Notes on the data modelling meeting in Crete July 1997

Introduction

This document consists of notes on a draft specification for the CIDOC OO reference model. This draft was compiled by an ad hoc group (Nick Crofts, Costis Dallas, Ifigenia Dionissiadou, Martin Doerr) at a meeting in Crete in July 1997 on behalf of the CIDOC documentation standards working group.

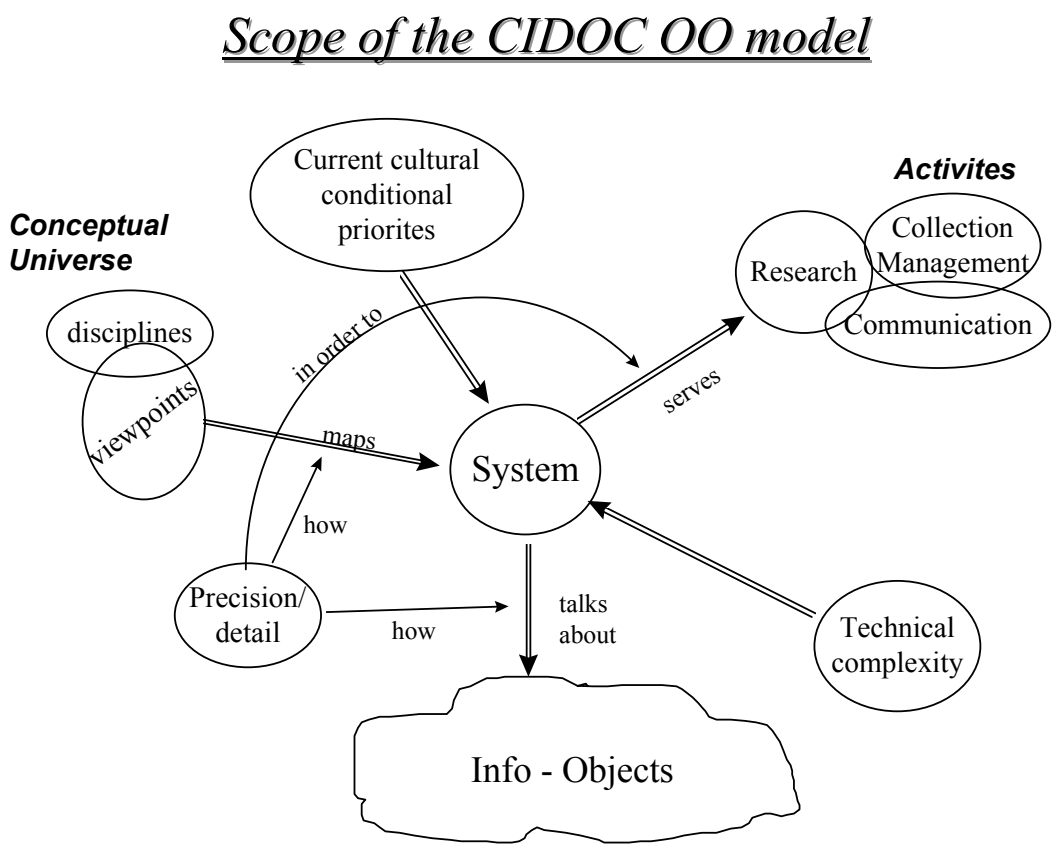
The draft OO model will be presented to the full working group for discussion, approbation and ratification. The current document should be considered as primarily as explanatory remarks and recommendations intended to clarify the model. The intended audience is the CIDOC documentation standards work group. This document is not intended for distribution to the wider CIDOC community

Goals of the Crete meeting

Following up on the work already done on the CIDOC OO reference model, the goal of the Crete meeting was to produce a semantic model, based on the *CIDOC International Guidelines for Museum Object Information*. This document defines a set of information categories but does not presuppose any particular data structure. It provides a useful basis for elaboration of the data model since it covers a broad range of information needs without imposing constraints on the way in which these needs are met. The data model can be seen as complementary to the *Guidelines*, providing a structured semantic framework which reveals the implicit relationships between the information categories.

Scope of model

The following definition of the scope of the model represents the position adopted by the group in Crete. The formulation is intended to make explicit both the current and potential scope of the model.



The general domain of the model is information systems for cultural and scientific collections; it is *not* intended as a universal description of reality. The scope of the model is further constrained by :

1. The **conceptual framework** (viewpoints) of the intended users (scholars, museum professionals and museum visitors, etc.)
2. Common museum **activities** (collections management and conservation, research and analysis, promotion and communication)
3. The **objects** collected by museums
4. The level of **detail** and **precision** required to provide an adequate level of quality of service.
5. Considerations of **technical complexity**.

All the above are conditioned by **current cultural priorities**. The pragmatic needs of museums, as reflected by the subset of potential information which is actually covered by current standards. (In general, information elements should not be introduced in the model which have no corresponding information category in any existing standard. However, where necessary for clarity, the model may need to be completed to maintain semantic coherence.) The model will have to evolve over time.

Clearly the *CIDOC Guidelines* are a subset of the museum community's potential information requirements. Modelling the *Guidelines* represents a first step. Future work on the model should extend its scope to encompass other information standards such as Africom, Archaeological sites, Ethnography, etc.

Initially, the scope will be further constrained by the practical need to organise the work involved. The group intends to adopt a 'top down' approach - avoiding unnecessary detail and giving priority to specific areas of interest (and continuing just so long as the chocolate holds out).

Nature and purpose of the model

The model can be considered as a formal knowledge representation of high expressive quality.

The model is a « semantic model » in that it aims specifically to clarify and define concepts. Work-arounds and compromises associated with particular implementation constraints are specifically avoided. The reference model is not intended *primarily* as a basis for implementation (although a direct implemented using an OO database system is perfectly feasible.).

The model can be used for :

- Global queries - with guaranteed data recall and precision properties.
- Data interchange with a defined degree of data preservation and precision properties
- Certification - allows comparison and evaluation of different systems.
- Implementation - as a reference design for the development of database applications.
- Resource document - as an aid to understanding museum information.

The OO model aims to preserve the highest degree of expressive power and to define ways of scaling down to 'poorer' systems rather than limiting the scope to the lowest common denominator. It is hoped that this will enable the OO model to correlate consistently with other, less detailed or more restricted standards. (To resolve certain interoperability tasks.)

Methodology

The approach adopted by the Crete group was essentially *top down.* Each information group was considered in turn and an attempt was made to construct a partial model of the information categories it contained. At the end of the week these partial models were combined into a global model and scope notes were added. Although it was generally avoided, the OO model does, in some instances, go beyond the requirements of the *Guidelines*. This was thought necessary to ensure a consistent treatment of similar problems and to allow for future extensions.

Comparison with other standards

Z39.50

Z39.50, uses a 'flat' access model, or set of attributes, designed for ease of use and maximum coverage rather than a structured data model, so that even a well defined mapping entails loss of information with respect to relational or OO schemata. A 'recall preserving' mapping is possible. Obviously, mapping between the CIDOC OO model and the CIMI Z39.50 profile should be investigated.

Dublin Core - Warwick framework

DC WF allows for structured records but concerns meta-data. There is some overlap of semantic categories with the current state of the Guidelines and OO model : recorder information, subject information and other information, in the case of museum objects which are documents. Mapping needs to be investigated. There is a particular interest in enabling access to information about specific museum objects referred to in documents.

Thesaurus

There are four types of authorities :

- Subject and types (classifications), such as AAT,
- names (individuals) such ULAN,
- geographical hierarchies e.g. TGN,
- chronological hierarchies.

Names, can be characterised as instance authorities, whereas subject and type are comparable to class hierarchies.

In order to correlate subject and type thesaurus hierarchies, or parts of hierarchies with the OO model, a specific level of concordance is required : each class in the OO model should have a corresponding class in the thesaurus and the BT links should not be in contradiction with the isA hierarchy.

The OO model requires the use of at least one correlated type hierarchy (which may be considered as a set of thesaurus) for each class hierarchy, or *facet* of the OO model. However it allows the use of multiple, correlated but different thesauri for each class hierarchy. Uncorrelated thesaurus may also be used although functional equivalence with the OO hierarchy is lost.

The documentation standards WG should consider the need to harmonise work on terminology and thesauri with the development of the OO hierarchy.

General design considerations

'Type' Class hierarchy

- Preservation of data between heterogeneous systems
- Reference points within the model allowing uniform access of relational and OO systems

An special 'type' class has been included in the model which is systematically referenced by a type attribute in each class. The type class defines a parallel isA hierarchy which duplicates (and may extend in detail) the structure of the model's class hierarchy. This type attribute included in each class references, often redundantly, the name of the class ; however, it may also reference a more detailed subclass which is not present in the OO structure. This allows specific implementations to extend the level of detail and precision without compromising the coherence of the OO model : unsupported 'local' subclasses can be collapsed into the their corresponding reference class in the CIDOC model.

This hierarchy of types can also used to indicate other, non-standard hierarchies and conceptual classification systems.

'Short-cut' joins

Joins eliminate intermediate classes.
Developers are encourages to implement either the short cut or the expanded version.
The global model contains both alternatives and identicates the relationship between them.
A default expansion rule can be defined for global queries and data transfer, typically be introducing a 'unknown' object of the most general compatible state and type.

e.g. a condition assessment event can be generated, for systems which do not keep a condition history, which states that the date of assessment took place on or before the date of the query

'current' information

The model should not contain fields for 'current' information. Instead, a 'history' is provided, the most recent entry is the current one.

Short cut joins are often used to indicate frequently updated 'state' information, used primarily for collections management : e.g. location, owner, physical custody, condition.

HISTORY information. Not all fields require history. Location has one, ownership too. Condition information. Most recent item in history is current *state* (calculated from). Some systems may store only current state and not history.

Constraints

Good practice constraints have been deliberately omitted. Implementors can enforce constraints. The model is explanatory and descriptive and not prescriptive.

Cardinality

Cardinality constraints are logical, not physical. Implementations may want to take them out.

Typed relations (Intersection entities)

Relations between classes are typed, in order to achieve a formal definition of the semantic correlations between different relations. This gives rise to relation hierarchies, analogous to object hierarchies. These families of relations are called ‘meta-categories’.

Implementers of OO databases are encourages to expand typed relations into sets of distinct attributes rather than using a variable type associated with a relation.

Systems which do not have this possibility may use intermediate classes to implement typed relations.

Direct links between objects

Certain types of direct links between objects are discourage the use of direct links between objects. Unnecessary and hinders chronological information.

‘Lost object’ information needs to be introduced in order to normalise the model. NB prototype objects. If the prototype object is missing. ie create hypothetical prototype object.

Any assessment of relations between objects should arise from
1. shared physical attributes
2. shared events (causality)... process, resource etc.
3. Component
4. Representation (subject - object)

Alternatively, relation is the result of intellectual judgement.

A collections management system ‘describes’ an object. A ‘study’ describes it too.

Negative information

NEGATIVE INFORMATION is not stored.

Text fields

In OO system, all data make reference to *real* objects. There is often a need to record more detailed and richer information than is captured by class attributes. A text field should be included to help disambiguate. A type field too.

Notes on the information categories :

Acquisition Information		
	Acquisition method	
	Acquisition date	
	Acquisition source	
Condition Information		
	Condition	
	Condition summary	
	Condition date	
Deaccession and Disposal Information		Specialisation of transfer of ownership. However, disposal is a change of physical state. (like creation) Deaccession and disposal are two separate notions. Legal and physical.
	Deaccession date	
	Disposal date	
	Disposal method	
	Disposal recipient	
Description Information		
	Physical description	attribut text of object.
	Specimen status	Out of scope - (ie natural history) nb does not depend on physical attributes of the object.
Image Information		Images are specialised cases of objects. (A collection object may be an image of another object). The implicit prescription that objects should be photographed is not represented in the model. Constraints *could* be included to enforce good practice. Sub class of reproduction... sound recordings. Reproduction or référence are two types of relation between objects.
	Image type	See note on Type meta-class. Specialisation of object type.
	Image reference number	Images are objects related to museum objects. A persistent link exists between the two. This link effectively means that a photo or picture is a good picture. (Information about the event leading to the creation of the photo is implicit, ie author and date, etc.). The reference number itself is a specialisation of object number.
Institution Information		A legal person (specialisation of agent) - short-cut link to object through 'current owner' to avoid acquisition event ? Acquisition implies ownership, having responsibility does not NB Champ pas très clair. Est-ce le propriétaire de l'objet ou le propriétaire de la fiche ? Dans les deux cas, modélisé par lien entre objet et agent - raccourci, via événement ? stockage physique de l'information. Copyright, ou 'signature de la base' provenance physique de l'information. Le plus interessant semble être l'info sur la gestion de l'information. (Lien directe sur agent.)
	Institution name	
	Institution sub-body name	
	Institution address	

	Institution country	
Location Information		Set of 'moved to' events and dates associated with this. See notes on history. « Placé à » événement - agent, et lieu, motif(événement). Localisation habituelle n'est pas un événement dans l'historique, mais un attribut de l'objet. (type lieu).
	Current location	
	Current location date	
	Current location type	
	Normal location	
Mark and Inscription Information		Intellectual (information) object (symbols) carried by a physical object. Isa hierarchy - inscription and mark. Field needed to indicate language used for translation. How systematic does this need to be ?
	Mark/inscription text	transcription of text - how do you transcribe a mark ?---voir IMAHGE. Rapport entre marques, signatures et inscriptions ? voir thesaurus IMAHGE marks et signatures. Attribut multiple de l'objet de type 'objet intellectuel'. Les attributs, position, technique etc. sont dans le lien. NB Une inscription n'est pas, en générale, associée à n objets.
	Mark/inscription type	
	Mark/inscription description	
	Mark/inscription technique	
	Mark/inscription position	A specialised piece of information if used as an access point.
	Mark/inscription language	
	Mark/inscription translation	
Material and Technique Information		A technique is a specialisation of plan. A 'text container' may be used in preference to a structured approach.
	Material	materials class
	Technique	
	Part or component description	Parts can be of different types.. integral, separable.
Measurement Information		
	Dimension	
	Measurement	
	Measurement unit	
	Measured part	feature measured
Object Association Information		History of events associated with the object. Open to specialisation.
	Associated place	
	Associated date	
	Associated group/person name	
	Association type	
	Original function	'original' ? points to type in type hierarchy. Used for - activity type (actual use) used for once - event instance made for - event instance (intended use)

		made for - activity type.(intended use) See diagram below... (function use)
Object Information	**Collection**	Could be seen as initial event in ownership history. (Provenance) Collecting is an specialisation of property event.
	Collection place	
	Collection date	
	Collector	
	Collection method	type of collection method - link to type hierarchy.
Object Information	**Entry**	When the object came in. Ownership is documented thru acquisition events. A short cut gives the 'current' owner. Entry is a physical transfer event. Another short cut gives custodial responsibility.
	Current owner	Current owner at moment of transfer.
	Depositor	
	Entry date	
	Entry number	
	Entry reason	types of object entry ? and other information.
Object Information	**Name**	'name' is misleading - type or class is preferable. The object is classified by its position in the class hierarchy. This classification can be modified, or even contradicted, by the classification event. The classification event should be in a textual description. However, the classification itself needs to be queried.
	Object name	
	Object name type	Which classification system is used. (authority type)
	Object name authority	not interesting for interrogation Distinction between prefered type and other types. - to indicate 'authoritative' opinion.
Object Information	**Number**	Only one identifier should be used. Associating date and type complexifies the identifier. Should be separated into other, non prefered, numbers.
	Object number	The number and type become the identifier.
	Object number type	
	Object number date	
Object Information	**Production**	Creation of an object. - in intermediate 'rôle' class is needed between creation and agent. This allows specialization of relations.
	Production place	
	Production date	
	Production group/person name	
	Production role	How someone contributed to the creation.
Object Information	**Title**	Appellation node with title entity (naming) relation with typing. Same problem with translation with language - textual object with translations (like inscriptions).
	Title	
	Title type	
	Title translation	Is the language of the title an access point ? Title is a subclass of intellectual object, itslef a sublass of text objects which inherits the 'language' attribute. The type of title is contained in the relation with the object. Translations are also subclass of intellectual objects (with language attribute). The relation between object and title is a subclass of 'name calling'. (see title model)
Part and Component		Should be a sub field of physical description. Could be

Information		text field
	Number of parts or components	
	Description of parts and components	
Recorder Information		• Per Record (optional) • Generate dynamically a 'database' signature : ex SIBIL GENEVA 20/4/1997, at moment of retrieval • Finer granularity ... text format. Who is the 'author' ? The authority should be distributed, rather than the data entry person.
	Recorder	Optionally per object record Dynamically per database
	Record date	
	Authority	contrôler, supervisor,
Reference Information		
	Reference	
	Reference type	
Reproduction Rights Information		Generalise to 'rights' information : given to Agent. About, object or intellectual object, type of right and description. (text)
	Reproduction rights note	
	Reproduction rights owner	
Subject Depicted Information		
	Subject depicted	
	Subject depicted description	

Name is used to refer to individuals
Class is used to identify groups of objects.

Relation with categories - the model is not better, it is different.

Philosopher, physicist, archeologist and diplomat.

Damassio - Descartes' error (or mistake).

Flat systems (ex z39.50) have to specify relations as separate information groups, eg the owner of the object, the artist, the subject of the painting ... all are 'people' but the model loses this fact. An interrogation on persons linked to the object requires multiple 'or'.

AUTHORITY information. Can be recorded for each piece of information - which level of granularity is needed ? Ex. attribution. NB an absolute authority is required to provide STATE information.

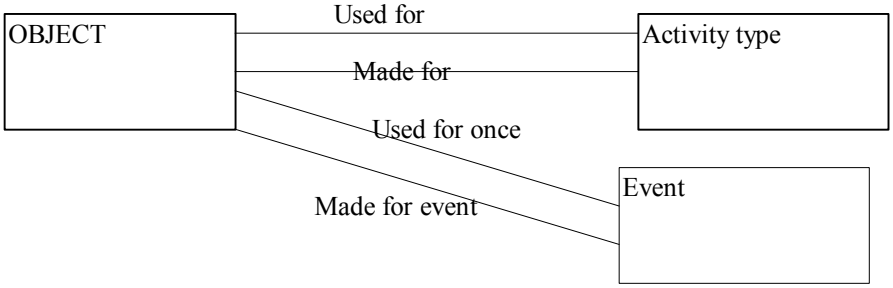
It is assumed that the cultural universe being modelled is sufficiently coherent to be dealt with without explicit ...

The type or classification of an object depends on the intention (purpose) of the description. Ex arable farms, profitable farms.

Plans are potential events. (abstract events). ?

Changes of detail to information categories... the categories contain some fields which are not access points - ex date of inventory number. This distinction is of greater importance in the model than in the IC and becomes apparent as structural differences (attribute or text field).

Function use



Representations

- real objects
- impressions of real objects
- Concepts, symbols
- fictitious reality. (fantastic constructions).

Apparent 'realism' of object is distinct from factual realism. 1 cannot be physically distinguished from 4.

Object
Physical/intellectual.