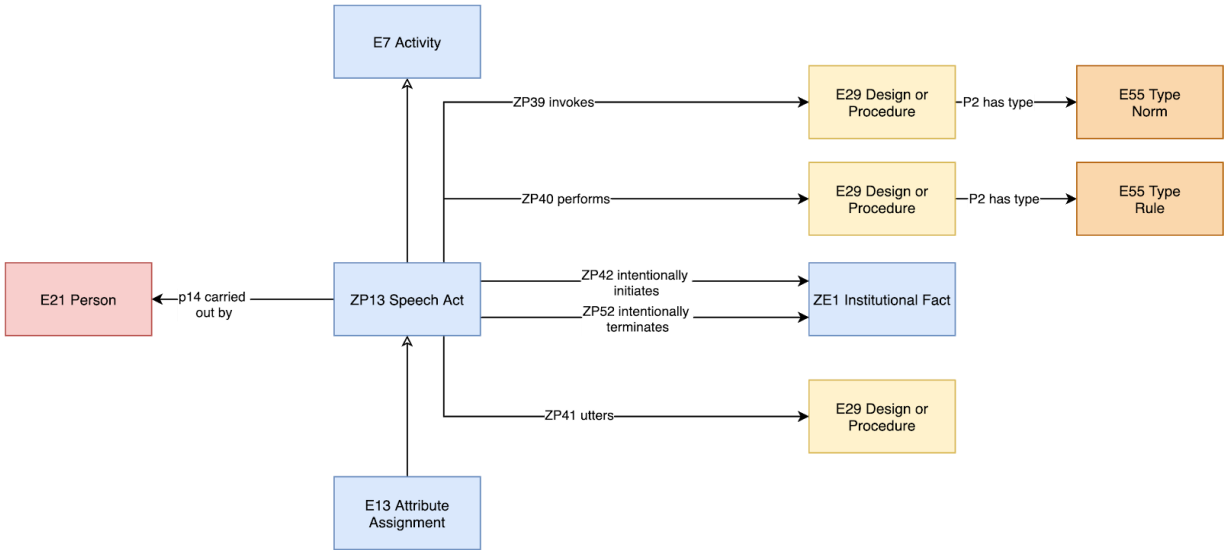
The institutional fact modelling provides a means to represent ‘what is the case’ of historical collective intentionality which is, often, the object of representation of the historical endeavour. In order to provide, however, a complete ontological picture of social facts, we must also provide a high level means to document the events which lead to the coming into existence or passing out of existence of these collective intentions. For this we produce the following model aligned to the CIDOC CRM of the Speech Act:

The Speech Act Modelling Pattern

Translating the notion of speech act into CRM AAA and aligning it to the CIDOC CRM event centric model creates a crucial join between the basic data that can be produced using the CRM standard and the specialized data that may be required to be generated in historical research regarding past social states. The Speech Act is modelled as subclass of E7 Activity. Here we signal it as being a characteristically human action. Speech Acts are in fact an essentially human, collective activity, in which a ritual is performed that empowers the realization of a token meaning onto a thing, an object or a situation. Adding the speech act to the model adds a generalization to all kinds of official acts which are the cause of the existence of institutional facts. This means that while the historical evidence may only provide partial reports of

institutional facts the model is still capable of documenting and accounting for the generative events which caused them in a consistent manner. In fact, the CIDOC CRM already has within it a hidden notion of speech act in the function of the E13 Attribute Assignment. Attribute Assignment and its subclasses describe a special subset of speech acts particularly relevant to museum and cultural heritage management. It has to do with the events that lead to the attribution of things like identifiers, classifications and evaluations of condition state. Whereas the CIDOC CRM, however, is interested in supporting latest state of information data integration for practical inventorying systems, the historical data imperative is to capture not only the event of attribution but the social state or institutional fact that such events create. The ascription of a name or a classification is far from a neutral practice and such actions have consequences in a society far beyond their seeming insignifiance on an individual level.

The conceptual modelling structure proposed in CRM AAA looks as follows:

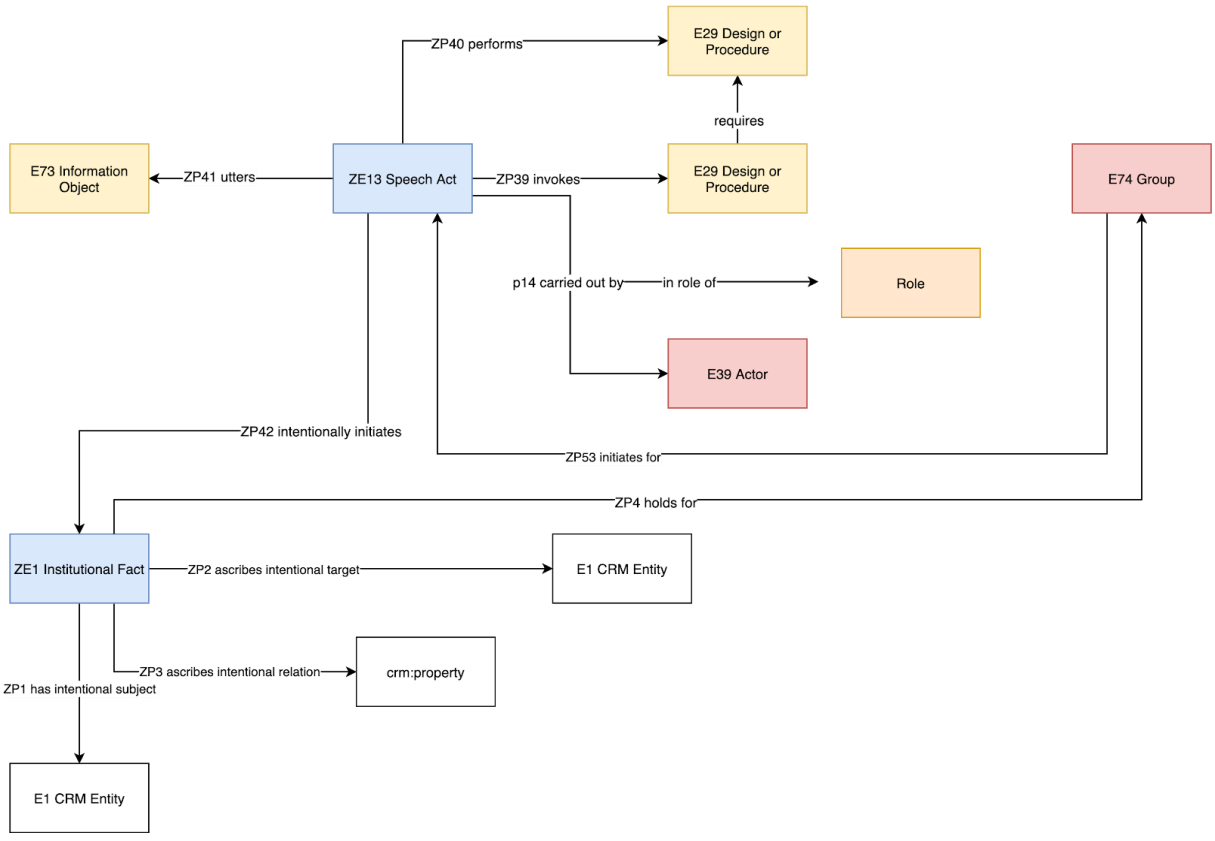


Speech Acts are therefore introduced as a super class of attribute assignment and provide a series of properties essential to capturing a speech act. As described in Austin, the Speech Act is a ritual with certain essential elements that must be carried out by its participants following a social code in order for it to realize its aim: a change in social state. Performed correctly, it creates a new social state, the institutional fact. We therefore provide for two properties related to the expression of the use by the speech act of social norms. The property ‘invokes’ links the speech act to a general law or social norm that the speech act is meant to appeal to for the general principles and forms of the institutional fact it aims to bring about. That is to say, it points to ideas or plans formulated by the social group of its own operation and continuation. In the typical case instances of the range of this property would be laws or mores. Here we have designated the range of the property as E29 Design or Procedure which is somewhat too narrow in its definition of a set of rules but which captures the idea in the interim. Continued work on CRM AAA would suggest a need here, as with Group, to provide a better generalization of this class to cover the historical case more accurately. The second property declared for this class is the ‘performed’ property. This property again links the speech act to a plan. Here the

intended range instance of this property would be the very specific plan specified for the kind of speech act that is being performed, for example the essential steps for a civil wedding. In deference to the speech act theory itself, the recording of the actual wording used in the speech act is made possible through the 'utters' property. Finally and most importantly, we provide the property 'intentionally initiates'. This property links the speech act to an institutional fact. This property marks a clear departure from the modelling principles of CIDOC CRM base which remains highly agnostic of causal relations, given the possibility for different explanations of the same underdetermined events and for multi-factorial causation etc. Speech acts, however, represent a distinct class of declarative, agent based actions which are direct real causes of their ontological subjective but epistemologically objective outcomes: institutional facts.

The Hidden Social Facts of CIDOC CRM Surfaced

Together the speech act and the institutional fact present a consistent modeling pattern that brings together the event orientation of the CRM base with the proposed collective intentionality orientation of CRM AAA. Seen together the two patterns connect as follows:



With these two major modeling patterns available, we are able in the first instance to do a reassessment of the CIDOC CRM with regards to thinking how its expressivity can be enriched consistently for the historical domain. In order to do so, we have to begin with consistently

dividing off purely social facts as documented in the CIDOC CRM from those which are physical in nature. The point of this exercise is to render the existing model with the new approach consistently in the extension and develop a pattern which is extensible for further historical study and questions using a consistent methodology. Therefore we set about to perform an analysis of CIDOC CRM separating social from physical properties in order to understand what should already receive additional expression in the CRM AAA extension. The method for finding these properties amongst the existing properties was to ask whether the property type asserted could or could not in principle be understood by a non or low level symbol processing agent. In a more simplistic way, one could ask, would by dog recognize this property. Physical facts in principle are accessible to any perceiving agent and verifiable from an external inspection. Social facts to be recognized and understood require the apperception of a symbol processing mind, a human agent belonging to a symbolic and linguistic system capable of understanding and intepreting the abstract connections projected on the physical world through social processes such as speech acts.

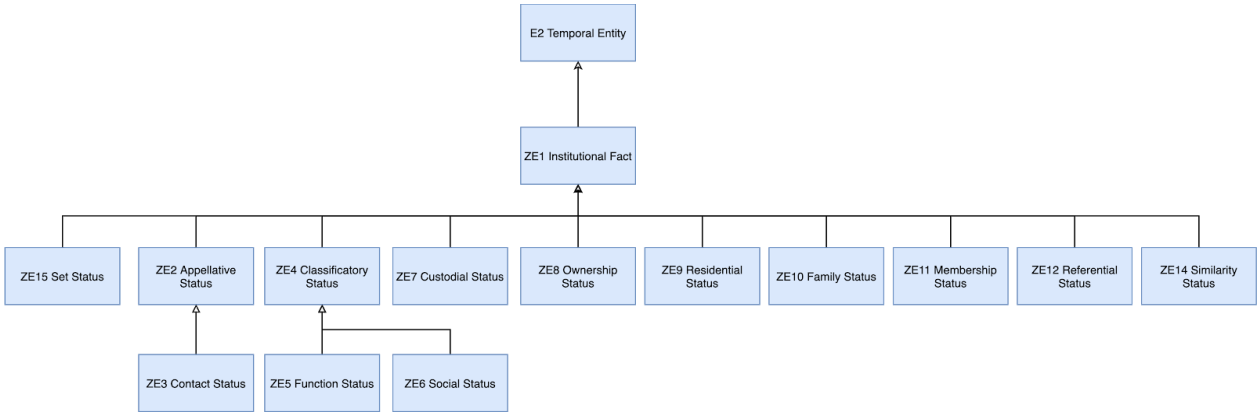
In doing so we created the following list outlining the properties extant in the CIDOC CRM bearing a specific social character and identifying particular classes of institutional fact to which they belong.

Property	**Property Type**	**Institutional Fact Type**	**CRM AAA v0.9**
P20 had specific purpose	Social	Aetiological Status	No
P21 had general purpose	Social	Aetiological Status	No
P1 is identified by	Social	Appellative Status	Yes
P48 has preferred identifier	Social	Appellative Status	Yes
P2 has type	Social	Classificatory Status	Yes
P137 exemplifies	Social	Classificatory Status	Yes
P76 has contact point	Social	Contact Point Status	Yes
P109 has current or former curator	Social	Curatorial Status	Yes
P49 has former or current keeper	Social	Custodial Status	No
P50 has current keeper	Social	Custodial Status	No
P152 has parent	Social	Family Status	Yes
P101 had as general use	Social	Function Status	Yes
P54 has current permanent location	Social	Locative Status	No
P107 has current or former member	Social	Membership Status	Yes
P51 has former or current owner	Social	Ownership Status	Yes
P52 has current owner	Social	Ownership Status	Yes
P67 refers to	Social	Referential Status	Yes

P70 documents	Social	Referential Status	Yes
P129 is about	Social	Referential Status	Yes
P138 represents	Social	Referential Status	Yes
P74 has current or former residence	Social	Residential Status	Yes
P75 possesses	Social	Right Status	Yes
P104 is subject to	Social	Right Status	Yes
P105 right held by	Social	Right Status	Yes
P130 shows features of	Social	Similarity Status	Yes
P150 defines typical parts of	Social		No

Each of these properties is characterized as belonging specifically to the social world and having a validity that is inherently tied to a social grouping and perspective. Here the materialization of the relevant institutional fact as a kind of documentable social phenomenon would be of value in properly historically contextualizing the fact and making it an object of provenanced understanding. In this process we did not generate new institutional fact classes for all identified properties but for those of particular interest in a first found for representation. The identified properties become the initial set of properties available for assertion within their relative institutional fact class. Such lists of properties could be deeply enriched for historical research, enabling a more fine grained approach to such statements, while maintaining a general consistency with the overall CIDOC CRM.

We can see the resulting hierarchy tree for this initial analysis of hidden institutional facts in CIDOC CRM here:

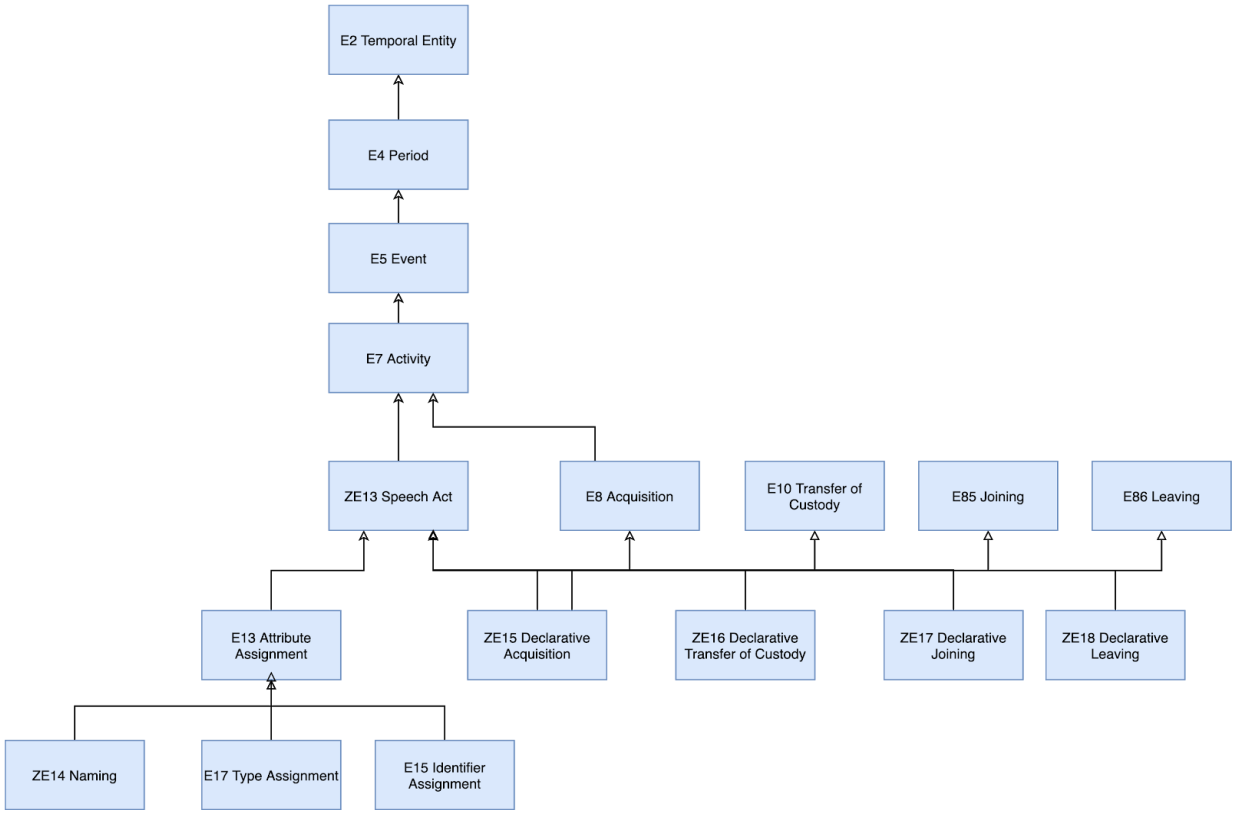


The new class structures of CRM AAA for representing institutional fact categories begin to paint a picture of how a more social historically oriented knowledge representation suitable for historical data management can be created as an extension to the CIDOC CRM.

CRM AAA enables the further expression of the social facts already hidden in CIDOC CRM in a robust manner which supports the documentation of the temporal limits of the social fact in question, the ‘for whom’ of the social fact as well as making the social fact itself a point in the knowledge network, referenable in its own right in order to indicate both how it came to be (a relevant speech act) or to make it an object of provenanced documentation (who has historically referred to or has presently inferred it).

The accompanying conjunct of each class of institutional fact is the existence of a class of speech act that either brings into existence or terminates an instance of institutional fact. Here the chief new class to be made was that Speech Act. A new class had to be made under identifier assignment to create a generic class for naming which is not covered under the general CIDOC CRM standard. After that the creation of appropriate classes was guided by specializing existing CRM classes for the relevant institutional facts. For example, the extant CIDOC CRM class E8 Acquisition provides a general class that describes the change of social state in which ownership of an object passes from one individual to another. As a general ontology CIDOC CRM allows also the registration of an instance of theft or loss as an instance of change of ownership. Such physical acts however do not lead necessarily to a change of social state. In principle a stolen or lost item may still belong to its owner depending on the social rules of the society in question. In CRM AAA we declare a subclass of this class and of Speech Act in order to isolate the specific social event which officially declares a transfer of ownership and thus causes a new social state which is completely and totally determined by the act and initiates a genuine change in social fact. This way mutatis mutandi we specialized the general CIDOC CRM classes to handle the special cases of social act which make this possible.

The proposed new class structure of CRM AAA is as follows:



The characterization of the ontologically subjective but epistemologically objectively available social facts related to historical objects is one of the core interests of art and architectural historical research and is of increasing general public interest as movements for social justice and a reevaluation of colonial pasts aims to surface suppressed perspectives and knowledge. Towards these tasks the CRMbase is not well suited as such but, supplemented by CRM AAA or an analogous modelling effort, there are important possibilities for extending the expressivity of CRM to capture the nuance of social fact and the contented social projections of objects, ideas and situations over time.

We present here the basic scenarios arising from initial modelling efforts within this project.

Appellative Status and Naming

The act of naming plays a fundamental role in the relation of human being to the thing named. Names are not neutral but articulated in a language, according to a point of view, extending over a more or less certain named object. The contentiousness of names in history has far too many examples. The introduction of semantic machinery to represent this fact is of utmost priority for research into changing and contested names. Examples can be sought for in the contested naming of territories in a historical sphere or the changing titles ascribed to museum objects in a museological context.

Classificatory Status and Type Assignment

The act of classifying is of fundamental social and scientific importance. The identification of an object, concept, situation or even person under a certain classification can have long standing ramifications on knowledge and ethics. Scientific examples such as the misclassification of the so-called brontosaurus is an interesting case in the history of science and ideas while more sinister examples may come from such domains as the history of slavery with the classification of certain human beings as chattel.

Contact Point Status and Identifier Assignment

Mundane but not without practical consequence, contact points are the means of finding individuals through communication services. The means by which an individual can be contacted and under which context is, on the one hand an administrative affair, but is also a question of following the provenance of knowledge directly to a human source. Representing such information positively enables the study of the availability of an individual through time.

Custodial Status and Declarative Transfer of Custody

The notion of custody is a deeply social matter related to the responsibility for care of an object. The proper and official change of custody of an object, under what jurisdiction and for what time is of high historical and legal interest.

Family Status

The CIDOC CRM purposely avoids the modelling of familial relations as a distinctly social phenomenon. The materialization of a family status class allows for the tracing of this social phenomenon in a consistent fashion. Recognizing the constructed nature of the familial relation and its culturally embedded articulation, the family status class allows for the introduction of family relation properties from developed ontologies for such purposes (e.g. Agrelon) in a systematic way consistent to the overall CIDOC CRM but capturing the specificities of different social formations and enabling the documentation of the acts leading to the acquisition or loss of certain of these relations.

Function Status and Type Assignment

The case of function status is a core example of the utility of modelling social facts. Function statuses are attributions of use types to objects. Searle mentions these as key examples of the attribution of objective social meaning to a physical carrier object for a group. He gives the example of a screwdriver that has the functions of screwing and unscrewing things and the related operations projected over its pure materiality. While the deliberate production of the tool to fit such a function makes it adapted to the task, it must also be assigned such a function by society, leading to relevant relative comportments to it qua thing with a purpose.

Membership Status and Declarative Joining / Leaving

The CIDOC CRM presently allows the expression of facts of membership through its event model or through the presentist mode of a binary/non-temporal property. CRM AAA extends the expressivity of the base model to handle the ‘fact of membership’. To belong to a group is a temporally bound phenomenon that relates back to the assignment into a group, which relates the individual with the membership back to the collectivity attributing membership in certain distinct social relations. The fact of being member at such and such a time is a positive fact that has consequences in social relations, allowing or disallowing different acts and speech. This is a first class topic of research and such requires positive representation in the ontology. The membership status class allows the documentation of these cases.

Ownership Status and Declarative Acquisition

Ownership in its fully legal sense is a social phenomenon. While theft and loss can lead to de facto illicit ownerships, such an ownership would not be recognized de jure. Thus ownership is core case for documenting the social fact of an ownership status. Ownership is conferred under certain norms and certain laws and hold for the groups that support and accede to them. Legal systems, however, have cultural, temporal and geographic boundaries. There can therefore be contested and multiple social facts of ownership claim upon any given object at any time. These are positive realities, objects of historical documentation and study. The explicitation of an ownership status class and its declarative acquisition conjunct enables the tracing and understanding of such claims.

Referential Status

CIDOC CRM base allows the attribution of a number of properties starting from ‘p67 refers to’ which allow the documentation of the referential status of an intellectual object. Here again the presentist paradigm is not interested in the evolution of meaning, but aims to support an

objective understanding of a collection of objects according to a present paradigm of knowledge. But within a historical study we can be interested in the changing meaning of a given concept, information object or visual content. What something refers to changes dramatically in context, for groups and through time. Introducing a referential status node enables the tracing of such changes. Examples of such changes may be systemic and institutional like the change in position relative to a symbol. Here we may have, by way of illustration, statues of national figures of historical import who once stood for ‘freedom’ or ‘good government’. Re-evaluations of such figures may lead to the investment of such objects with new meanings. Such meanings could indeed be reversals with the same object now be characterized as referring to ‘racism’ and ‘slavery. These are social phenomenon of objective status that are the object of art and architecture historical investigation. The referential status class enables the construction of documentation and argumentation around such questions. At this point the project we have not arrived at a specific sub-class to describe the nature of investing or divesting an object of a particular meaning. For this reason, the initiating act for such an institutional fact at this moment can be modelled with the ZE13 Speech Act class. Further research will attempt to define a usable ‘declarative meaning assignment’ class for capturing just such acts.

Residential Status

While the CIDOC CRM aims to track objectively retrievable facts from the empirical world, location as a social concept is not amenable to such an approach. Having residence in a de jure sense need in no way necessarily directly associate one physically with a place. While, at the end of the analysis, the fact of a residence status, that such and such a person or group had a residence here or there, may imply that some actor at some point passed by that physical location, the import of such information does not lie in tracking the physical movements of agents. Rather, residential status has to do with a social phenomenon of where a person or group is related to as having a place of operations and thus having commensurate duties and responsibilities. Residential Status has to do with where the person or group ‘ought’ to be and suggest potentials for them ‘being able’ to act in certain ways in a community or place, regardless of their physical whereabouts (in the case of groups such whereabouts are typically only conceptual. I.e.: a company is not physically anywhere, it is only an idea. The class of residential status allows the tracing of places that are the declarative locations of individuals or groups over time. This is particularly useful for tracing the ‘movements’ of such entities as companies which may setup offices in multiple localities etc. Here again for lack of primary evidence we did not arrive at a specific class for causing residential statuses to come to be. When documenting such events the overall speech act class may be used. Future research may uncover a proper specialization.

Right Status

Right status is a particular complex issue in CIDOC CRM. The E30 Rights class that exists in CIDOC CRM base shows the lack of a means to model entirely social facts. An instance of E30

Right in CRM base is ultimately an instance of E28 Conceptual Object. That is to say, rights are modelled as concepts. But rights have a much fuller ontological nature than this. They have temporal, social and deontic aspects that are not covered by this very cursory class. The notion of right is further conflated by the high level class E72 Legal Object which introduces into the CIDOC CRM a strongly cultural notion of what does and does not count as a legal object. While, again, this perfectly matches to the declared limits and functional aims of the CRM base it leads to problems in application for historians who are committed to work from a representation not of what we think is necessarily the case but a presentation of what historically was seen to be the case. The complete correction of the modelling around the Rights class in CIDOC CRM requires further work and analysis and was not the focus of this project. Nevertheless, this project offers some headway in providing a consistent framework for treating right according to its own nature, not as a fact of nature, but as a fact of human consensus. By materializing Right Status as an institutional fact, we can talk about the temporal and social aspects of a right. When did a right come to be, when did it end, what object did it hold over and for whom did it hold? Here it seems clear that a new higher level class above ‘acquisition’ which deals with the trade in physical objects, must be created to talk about the general transfer or rights of which acquisition is a subset of cases. This would be work for a project with a more rights focussed aspect.

Similarity Status

This class arose out of the modelling of the Bernini Tagging Project use case. Similarity can be seen as a concept on the border of physical and social facts. There must be some sort of objective similarity for a similarity to be asserted, but similarity - seeming - is always a seeming to someone and under some context. In the context of historical argumentation the question of similarity has to do not with the general assertion ‘x is like y’, but a series of argumentation steps around what grounds and supports a statement of the type ‘x is like y’. Typically in historical contexts a similarity status may be proposed between two objects and the nature of this similarity can be the subject of an extended dialogue and conversation. Materialization similarity status as an institutional fact makes such an assertion a first class entity whose provenance can be followed and about which argumentation can be put forward for and against. We did not derive at this point a conjunct declarative speech act; this will form an object of further research.

Set Status

The CIDOC CRM explicitly does not have a set class on the now familiar argument that it doesn’t serve to support a state of present knowledge and is not suitable for integration. Sets in the sense of arbitrary collections of things, a list of entities regardless of their ontological status, are often cursory and ephemeral objects. Therefore, they often fall outside of official documentation. But the understanding of the development of ideas and the organization of events weighs often heavily on an understanding of just such cursory and ephemeral states. The exact nature of a a set of potential objects to form an exhibition or the set of information objects that constitute a complete area of argumentation in a text are objects of scholarly

documentation and elaboration and form part of the understanding of history from a social development perspective. The introduction of this class was motivated by work on the Bernini Tagging Project which aims to document potential connected areas of thought within texts. This is just one example of many that could be given for this class. We did not derive at this point a conjunct declarative speech act; this will form an object of further research.

Conclusion

The proposal for CRM AAA comes as part of a general project, “Sustainable Art and Architectural Data Reuse in Teaching and Research.” CRM AAA attempts to propose an important new, systematic foundation which builds from the modelling foundations and principles of the CIDOC CRM but offers an extension which expands the expressivity of the existing model to cover characteristic areas of research and documentation in history that CRM base, because of its declared scope and functional aim (serving a prentist knowledge paradigm) cannot support. CRM AAA leans strongly on the theoretical foundations of Searle and Austin while trying to take into account the work of Gangemi as well. In order to move from a proposal to a concrete system to be used and tested, the ideas outlined and illustrated in this report have been systematically elaborated into an official specification and serialization of CRM AAA. These are delivered as part of the overall deliverables of this project.

References

Austin, J. L. *How to Do Things with Words: Second Edition*. Edited by J. O. Urmson and Marina Sbisà, 2 edition, Harvard University Press, 1975.

Searle, John R. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge University Press, 1983.

Searle, John R. *Making the Social World: The Structure of Human Civilization*. Oxford University Press, 2010.

Searle, John R. *The Construction of Social Reality*. Free Press, 1997.