Archeologia e Calcolatori
7, 1996, 1047-1060

MODELLING OCCURRENCES IN CULTURAL DOCUMENTATION

1. Introduction

The documentation requirements of cultural goods range from keeping a simple log of objects to the management and conservation of collections and even to the recording of a variety of cultural information. We distinguish documentation into administrative, using a fixed set of data mainly to support administrative functions and to provide basic information, and cultural, aiming at organizing an evolving body of knowledge about objects, to be used in scientific study and research. Cultural documentation requires a system capable of recording the entire variety of information, which constitutes the current knowledge about a set of objects. This includes formatted data and other, multimedia data (images, audio and video recordings, text...) and are characterized by a high degree of linking, large variety of references and classifications, broad usage of abstract relations and the need for multiple, mainly referential access. Moreover, this information is in general enriched but rarely modified. Ordinary administrative documentation systems and hypermedia information bases do not meet all these requirements.

The CLIO system was developed at the Institute of Computer Science, Foundation of Research and Technology – Hellas to fulfill the requirements of cultural documentation. Information is organized in CLIO as a knowledge base according to a specifically designed semantic model. The functional kernel of CLIO is the Semantic Index System (SIS, see Constantopoulos, Doerr 1993), built at the Institute of Computer Science. The construction of CLIO allows extremely dense linking of information, access by unlimited chained references, expression of historical and cultural context as well as of abstract properties, joint temporal and spatial assignment in absolute or relative terms, and recording alternative, possibly conflicting information along with the respective sources. Information is presented in graphical or textual form. An extensible list of predefined queries is offered. A particularly important feature is the uniform treatment of schema and data, enabling the immediate extension and modification of the schema by the users themselves. The analysis of requirements for the CLIO system has been performed in close cooperation with the Benaki Museum and the Historical Museum of Crete. The development of the system was partly funded by the STRIDE and ESPRIT programmes.

A central problem in recording historical information and cultural documentation in general, is how to render the fundamental notions of time, space, events, existence, life, activity, causality, etc. Particular kinds of events, sequences of events, or even temporally unordered sets of events, each with their own characteristics and significance can be distinguished. In this paper we address these notions, referring to them collectively as notions of occurrence, within a conceptual modelling framework and in the context of developing a general ontology for cultural documentation. Particular attention is given to the representation of relations on which historical and other inferences can be based.

2. Modelling framework

2.1 *Knowledge representation*

The structural part of the Telos knowledge representation language (Mylopoulos *et al.* 1990) is used for representing knowledge in CLIO. This offers the general mechanisms of attribution, classification and generalization, common in all semantic, conceptual or object-oriented representation schemes, and, in addition, it supports unbounded classification hierarchies and equal treatment of relations and entities, resulting in great expressiveness and flexibility. Higher levels of classification enable the expression of abstract or general properties which are of interest as such even when instantiated differently in particular cases (e.g., partition, overlapping, creation, measurement interval). The usage of multiple generalization enables the creation of entangled hierarchies, suitable for representing complex terminological systems and faceted classification schemes, in addition to great economy of expression.

2.2 *Ontology of CLIO*

The knowledge representation model of CLIO expresses a rather general ontology which can be complemented by derived concepts, thus specialized to particular fields. This ontology includes concepts of matter, location and chronology, occurrence, quantity, mankind, conceptual creation and naming, as well as relations between them, and is presented in detail in Christoforaki *et al.* 1992. A similar ontology has been developed in Lenat *et al.* 1990 for natural language processing purposes. Here, we are restricted to a very brief review.

The concepts of matter account for the kinds and the structure of physical objects, the elements of their appearance, the distinction into natural and artificial, composition or construction materials, special information concerning fine art objects and museum objects, descriptions of style and ornamentation, and kinds of tools.

The concepts of location define absolute and relative location, orientation and various topological relations. The concepts of chronology define absolute and relative chronologies, succession and inclusion relations.

The concepts of occurrence introduce certain fundamental notions, such as event, as a mean-value point in time and space; period, as a set of events delimited in time and space; and existence, as a directed, causal line of events within a period. They include categories of events of particular significance and such concepts as activity, use, creation, technique, etc.

The concepts of quantity distinguish between physical, arithmetic and numismatic quantities, elaborate their kinds, and define ways of measurement. The latter are distinguished on one hand into absolute, defined with respect to conventional coordinate systems (e.g., chronology, location), and relative, expressing magnitudes in given measurement systems (e.g., time, distance, area, volume, weight), and on the other into exact, yielding a single value, and approximate, yielding a range of values.

The concepts of mankind concern the description of human persons and organizations, such as actor, role, group, membership, institution. The concepts of conceptual creation essentially differentiate the product of spiritual creation from its physical embodiments. Finally, a system for naming and identification is defined.

The CLIO knowledge representation model subsumes the notions and relations of the CIDOC/ICOM fine arts documentation standard of 1990.

3. Events and other notions of occurrence

Events, their relations and mutual relevance play an essential role in history. The selection of events and relations under consideration are always subject to the aspect under which the respective research is done. Nevertheless we may regard the events themselves to be among the most reliable knowledge different researchers agree on.

In classical mechanics, an event is defined as a dimensionless point in absolute space and time. Just as in modern physics this view is abandoned in favor of a concept of interactions, it is also not very useful for a consistent description of history for two reasons.

First, historical events can never be precisely defined to a point in space and time due to their nature. We must by far more regard them as processes. A birth e.g. lasts some hours and may take place in a room. A battle may last days or months, and span over a thousand kilometers. There is no clear meaning of the begin and the end, and the spatial confinement. References are given with a precision at most close to the duration and spatial spread of the event, but usually in the order of magnitude of the event's consequences. For a birth, these consequences are the new existence of a human, new family relations, a new inheritance right. Events of historical interest have a result, they denote at least a change of status or knowledge. We are interested in phenomena, not in plain points in time and space, as e.g. birth and death, creations, information exchange, natural catastrophies. They all can be seen as interactions of one or more alive or dead (e.g. a bullit) actors resulting in changes on the participants (knowing, death, murder) or creation of something new (baby, painting, poem). Modern science regards even information transfer and storage as to be mediated by energy, thus obeying physical restrictions.

Second, for most events, space and time are unknown or uncertain. If given, often enough they are conclusions from other facts, opinions, or based on unreliable sources. The most certain facts about events concern their participants and outcomes. They are more often decidable as true and false, and as such rather suitable for the limited possibilities of digital computers. For instance, there is really no doubt about the existence of El Greco, the painter, which implies that he was born. Opinions on date and place of his birth are subject to change, or not commonly accepted. The first historical reference of El Greco, however, allows us to derive estimates on time and place of his birth.

There are characteristic interrelations between participants and interactions on the one side, and time and place on the other. Temporal succession within a certain (cultural) space lets us conclude on a possible influence or causality. Most Japanese soldiers e.g. learned about the end of the World War II within days and hence stopped military actions. Some however did not for a long time. Temporal precedence excludes any influence or causality, with precautions however about the precision of the data given. Only if an event was earlier than another by more than the characteristic duration of the process, we can be sure about it. See e.g. laws on fatherhood in cases of a longer absence of a husband. Causal relations between events, i.e. the outcome of an event in terms of a creation, thereby acquired status or knowledge, which participates in another event, provides us with absolute knowledge on temporal sequence. Moreover, if a low enough maximum transfer speed of participants from the one event to the other is known (e.g. the speed of a horse), relative knowledge of the distance between two causal related events gives lower bounds for the minimal time span elapsed and vice versa.

To create a stable base of knowledge, we must integrate causal relations as primary sources of temporal and spatial reasoning, which allows us at any stage to follow automatically or by hand the reasons which lead to an estimate of time and place. This stays in contrast to an alternative approach, which would normalize all events to an absolute chronology and a point on a graphical information system as primary step. In particular, we must be able to describe place and date of an event indirectly as a series of conclusions from the types of relations it has to other events. Such types are upper and lower bounds and identity ("same as"), which can be attached to categories as "created", "found" etc. or be directly given. We can achieve this way, to replace incomplete knowledge on time and space by "good" knowledge on interrelations between events. Any further fact added in the future will not invalidate the previous, but refine the precision of our knowledge. The first historical reference of a person e.g. is an upper temporal bound for its birth. Such a conclusion can be automatically drawn by a computerized system, and it will never become invalid.

Since any information on the past may be wrong, such a system results in a complex network of mutual dependencies, where the computer can do valuable work to validate hypotheses on erroneous information by tracing inconsistencies under various assumptions. By maintaining relative offsets between events, chains of relatively determined events can be "recalibrated" on reliable fixed points in time and space according to our trust in those.

We may define in a classical way the coordinate aspect of an historical event $e := (x, t)$, where the definition of x, the place and t, the date, is to be understood as "fuzzy", i.e. a medium value for the respective process. In the CLIO model however, we refer the place by usual geographic, political and cultural expressions as: Greece, Acropolis, Saaba, White House, which actually have an area. No attempt is made to model a dimensionless point. The chronology is given in arbitrary precision, e.g. 1453 AD, which is actually an interval of one year. We also foresee to define uncertainty bounds as: "between 1400 and 1500 AD". I.e. the event's coordinates are approximated by an area and a time-span enclosing them. For places, we model inclusion, overlap and adjacency, creating thus a primitive but efficient topology.

Historical sources often refer to important events as determination of place or date, the knowledge of which may be lost. Hence between events, we are also interested in partial equality, i.e. "has same date with", "has same place with". Equality is understood as an overlap of the relevant temporal or spatial spread of the events under concern. For unknown dates we also model "before" or "after" the date of another event. Finally, we include the distance and direction of unknown places to known or unknown ones, and temporal distances between unknown dates, if these are known. Since categories (fields in the traditional sense) in CLIO are optional, the user selects the appropriate fields for his kind of knowledge, and leaves other fields open. The relations for dates may be extended by more detailed notions of set intersections.

The intention of the work presented here was however only to capture formally usual historical references. Similar relations for time are used in other temporal models. They do not take into account the possible dependencies between time and place, e.g. EDELWEISS *et al.* 1993; TOLBA *et al.* 1992; ALLEN 1983. Temporal conditions can be separated from spatial conditions, if the characteristic speeds within the modelled system are high compared to the spatial size. More precisely, if the smallest time unit we use is big enough for an item under concern to travel across the system, a separation is justified. For example, if we use a year as time unit, we can assume that political news became "immediately" known even in an ancient state. But cultural characteristics like style, technologies, usage of letters etc. usually spread out very slowly, stop at frontiers and "leap" over them.

In archaeology, artifacts may be dated by style, material, location of finding, other artifacts found nearby, inscriptions and physical measurements, (radiocarbon or thermoluminescense). The basic property of notions like style, location of finding and usage of a material, which allows conclusion on the dating, is their confinement to a space and time interval, i.e. they define an event space. This confinement needs not to be given directly, analogous to the above, but it may be restricted through its events merely by the "belongs to" property, or other semantic relations, in the sense that a style ends with its last (true?) manifestation, if not better known.

Any set of time-place points can be an event space (period) in our sense. As in events, we are only interested in periods associated with certain cultural manifestations, i.e. events of a certain kind. In other words, a set of certain manifestations is understood as period, and not necessarily all events in a certain time and place. Usually, such periods are confined by a certain region, e.g. a country, a city, a battle field, and a time interval. But such separation of time and place often does not hold. The Gothic style e.g. in Spain, France, and Germany has different time limits. If we conclude from style to date, the knowledge of the place of creation may allow for more precise dating. In our model, we restrict the representation of the time-space aspect of a period deliberately to a collection of place / time-interval primitives, an approximation by outer bounds, which may be refined in the future.

We can now establish relations rather close to those given by history and archeology between events and periods, and periods with partial knowledge of time and place, or even without any such knowledge. From the set of possible relations we model for periods inclusion, and total temporal succession. Other relations are given through the respective place and date relations. Between events and periods we model inclusion. Any event may be "upgraded" to a period, if the description requires more details, without invalidating previous information. For example, a battle is an event with a winner, if any. If we are interested in the future in its phases, we see it as a small period. Summarizing, the CLIO model allows to correlate temporally and spatially events and periods without separating date and place and even without knowledge about the latter.

In the following, periods are associated with style, cultures, states, kingdoms, communities, cities, and technologies. Events are specialized to creations, production events, birth, acts, actions and uses. A technique is a class of production events. A usage, cooking e.g. is also a class of respective events, like "I cooked". In the CLIO semantic network, we can follow paths of the above relations, which provide us information for temporal and spatial constraints. These may be evaluated by interval arithmetic to resulting bounds. Historical references contain however errors or even lies. In a network of mutual references, there is often no formal way to decide which information is erroneous. Any historical reference is a fact by itself, we want to store, besides our confidence in its contents. Hence our model does not intend to enforce any constraints, and leaves the contradiction resolution to the researchers opinion.

As mentioned above, the primary aspect of an event in our model is the process, or interaction associated with it. Similar to the Feynman graphs (FEYNMAN 1965) known from quantum mechanics, we regard:

1. The initial state. It consists of the objects, persons, institutions participating in an interaction.
2. The interaction. This corresponds mostly to a verb in natural language, like bears, meets, cooks, produces, writes etc.
3. The final state. It consists of the objects, persons, institutions "surviving" an interaction, and the newly created ones.

Objects, persons and most institutions share the property, that they exist in a physical or legal sense. We introduce the notion of an "existence" which is characterized by a beginning, birth, creation, etc., that it is found at one place at a time, that it may move, and eventually terminates. With an existence, we can associate a "worldline" in the sense of the special theory of relativity (BORN 1962), which is the path of the existence through time-space.

Events in our sense are regarded as the known "fixed points" on the "worldlines" of the participants. The history of an existence in a narrow sense under this view is the set of known events on its worldline. This description is independent from the state of knowledge on time and place.

We distinguish three basic types of events, the "meeting" type, where participants survive, the "creative" type, which starts a new worldline, and the "destructive" type, which terminates a worldline. Real events are combinations of these types (Fig. 1).

Due to the causality principle, we are able to establish the following deduced temporal relations between events:

1. The "creative" event of an existence is a lower bound for its history. I.e. all events connected by a "meeting" or "destructive" type with an existence, are "after" its "creative" event.
2. The history of an existence is an upper bound for its "creative" event.
3. The "destructive" event of an existence is an upper bound for its history. I.e. the "destructive" event of an existence is "after" all events connected by a "meeting" or "creative" type with an existence.
4. The history of an existence is a lower bound for its "destructive" event.

If spatial boundaries can be derived for an existence, also spatial relations can be derived in an analogous way through assumptions on maximum speeds.

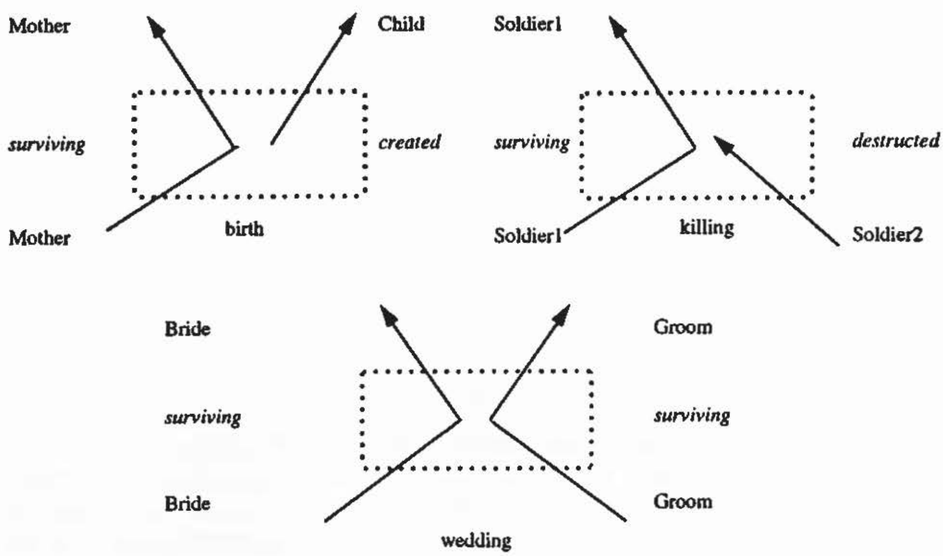

Fig. 1 – The three basic types of events

4. A MODEL OF OCCURRENCES

As Telos allows the user to extend the schema at runtime, our model contains only the basic notions. In particular it contains the full definitions at the meta-level, which allow to create specific event, existence and period classes at simple class level consistent with the process and coordinate view. An event class is characterized by a gerundial form of the respective verb. Attributes connecting items and events are grouped under meta-categories according to the role of the referred item in the process. The meta-categories distinguish between surviving, created and destructed participants. The attributes are named in the terminology of the specific event group. A soldier e.g. "was_killed_in" a fighting. "was_killed_in" must be an instance of the "destructive" category. Thus at token level, the process nature of an event is implicit in specific categories, and the user needs not to think about these structures. We are however able to query on this implicit knowledge.

We present now parts of the CLIO model dealing with occurrences (Fig. 2). Occurrence is the set of classes describing items limited in time:

```
TELL Individual Occurrence  in M1_Class, Notion with
  attribute
    history: Occurrence
 end
```

The category "historical" groups connections between worldlines and events. Occurrence is specialized into Existence, the set of classes describing items with worldlines:

```
TELL Individual Existence in M1_Class isA Occurrence end
```

Occurrence is further specialized into EventType, the set of event classes:

```
TELL Individual EventType in M1_Class isA Occurrence, SpatialMeasureType,
TemporalMeasureType end
```

EventType is besides others a compound of spatial and temporal definitions. We connect EventType with the following specializations under the "historical" connection group:

```
TELL Individual ActionType in M1_Class isA EventType end
TELL Individual CreationType in M1_Class isA ActionType end
TELL AttributeClass created_by
  component
    from: Existence
    to: CreationType
in M1_Class isA historical end
TELL Individual DestructionType in M1_Class isA EventType end
TELL AttributeClass destructed_by
  components
    from: Existence
    to: DestructionType
in M1_Class isA historical
end
TELL AttributeClass participating
  components
    from: EventType
    to: Existence
in M1_Class isA historical
end
```

We express above, that creations need an actor. Destructions may be spontaneous, or we are not interested in the existence of the destructor, as a virus in a natural death. The union of all event classes is represented by the simple class Event, which carries the categories date and place common to all event classes:

```
TELL Individual Event in S_Class, EventType with
  temporal
    date: Date
  spatial
    place: Place
  spatial, temporal
    within: Period;
end
```

The "temporal" and "spatial" category group from Temporal Measure Type and Spatial Measure Type respectively, characterize connections transferring temporal and spatial constraints. They are used in combination to denote transfer of both. Telos categories are optional. They need not be used at instance level.

The union of all period classes is represented by the simple class Period, which carries the common temporal and spatial categories, as well as the partitioning into space-time intervals:

```
TELL Individual Period in S_Class, Temporal Measure Type, Spatial Measure Type with
    temporal
        begin: Date;
        until: Date
    spatial
        region: Place
    spatial, temporal
        within: Period
    partitioned_in
        consistsOf: Period
end
```

Subject denotes physical and legal persons, we are interested in the actions of:

```
TELL Individual Subject in S_Class,Existence end
TELL Individual Person in S_Class,Existence isA Subject with
created_by
    born: Birth
destructed_by
    died: Death
end
```

We characterize an action as an event performed by a subject:

```
TELL Individual Action in S_Class,ActionType is A Event with
participating
    subject: Subject
end
TELL Individual Birth in S_Class,CreationType, ActionType is A Action end
TELL AttributeClass mother
components
    from: Birth
    to: Person
in  S_Class isA subject
end
TELL Individual Death in S_Class,DestructionType isA Event end
TELL Individual Killing in S_Class,DestructionType, ActionType isA Death, Action end
TELL AttributeClass was_killed_in
components
    from: Person
    to: Killing
in  S_Class isA died    end
```

Death includes natural death, whereas a killing has the aspect of a death as well as of an action.

For illustration we give an example of the token level (data level):

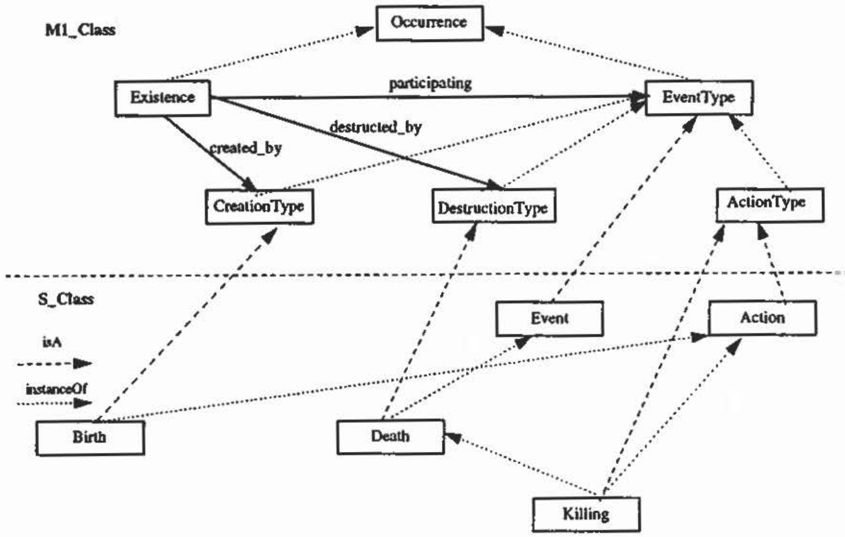

Fig. 2 – Modeling occurrences in CLIO.

```
TELL Individual Soldier1 in Token, Person end
TELL Individual Soldier2 in Token, Person with
was_killed_in
    : Soldier2Death
end
TELL Individual Soldier2Death in Token, Killing with
subject
    : Soldier1
in
    : WorldWarII
end
TELL Individual WorldWarII in Token, Period end
```

5. Usage illustration

In the following example we demonstrate the historical documentation of an artifact using events. Specifically, we try to determine the creation date of an ancient Egyptian scarab found in 1965 in the Minoan city of Lebena in Crete. For identification purposes the artifact is named "Scarab1".

Scarab1 was found in the MM1A excavation layer in Lebena, which is dated by the archeologists in the Prepalatial Minoan period, spanning from 2700 BC to 2000 BC all over Crete. On the other hand, the style of Scarab1 was identified as the one of Egyptian objects created during the XII Dynasty of Egyptian Kings. This period spans from 1991 BC to 1786 BC in Egypt, and is included in the Middle Egyptian Kingdom. The exact creation date

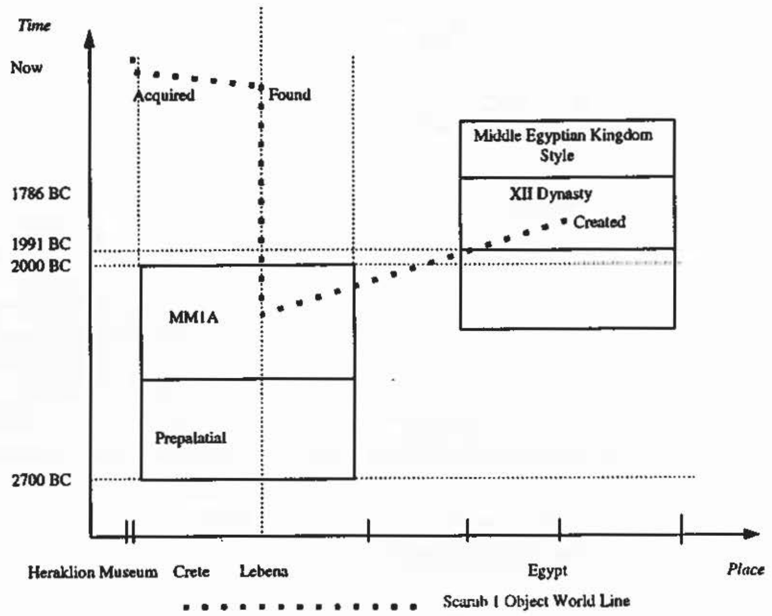

Fig. 3 – The worldline of scarab1

and place are unknown, and today the artifact is been held in the Archeological Museum in Heraklion. Fig. 3 shows the "Worldline" diagram of the object, as it is determined from the events known to us.

These are the creation (though not known when and where it was obviously created), the finding and finally the acquisition of the object by the museum. The dating of the object is based on the information deduced by the object itself (its style) and the dating of the excavation layer it was found. In the time axis, Fig. 3 shows a contradiction derived from the time limits of 9 years between 1991 and 2000 BC. Obviously, the end of MM1A is given with a precision of about 50 years, and an archeologist may not regard the difference as a contradiction. Perhaps the cultural phase MM1A in Lebena, at the south coast of Crete, lasted also longer than in the major Cretan cities. Also an intrusion may have happened, i.e. the scarab was mechanically transferred to deeper layer.

In Fig. 4 the same information is represented in the CLIO model. The MM1A excavation layer, Prepalatial period, XII Dynasty style and Middle Egyptian Kingdom style are represented as periods, while the creation, finding and acquisition of Scarab1 as events.

Using the terminology of TELOS, the rectangles in Fig. 4 represent individuals while the links between them attributes. The link labels are the names of the respective attributes.

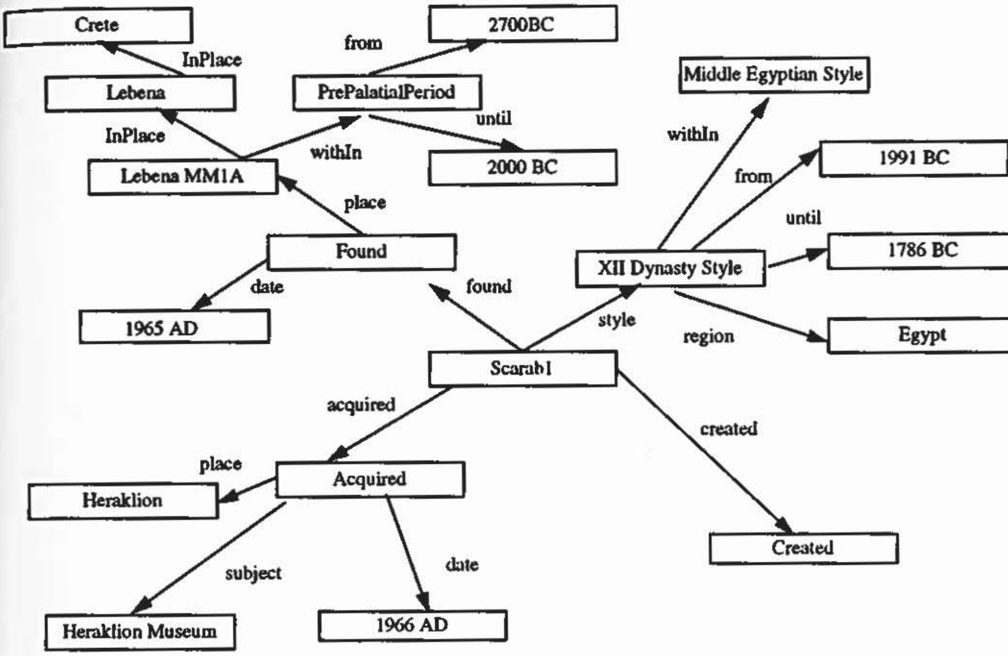

Fig. 4 – Historical documentation in CLIO

6. FURTHER WORK

A representation of periodic temporal conditions as in LI *et al.* 1992 would help in modelling periodic social events as Olympic games, etc. We plan to develop a tool running on top of CLIO, which evaluates by means of interval arithmetic resulting temporal limits and allows for the automatic detection of contradictory references. For that sake, notions of trust in a reference would be helpful.

MARIA CHRISTOFORAKI, PANOS CONSTANTOPOULOS
MARTIN DOERR
Institute of Computer Science
Foundation of Research and Technology – Hellas

BIBLIOGRAPHY

ALLEN J.F. 1983, *Maintaining knowledge about temporal intervals*, «Communications of the ACM», 26 (11).

BORN M. 1962, *Einstein's Theory of Relativity*, New York, Dover Publications.

CHRISTOFORAKI M., CONSTANTOPOULOS P., DOERR M. 1992, *Clio an Object-Oriented Model of Cultural Data Part I: General Concepts*, Project Report ICS-FORTH.MUIS.92.1, Institute of Computer Science FORTH.

Constantopoulos P., Doerr M. 1993, *The Semantic Index System – A Brief Presentation*, Institute of Computer Science FORTH.

Edelweiss N., Palazzo M.J., de Oliveira Pernici B. 1993, *An object-oriented temporal model*, in C. Rolland, F. Bodart (edd.), *Advanced Information Systems Engineering, 5th International Conference, CaiSE '93*, Paris, France, Springer-Verlag.

Feynman R.F. 1965, *The Feynman Lectures on Physics by Richard P. Feynman Robert B. Leighton and Matthew Sands*, Addison-Wesley Pub. Co, Reading Mass.

Lenat D.B., Guha R.V., Pittman K., Pratt D., Shepherd M. 1990, *Cyc: Toward programms, with common sense*, «Communications of the ACM», 33 (8).

Li R., Carmo J. 1992, *A modal approach to representation of periodical knowledge*, in *ERCIM Workshop on Theoretical and Experimental Aspects of Knowledge representation.*

Mylopoulos J., Borgida A., Jarke M., Koubarakis M. 1990, *Telos: Representing knowledge about information systems*, «ACM Transactions on Information Systems», 8 (4), 352-362.

Tolba H., Charpillet F., Haton J. 1992, *A new temporal combining qualitative and quantitative information*, in *ERCIM Workshop on Theoretical and Experimental Aspects of Knowledge representation.*

ABSTRACT

CLIO, developed by ICS-FORTH, is a system for cultural documentation purposes of museums. It serves as a scientific catalogue of museum artifacts, as opposed to the basic documentation and administrative purposes served by usual collections management systems. It supports artifact descriptions as temporal, geographical, cultural, historical contexts; style, technique, usage, and physical data information. It allows to express certain and uncertain knowledge as well as opinions. In this paper we address the notions of existence, events and causality, referring to them collectively as notions of occurrence, within a conceptual modelling framework and in the context of developing a general ontology for cultural documentation. Particular attention is given to the representation of relations on which historical and other inferences can be based. We present a new approach, which takes mutual dependencies between time and space into account.