The CIDOC CRM, a Standard for the Integration of Cultural Information

Stephen Stead

CIDOC Conceptual Reference Model Special Interest Group

ICS-FORTH, Crete, Greece
November, 2008

The CIDOC CRM

Outline

- Problem statement – information diversity
- Motivation example – the Yalta Conference
- The goal and form of the CIDOC CRM
- Presentation of contents
- About using the CIDOC CRM
- State of development
- Conclusion

The CIDOC CRM
Cultural Diversity and Data Standards

- **Cultural information is more than a domain:**
  - Collection description (art, archeology, natural history….)
  - Archives and literature (records, treaties, letters, artful works..)
  - Administration, preservation, conservation of material heritage
  - Science and scholarship – investigation, interpretation
  - Presentation – exhibition making, teaching, publication
- **But how to make a documentation standard?**
  - Each aspect needs its methods, forms, communication means
  - Data overlap, but do not fit in one schema
  - Understanding lives from relationships, but how to express them?

The CIDOC CRM

Historical Archives….

Type:	Text
Title:	Protocol of Proceedings of Crimea Conference
Title.Subtitle:	II. Declaration of Liberated Europe
Date:	February 11, 1945
Creator:	The Premier of the Union of Soviet Socialist Republics
	The Prime Minister of the United Kingdom
	The President of the United States of America
Publisher:	State Department
Subject:	Postwar division of Europe and Japan

Metadata

About…

Documents

"The following declaration has been approved:
The Premier of the Union of Soviet Socialist Republics,
the Prime Minister of the United Kingdom and the President
of the United States of America have consulted with each
other in the common interests of the people of their countries
and those of liberated Europe. They jointly declare their mutual
agreement to concert…
….and to ensure that Germany will never again be able to
disturb the peace of the world…… "

The CIDOC CRM

Images, non-verbose…

Type:	Image
Title:	Allied Leaders at Yalta
Date:	1945
Publisher:	United Press International (UPI)
Source:	The Bettmann Archive
Copyright:	Corbis
References:	Churchill, Roosevelt, Stalin

Metadata

About…

Photos, Persons

The CIDOC CRM
Places and Objects

TGN Id: 7012124
Names: Yalta (C,V), Jalta (C,V)
Types: inhabited place(C), city (C)
Position: Lat: 44 30 N,Long: 034 10 E
Hierarchy: Europe (continent) <– Ukrayina (nation) <– Krym (autonomous republic)
Note: …Site of conference between Allied powers in WW II in 1945; ….
Source: TGN, Thesaurus of Geographic Names

Places, Objects

About…

Title: Yalta, Crimean Peninsula
Publisher: Kurgan-Lisnet
Source: Liaison Agency

The CIDOC CRM

The Integration Problem (1)

- **Problem 1- Identity:**
  - **Actors, Roles, proper names:**
    - The Premier of the Union of Soviet Socialist Republics
      Allied leader, Allied power
      Joseph Stalin….
  - **Places**
    - Jalta, Yalta
    - Krym, Crimea
  - **Events**
    - Crimea Conference, “Allied Leaders at Yalta”,“… conference between Allied powers” “Postwar division”
  - **Objects and Documents:**
    - The photo, the agreement text

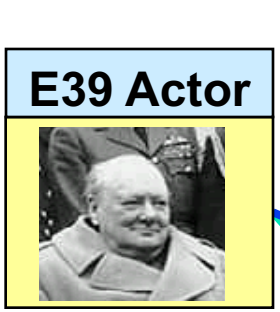
The CIDOC CRM

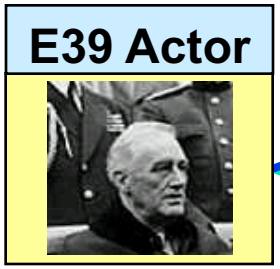
The Integration Problem (2)

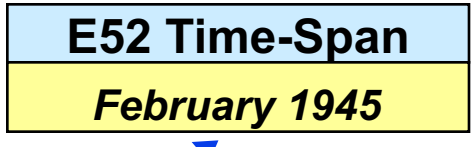
- **Problem 2- hidden entities (typically “title”):**
  - **Actors**
    - Allied leader, Allied power
  - **Places**
    - Yalta, Crimea
  - **Events**
    - Crimea Conference, “Allied Leaders at Yalta”,“… conference between Allied powers” “Postwar division”
- **Solution:**
  - **Change metadata structures: but what are the relevant elements?**

The CIDOC CRM

Explicit Events, Object Identity, Symmetry



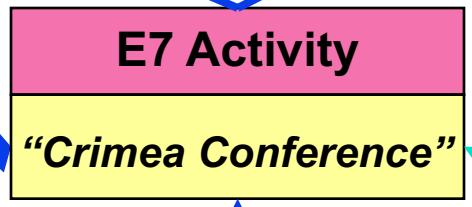




P82 at some time within

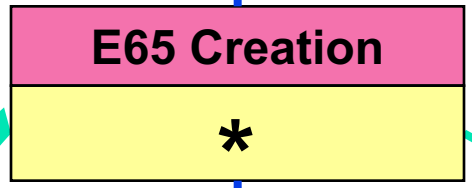
P7 took place at

P11 participated in



P67 is referred to by

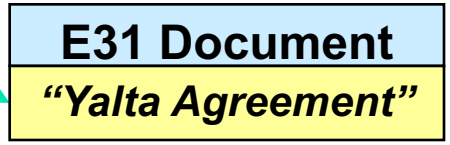
P86 falls within

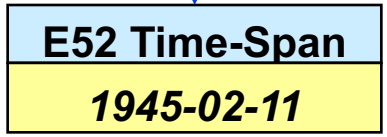


P14 performed

P81 ongoing throughout

P94 has created





The CIDOC CRM…

- …captures the underlying semantics of relevant documentation structures in a formal ontology
- Ontologies are formalized knowledge: clearly defined concepts and relationships about possible states of affairs in a domain
- They can be understood by people and processed by machines to enable data exchange, data integration, query mediation etc.
- Semantic interoperability in cultural heritage can be achieved with an "extensible ontology of relationships" and explicit event modeling
- This provides shared explanation rather than the prescription of a common data structure
- The ontology is the language that S/W developers and museum experts can share. Therefore it needed interdisciplinary work. That is what CIDOC has provided

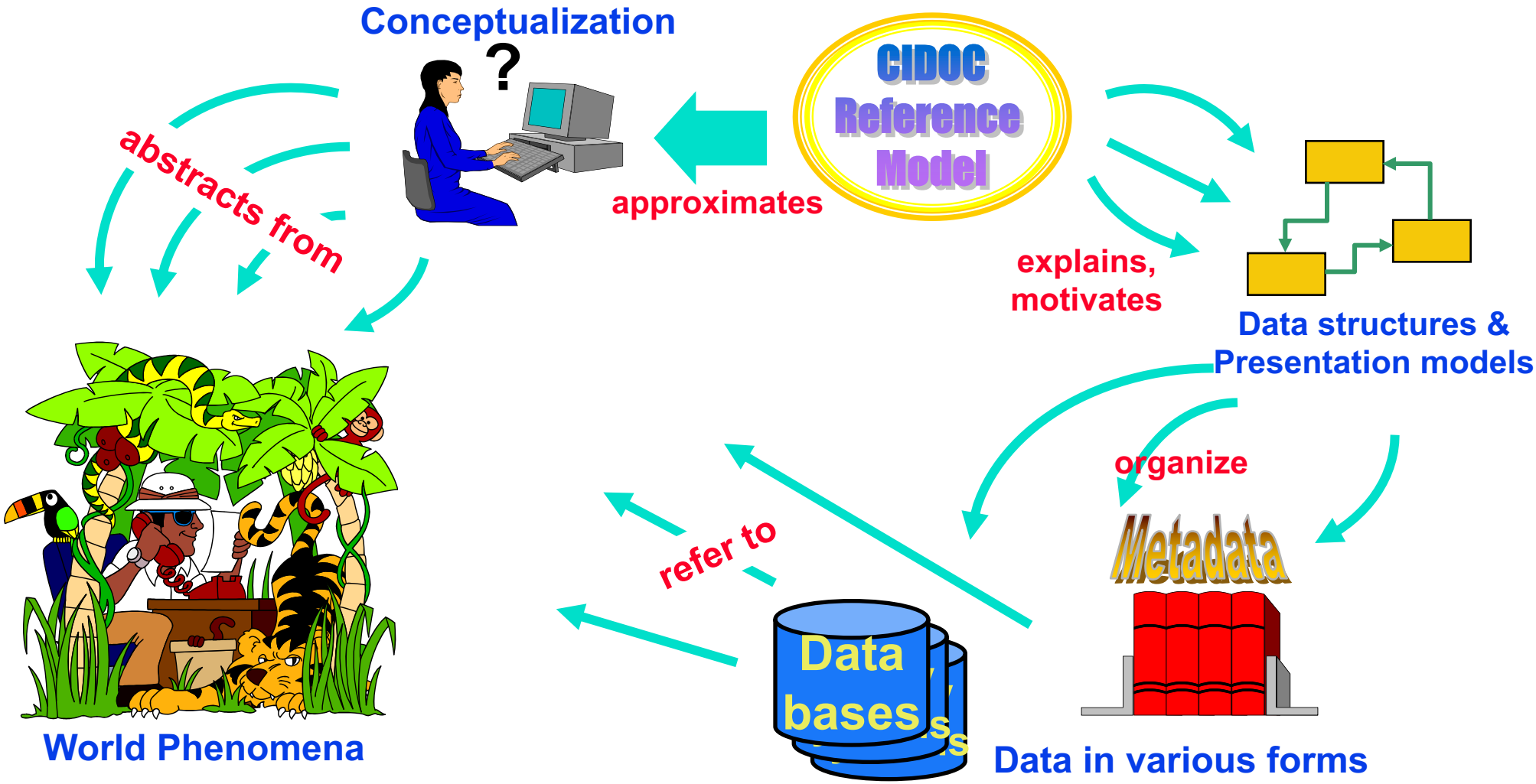
The CIDOC CRM
Outcomes

- The CIDOC Conceptual Reference Model
  - A collaboration with the International Council of Museums
  - An ontology of 86 classes and 137 properties for culture and more
  - With the capacity to explain hundreds of (meta)data formats
  - Accepted by ISO TC46 in September 2000
  - International standard since 2006 - ISO 21127:2006
- Serving as:
  - intellectual guide to create schemata, formats, profiles
  - A language for analysis of existing sources for integration/mediation

    "Identify elements with common meaning"
  - Transportation format for data integration / migration / Internet

The CIDOC CRM

The Intellectual Role of the CRM



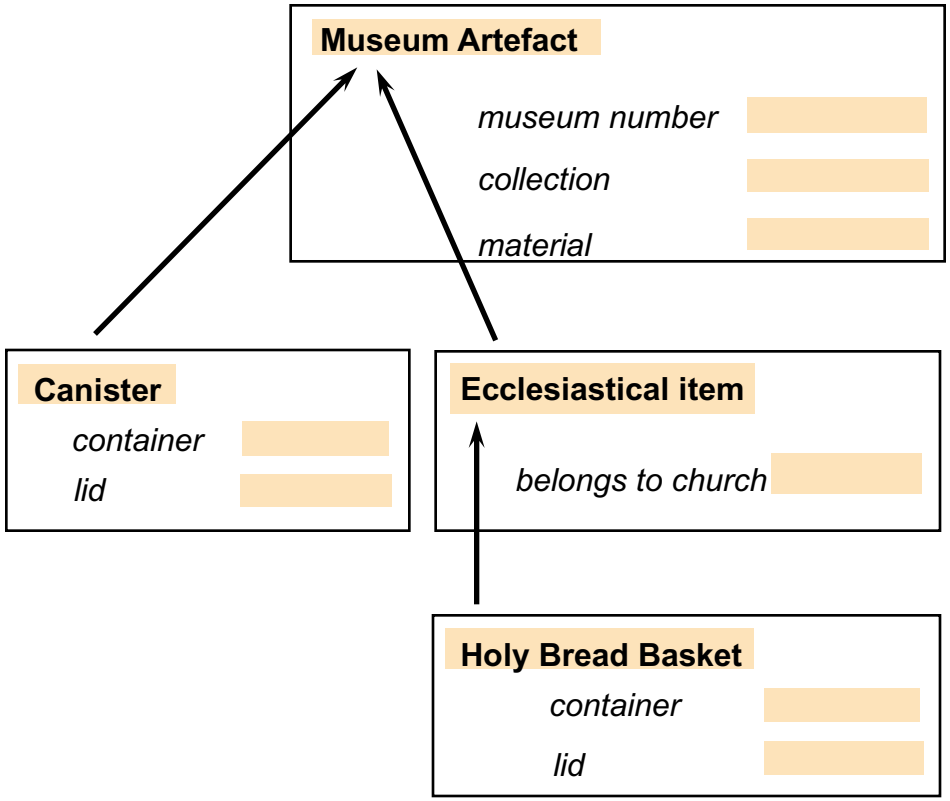
The CIDOC CRM
Encoding of the CIDOC CRM

- The CIDOC CRM is a formal ontology (defined in TELOS)
  - But CRM instances can be encoded in many forms: RDBMS, ooDBMS, XML, RDF(S)
  - Uses Multiple isA – to achieve uniqueness of properties in the schema
  - Uses multiple instantiation – to be able to combine not always valid combinations (e.g. destruction – activity)
  - Uses Multiple isA for properties to capture different abstraction of relationships
- Methodological aspects:
  - Entities are introduced as anchors of properties (and if structurally relevant)
  - Frequent joins (short-cuts) of complex data paths for data found in different degrees of detail are modeled explicitly

The CIDOC CRM

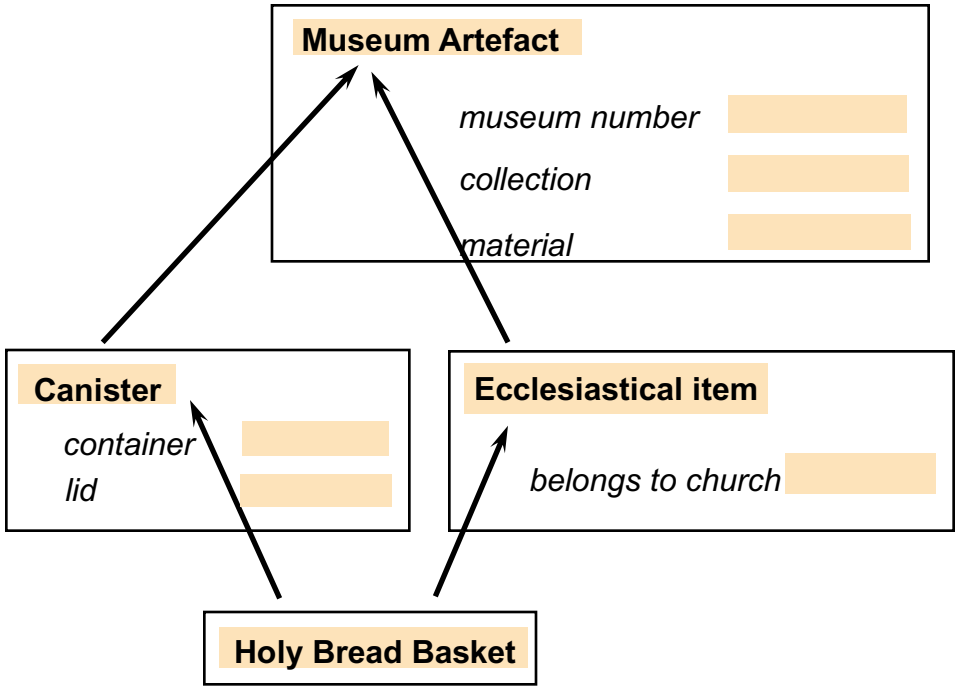
Justifying Multiple Inheritance

Single Inheritance form:



Repetition of properties

Multiple Inheritance form:



Unique identity of properties

The CIDOC CRM

Data example (e.g. from extraction)

Epitaphios GE34604 (entity E22 Man-Made Object)

P30 custody transferred through, P24 changed ownership through

Transfer of Epitaphios GE34604 (entity E10 Transfer of Custody, E8 Acquisition Event

Multiple Instantiation

P28 custody surrendered by

Metropolitan Church of the Greek Community of Ankara (entity E39 Actor)

P23 transferred title from

Metropolitan Church of the Greek Community of Ankara (entity E39 Actor)

P29 custody received by

Museum Benaki (entity E39 Actor)

P22 transferred title to

Exchangeable Fund of Refugees (entity P40 Legal Body)

P2 has type

national foundation (entity E55 Type)

P14 carried out by

Exchangeable Fund of Refugees (entity E39 Actor)

P4 has time-span

GE34604_transfer_time (entity E52 Time-Span)

P82 at some time within

1923 - 1928 (entity E61 Time Primitive)

P7 took place at

Greece (entity E53 Place)

P2 has type

nation (entity E55 Type)

republic (entity E55 Type)

P89 falls within

Europe (entity E53 Place)

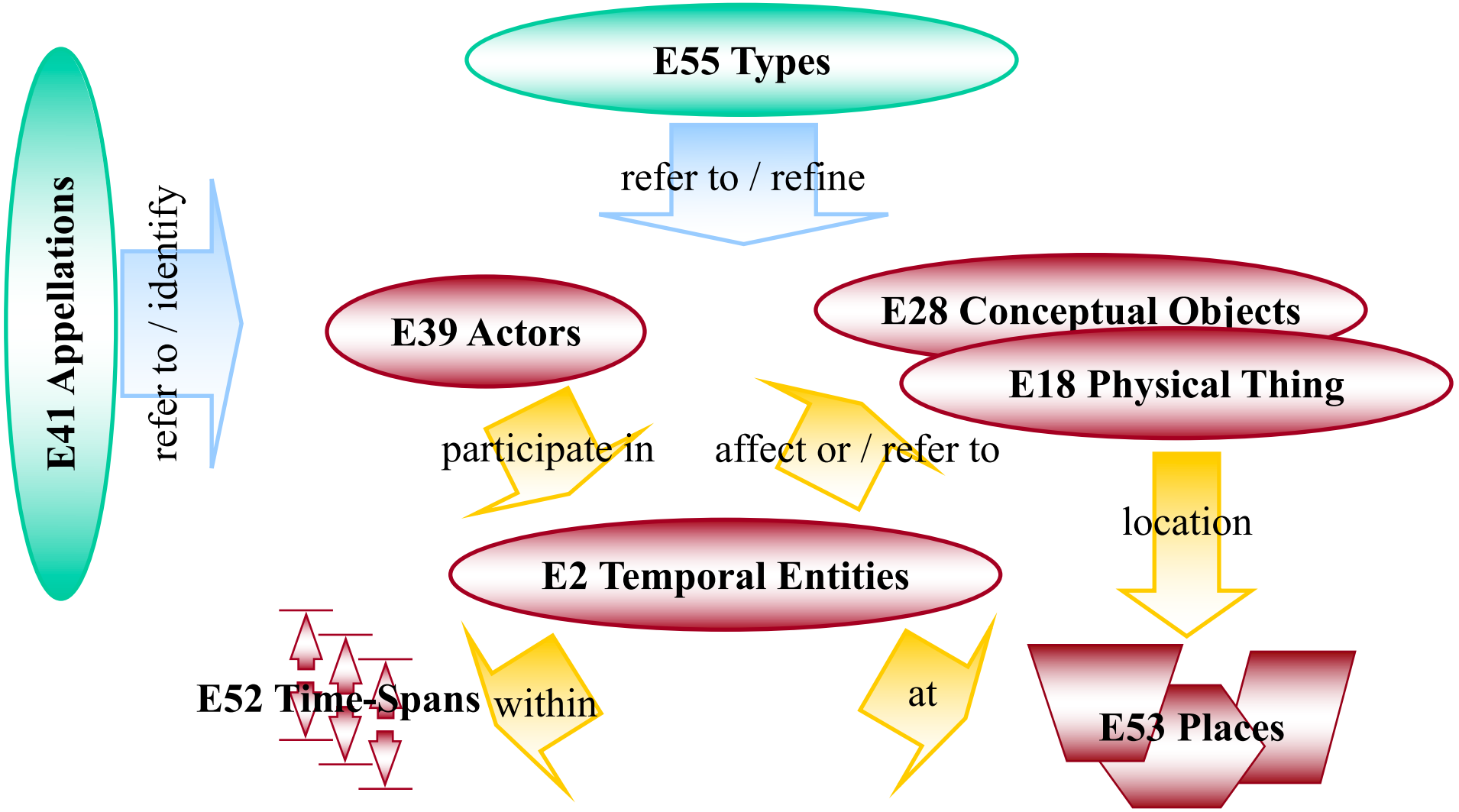
P2 has type

continent (entity E55 Type)

TGN data

The CIDOC CRM

Top-level classes useful for integration

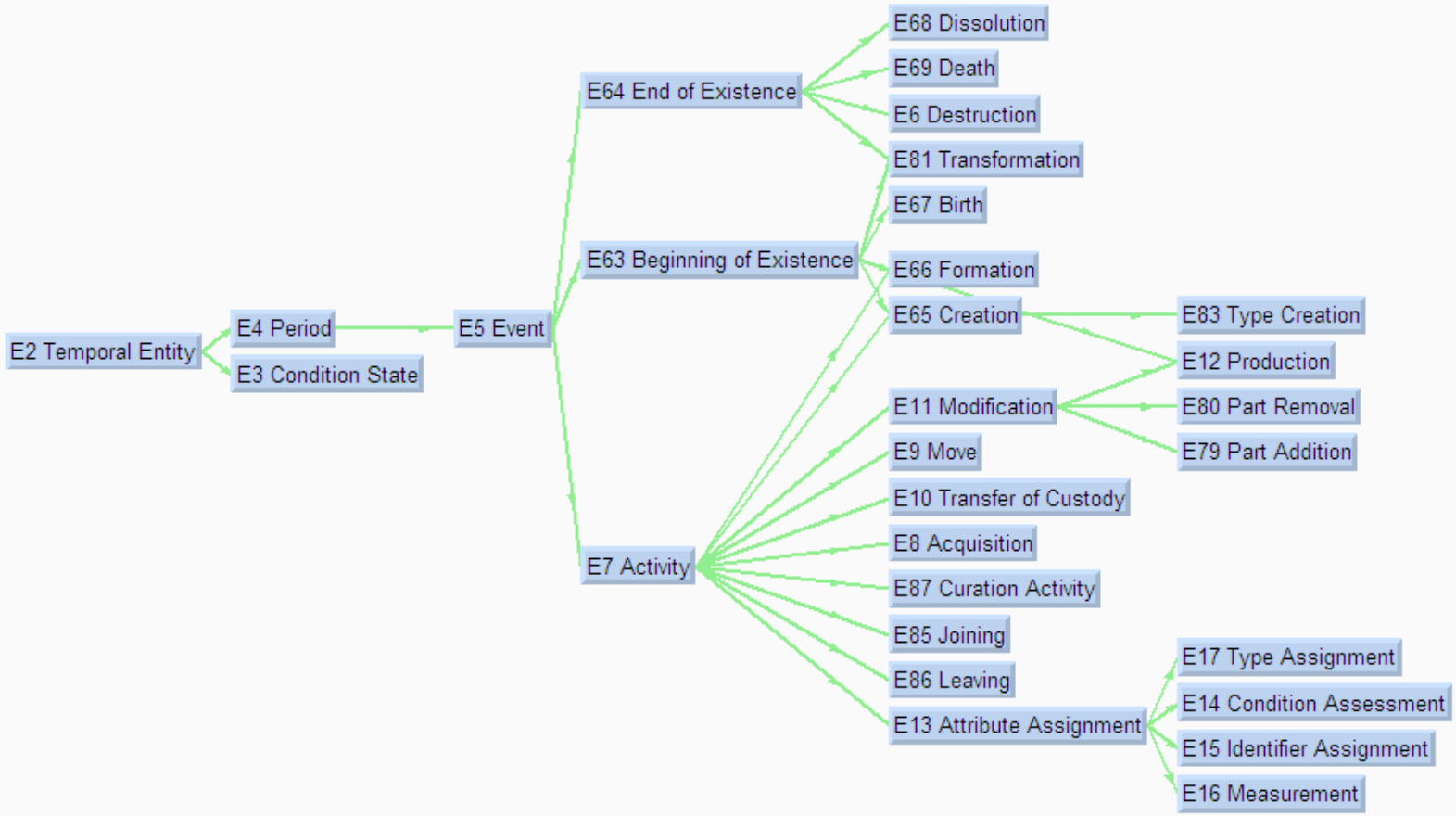


The CIDOC CRM
The types of relationships

- **Identification of real world items by real world names**
- **Observation and Classification of real world items**
- **Part-decomposition and structural properties of Conceptual & Physical Objects, Periods, Actors, Places and Times**
- **Participation of persistent items in temporal entities**
  - creates a notion of history: “world-lines” meeting in space-time
- **Location of periods in space-time and physical objects in space**
- **Influence of objects on activities and products and vice-versa**
- **Reference of information objects to any real-world item**

The CIDOC CRM

The E2 Temporal Entity Hierarchy



The CIDOC CRM

Scope note example: E2 Temporal Entity

E2 Temporal Entity

- **Scope Note:**

  This class comprises all phenomena, such as the instances of E4 Periods, E5 Events and states, which happen over a limited extent in time.

  In some contexts, these are also called perdurants. This class is disjoint from E77 Persistent Item. This is an abstract class and has no direct instances. E2 Temporal Entity is specialized into E4 Period, which applies to a particular geographic area (defined with a greater or lesser degree of precision), and E3 Condition State, which applies to instances of E18 Physical Thing.

  - Is limited in time, is the only link to time, but is not time itself
  - spreads out over a place or object
  - the core of a model of physical history, open for unlimited specialisation

The CIDOC CRM

Temporal Entity- Subclasses

- **E4 Period**
  - binds together related phenomena
  - introduces inclusion topologies - parts etc.
  - Is confined in space and time
  - the basic unit for temporal-spatial reasoning
- **E5 Event**
  - looks at the input and the outcome
  - introduces participation of people and presence of things
  - the basic unit for weak causal reasoning
  - each event is a period if we study the process
- **E7 Activity**
  - adds intention, influence and purpose
  - adds tools

The CIDOC CRM

Temporal Entity- Main Properties

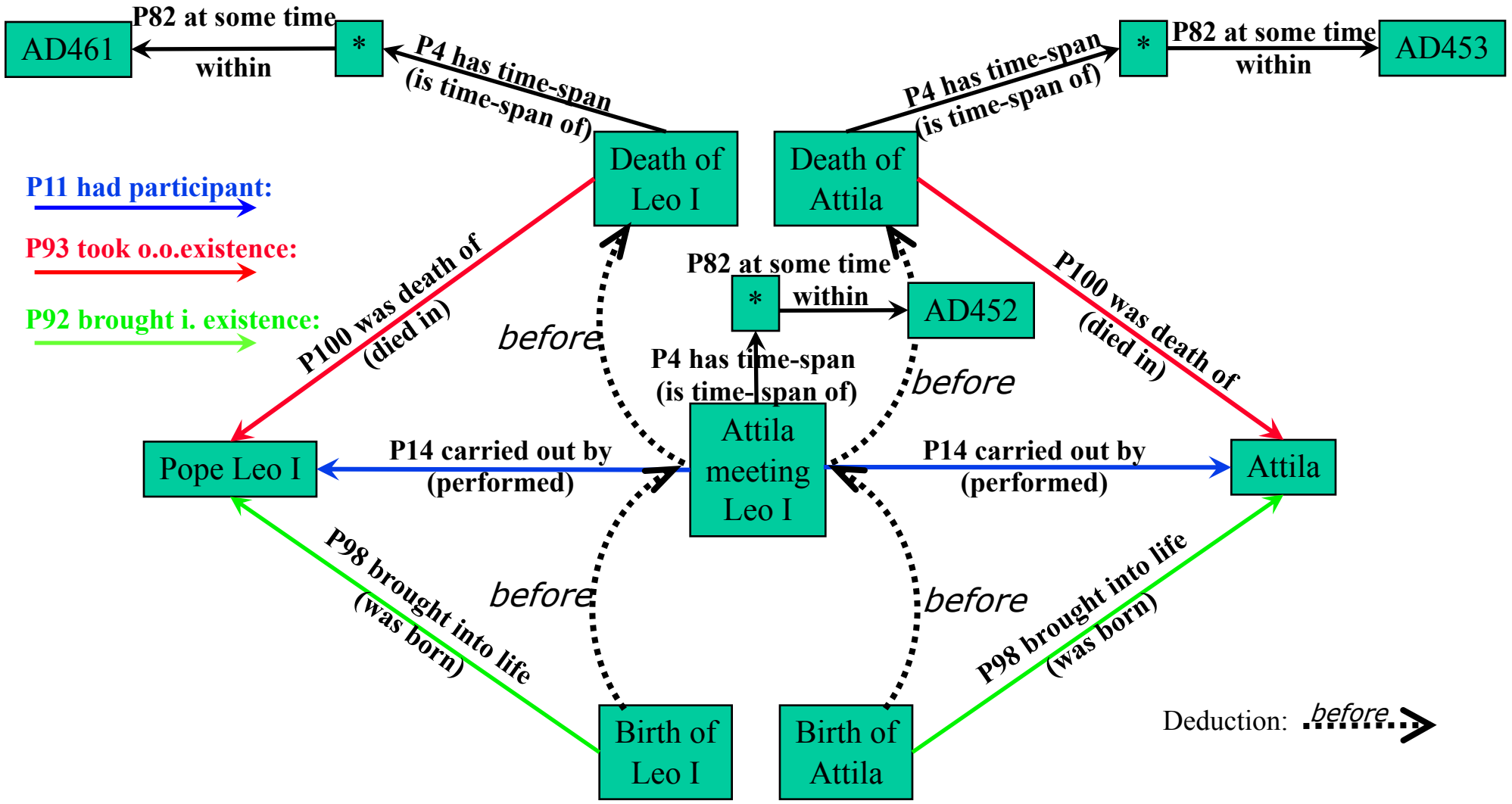
- **E2 Temporal Entity**
  - Properties: P4 has time-span (is time-span of): E52 Time-Span
- **E4 Period**
  - Properties: P7 took place at (witnessed): E53 Place
    P9 consists of (forms part of): E4 Period
    P10 falls within (contains): E4 Period
- **E5 Event**
  - Properties: P11 had participant (participated in): E39 Actor
    P12 occurred in the presence of (was present at): E77 Persistent Item
- **E7 Activity**
  - Properties: P14 carried out by (performed): E39 Actor
    P20 had specific purpose (was purpose of): E5 Event
    P21 had general purpose (was purpose of): E55 Type

The CIDOC CRM

The Participation Properties

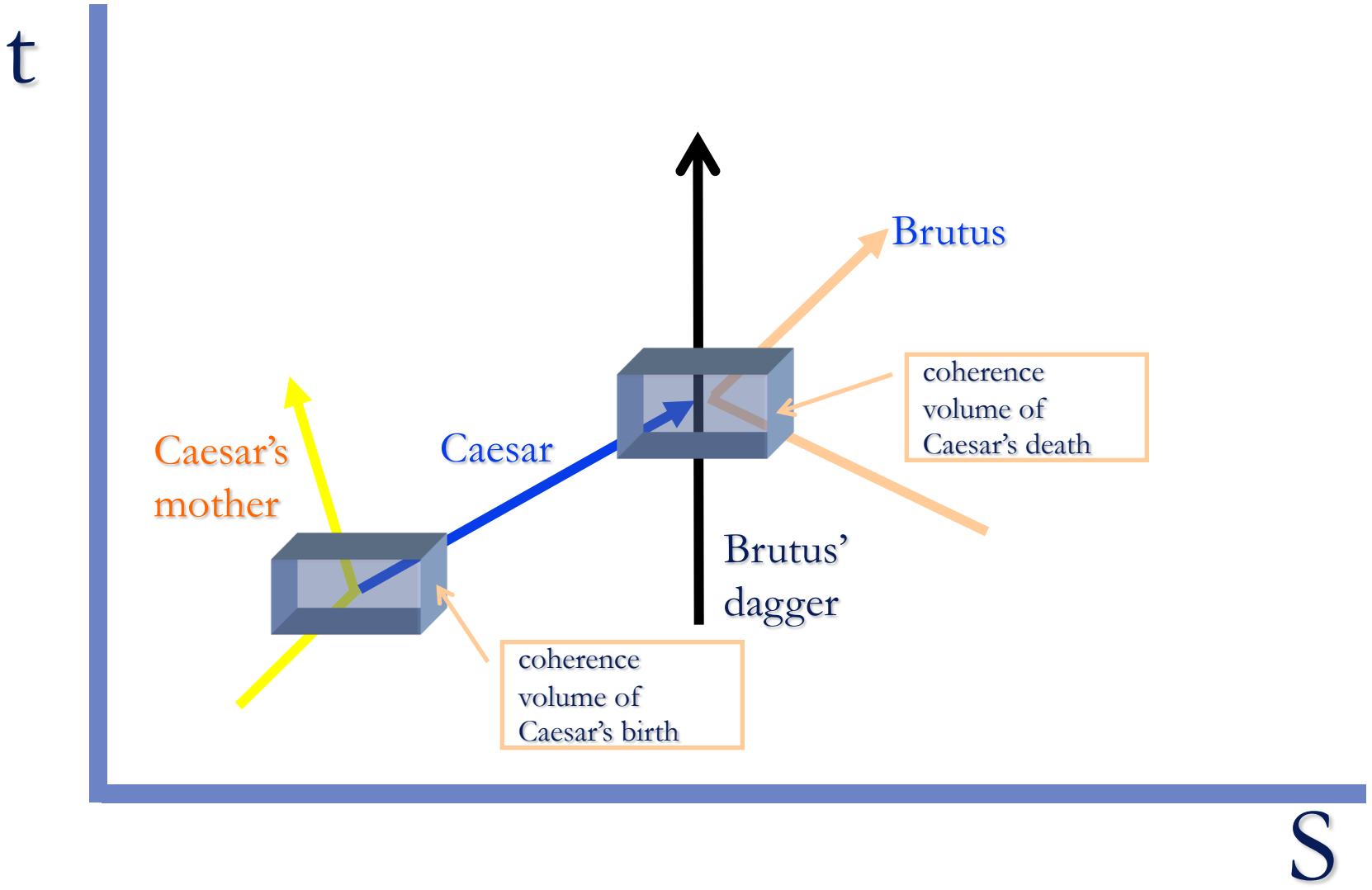
The CIDOC CRM

Termini postquem / antequem



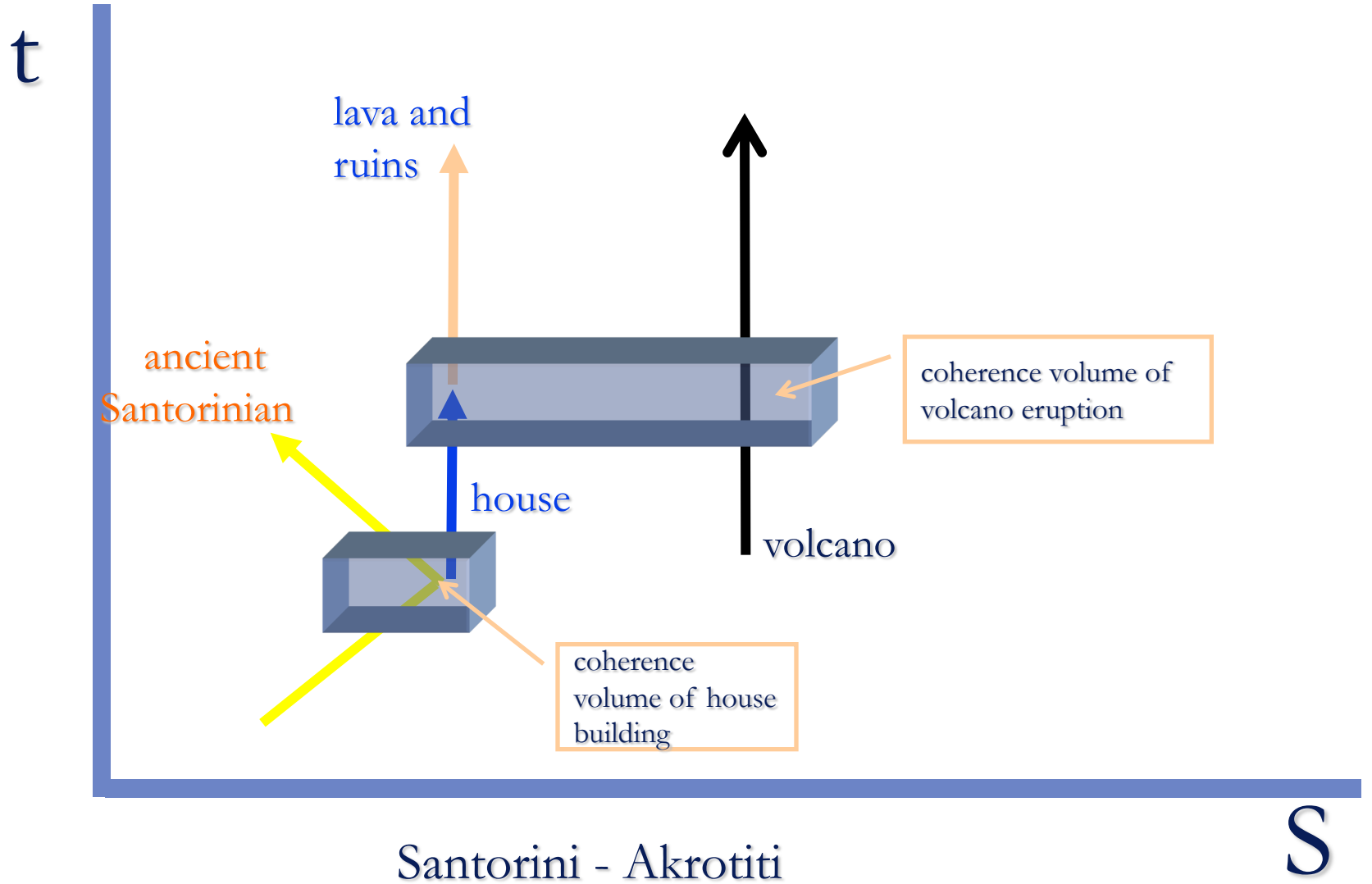
The CIDOC CRM

Historical events as meetings



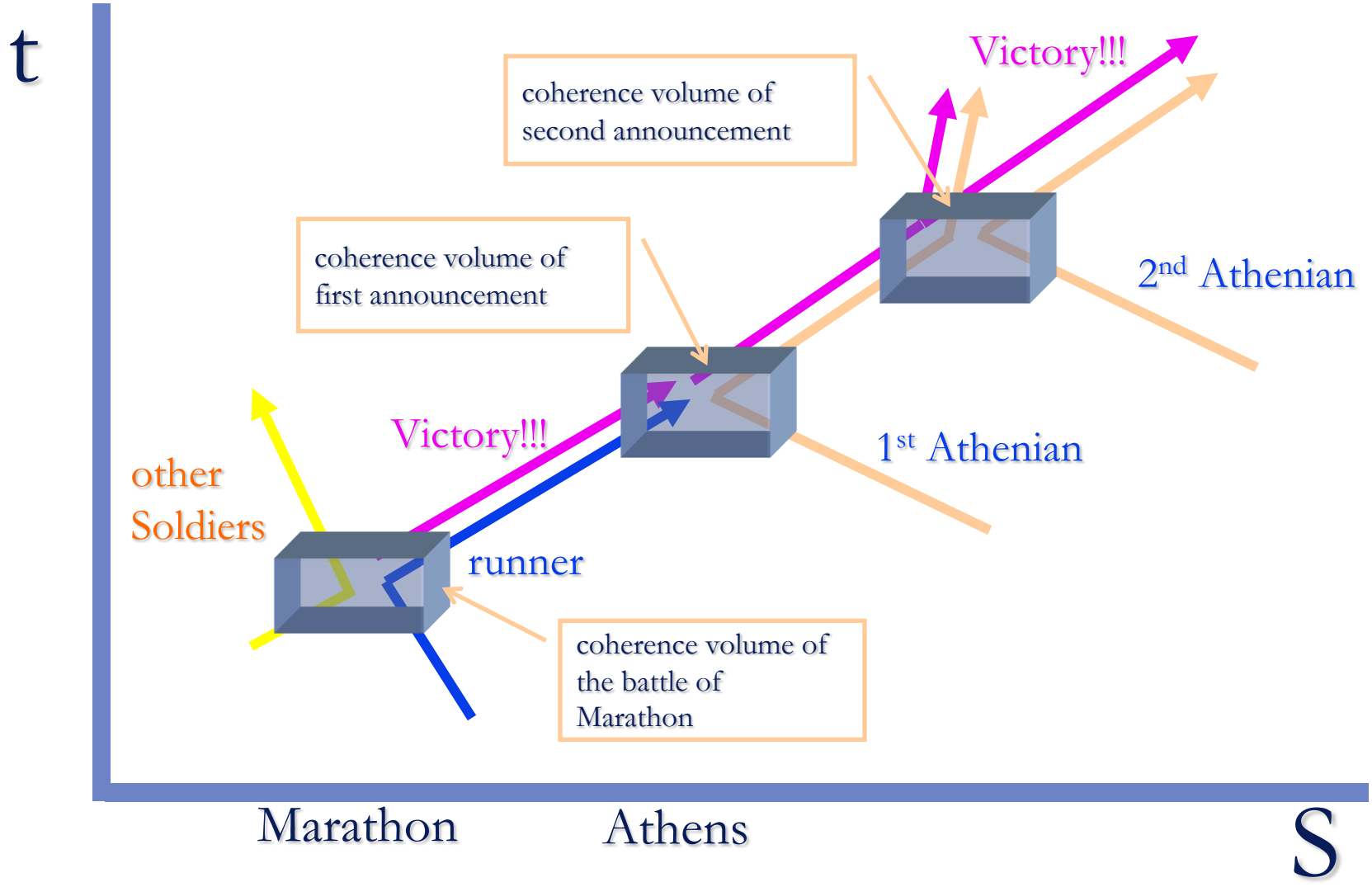
The CIDOC CRM

Depositional events as meetings



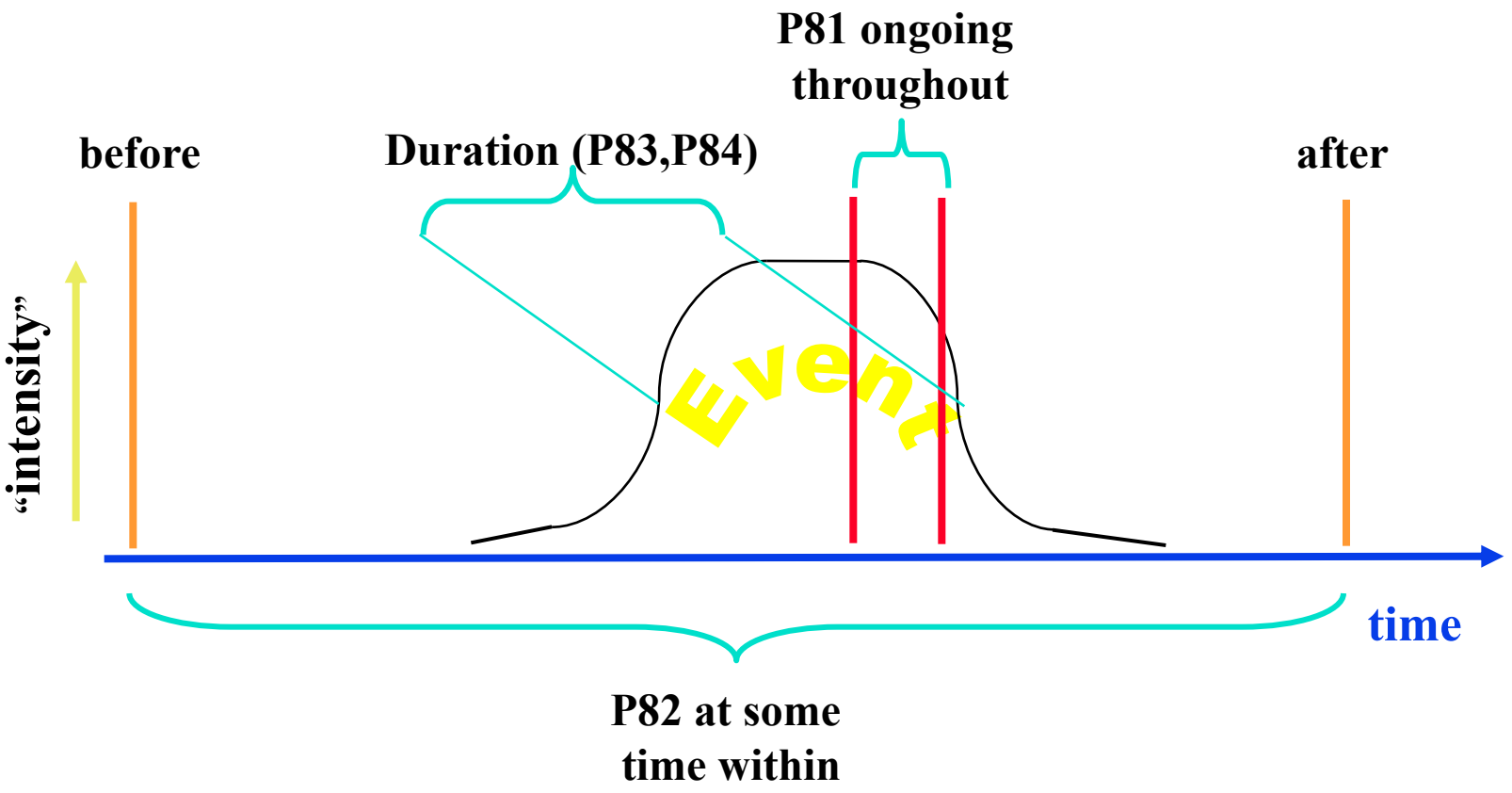
The CIDOC CRM

Exchanges of information as meetings



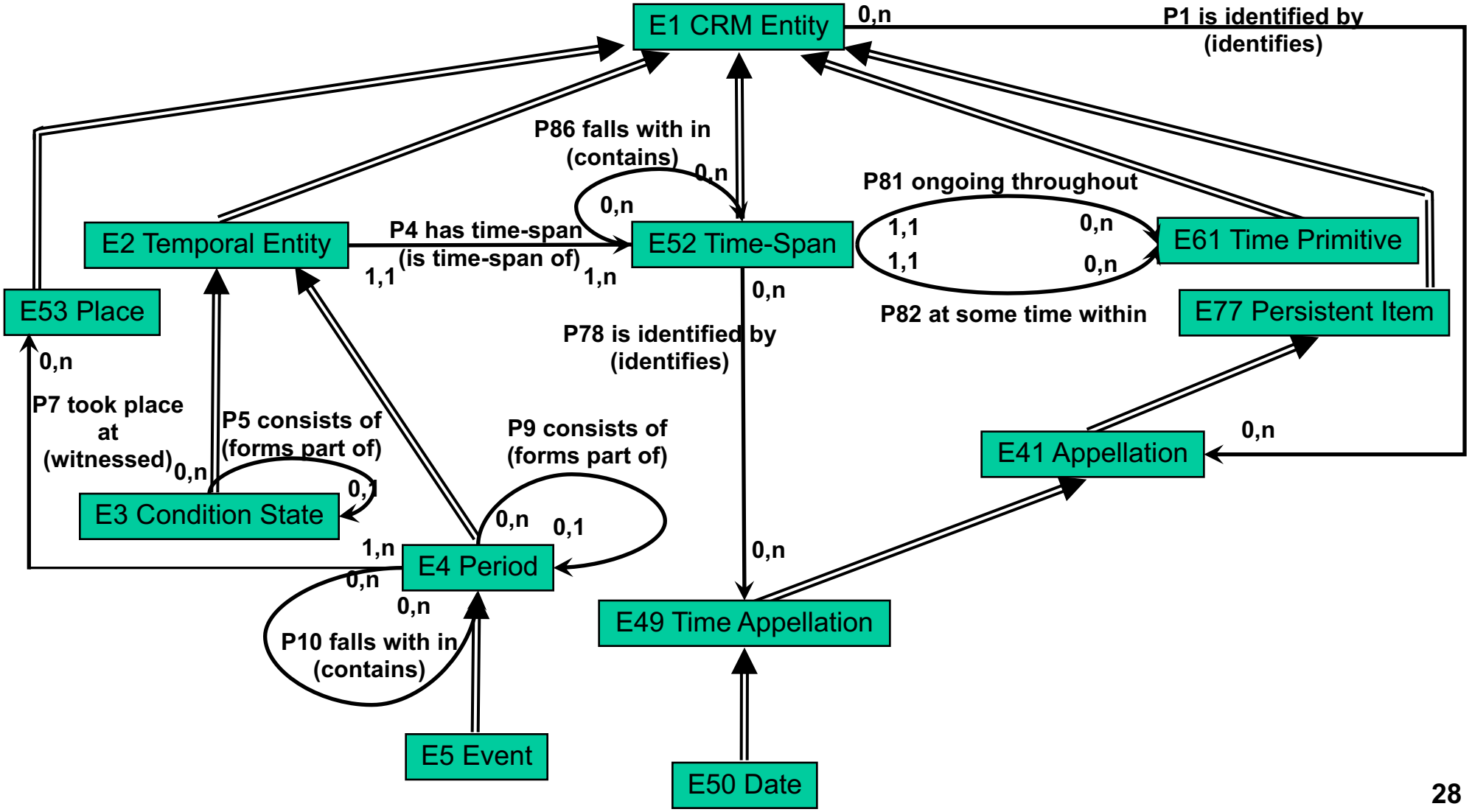
The CIDOC CRM

Time Uncertainty, Certainty and Duration



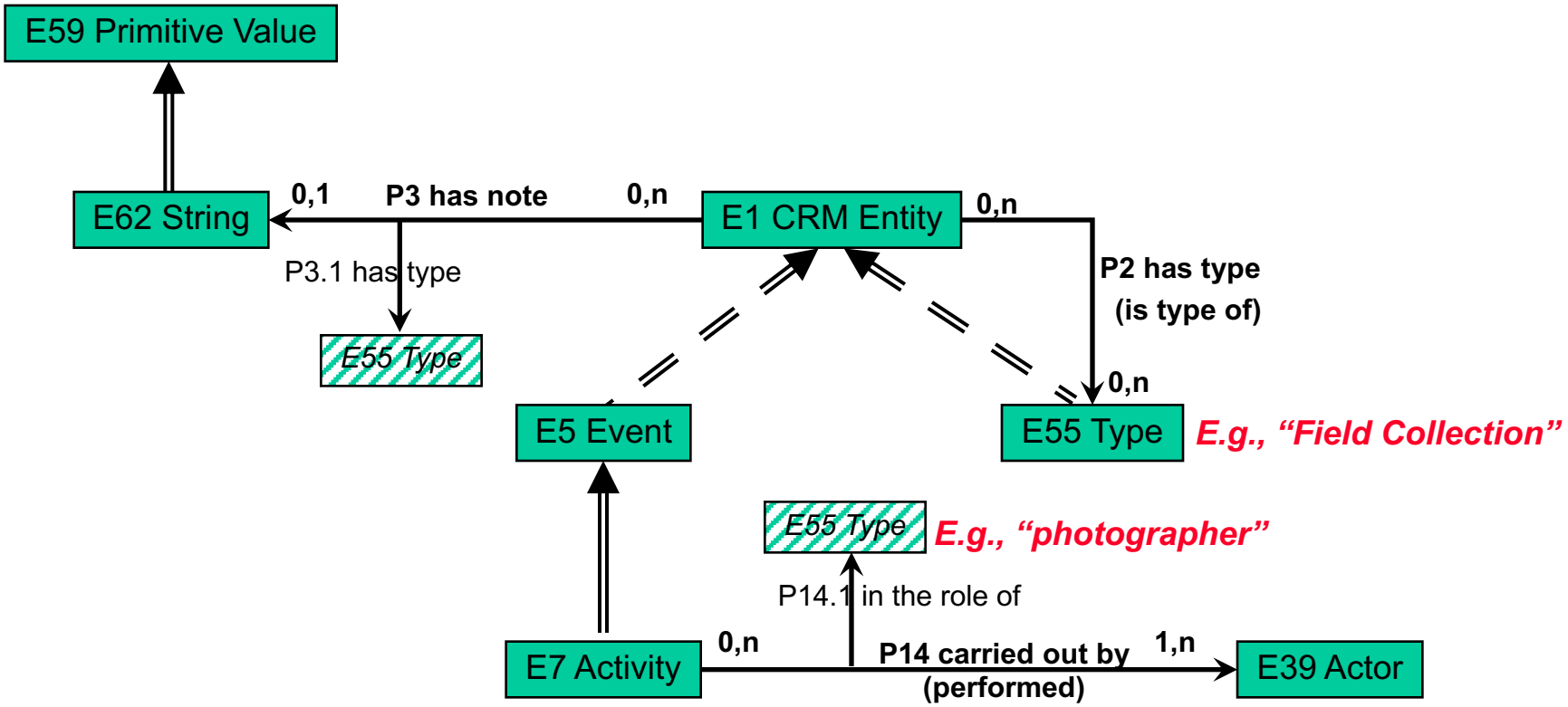
The CIDOC CRM

E52 Time-Span



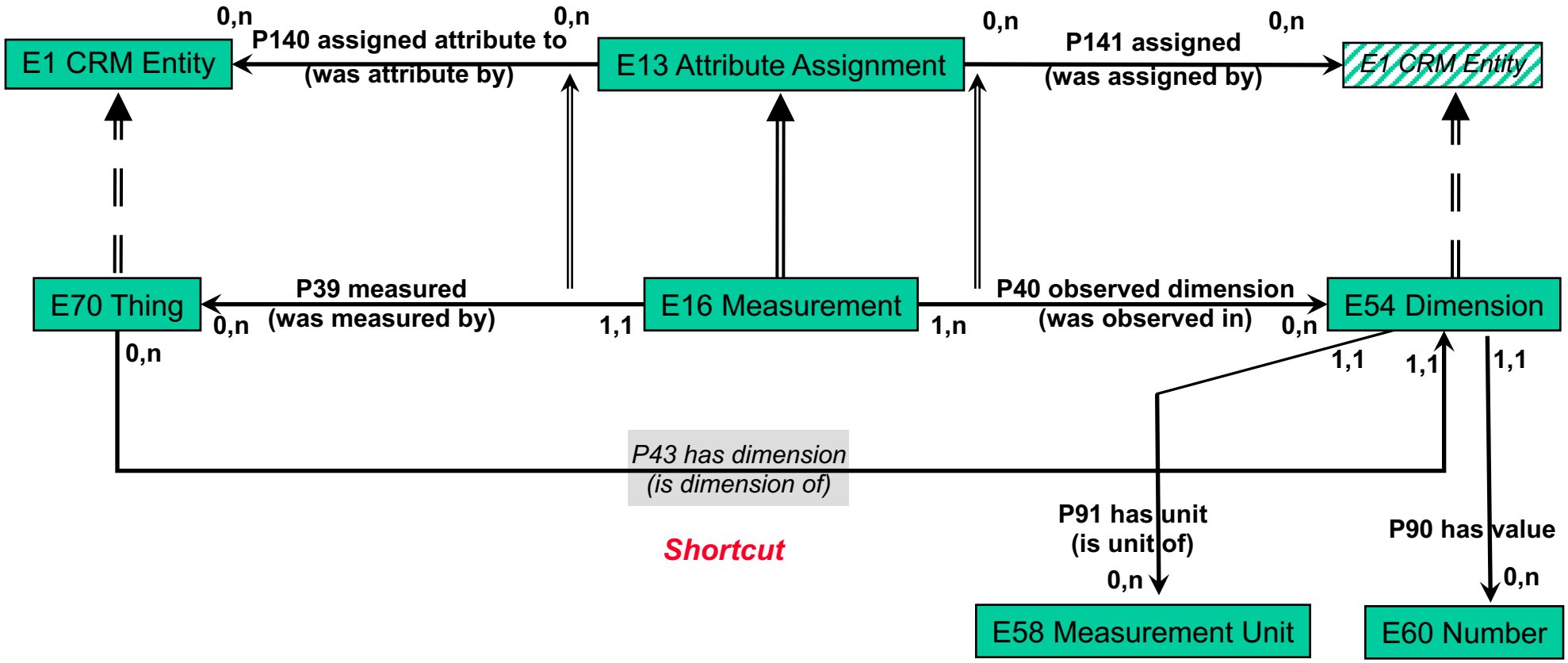
The CIDOC CRM

E7 Activity and inherited properties



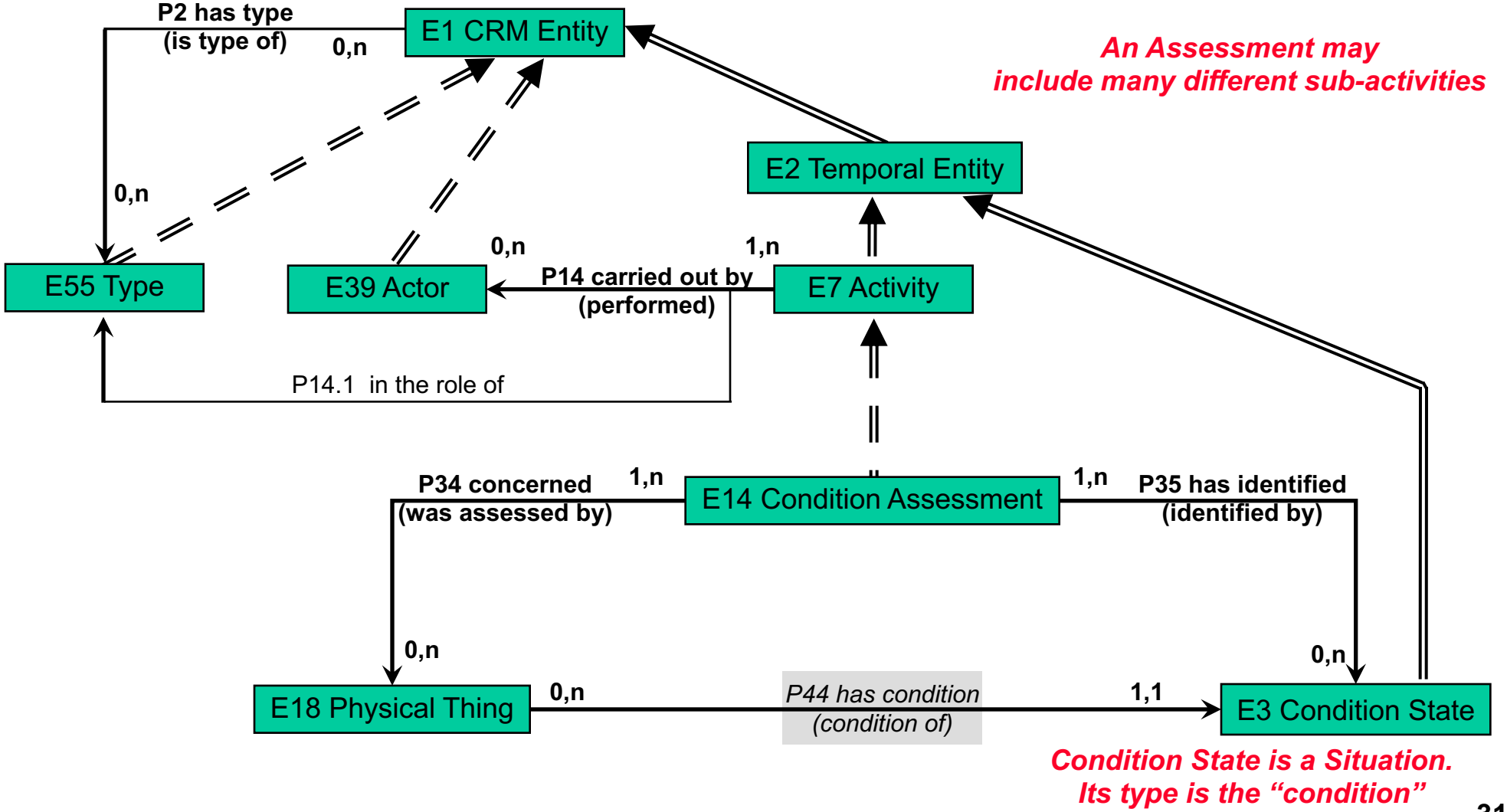
The CIDOC CRM

Activities: E16 Measurement



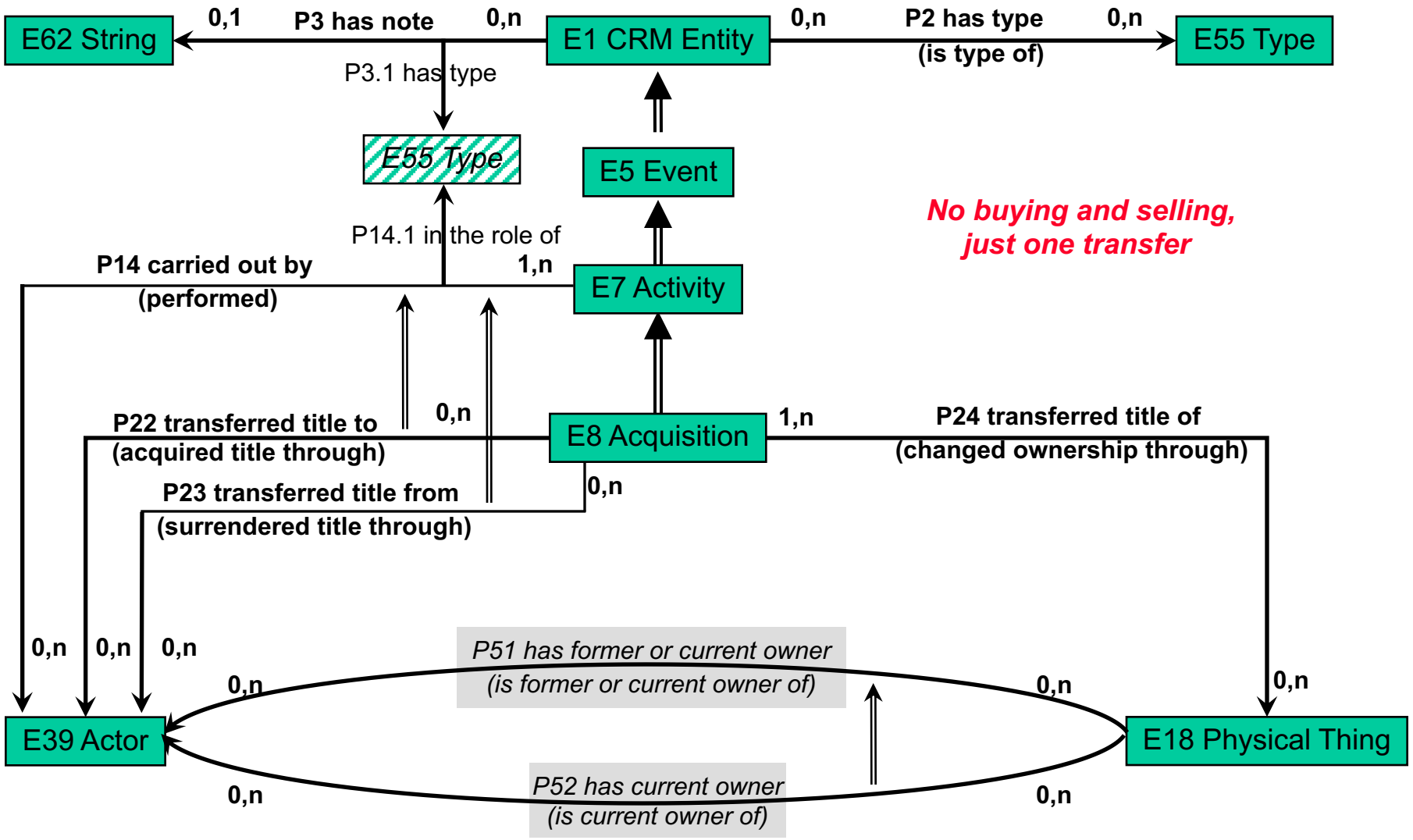
The CIDOC CRM

Activities: E14 Condition Assessment



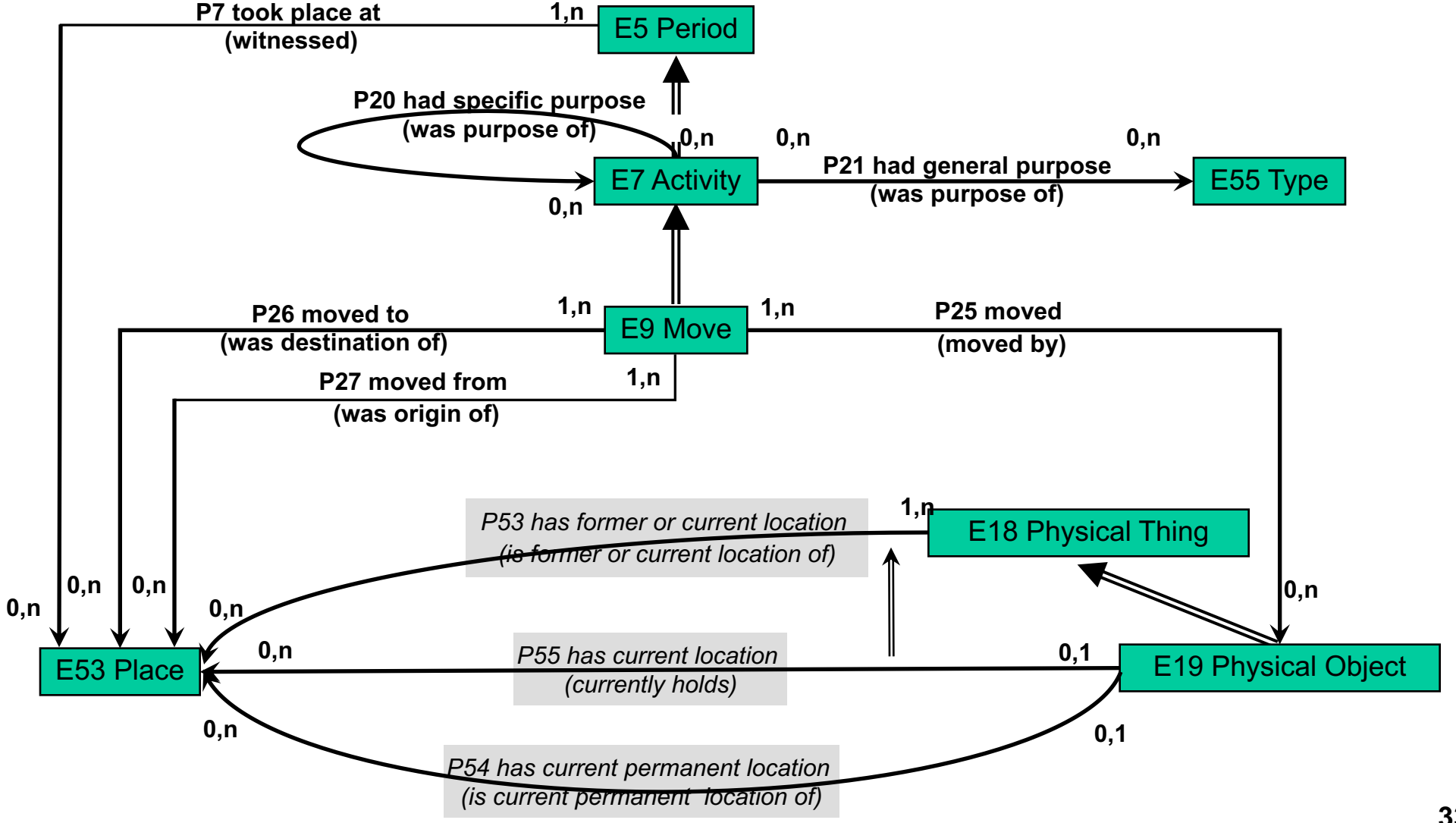
The CIDOC CRM

Activities: E8 Acquisition



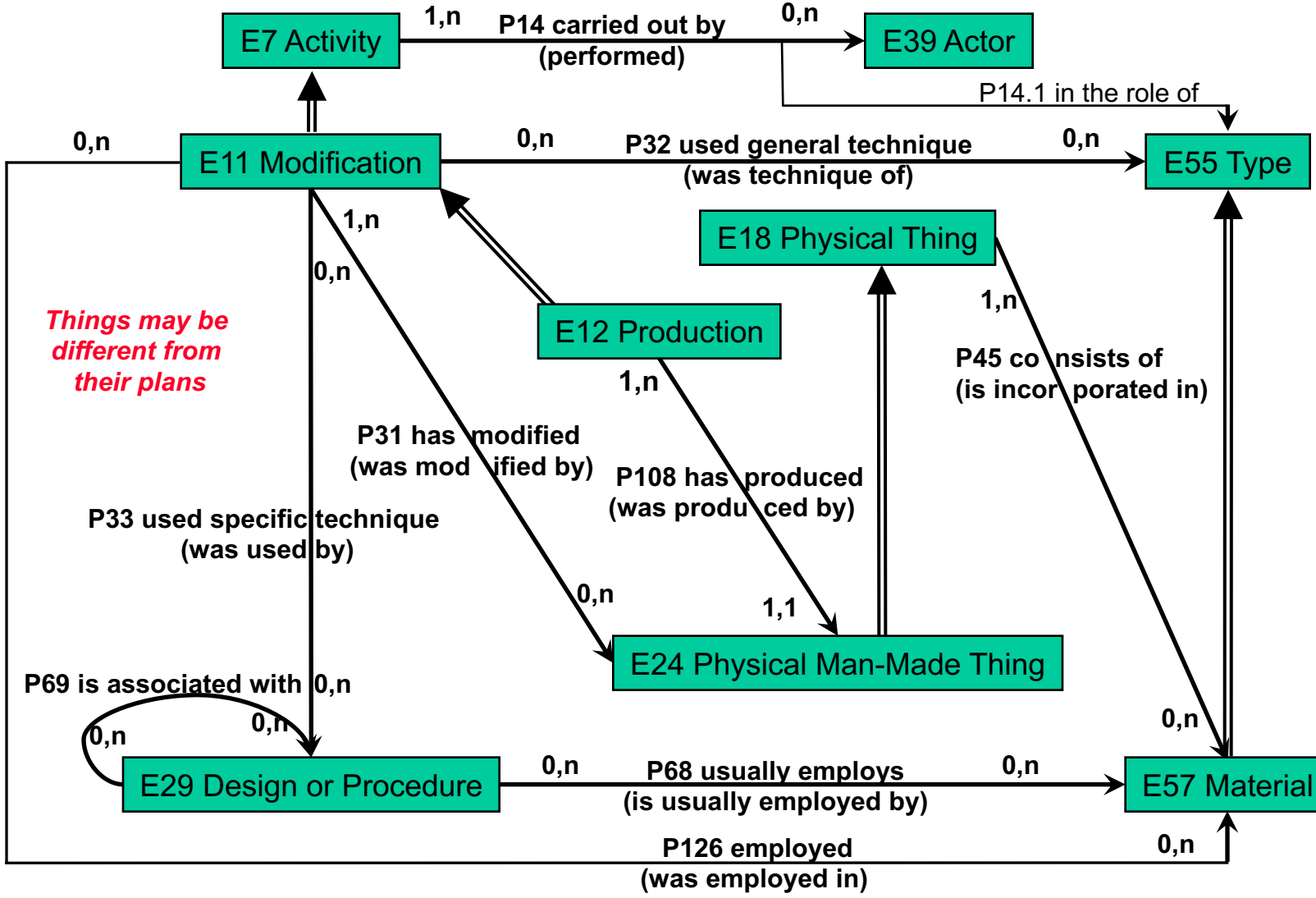
The CIDOC CRM

Activities: E9 Move



The CIDOC CRM

Activities: E11 Modification/ E12 Production



The CIDOC CRM

Inheriting Properties: E11 Modification

Properties:

P1 is identified by (identifies): E41 Appellation
P2 has type (is type of): E55 Type
P11 had participant (participated in): E39 Actor
P14 carried out by (performed): E39 Actor
(P14.1 in the role of : E55 Type)

inherited properties

P31 has modified (was modified by): E24 Physical Man-Made Thing

declared property

P12 occurred in the presence of (was present at): E77 Persistent Item
P16 used specific object (was used for): E70 Thing
(P16.1 mode of use: E55 Type)

inherited properties

***P32* used general technique (was technique of): E55 Type**
P33 used specific technique (was used by): E29 Design or Procedure

declared properties

P17 was motivated by (motivated): E1 CRM Entity
P19 was intended use of (was made for): E71 Man-Made Thing
(P19.1 mode of use: E55 Type)
P20 had specific purpose (was purpose of): E5 Event
P21 had general purpose (was purpose of): E55 Type

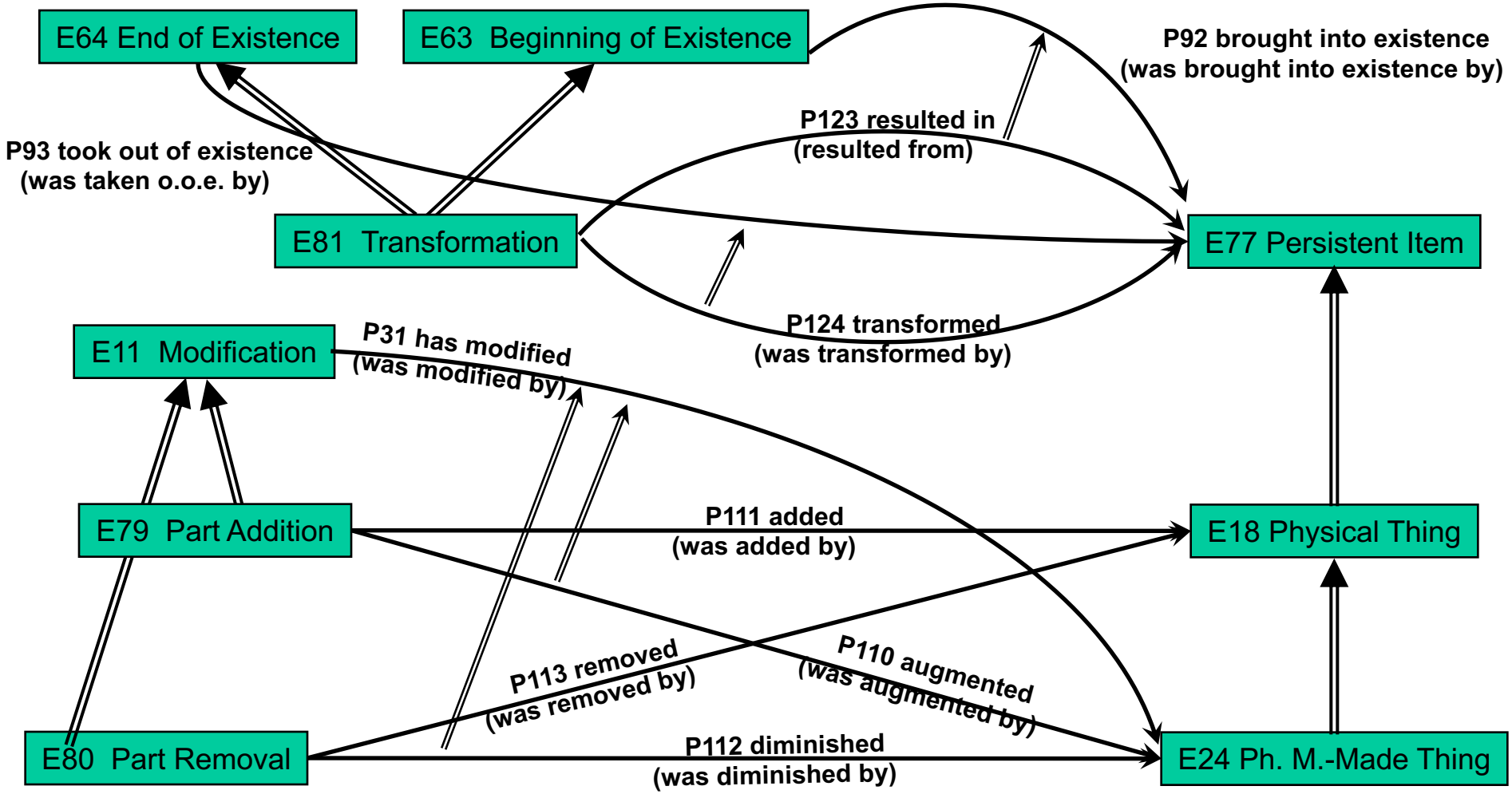
inherited properties

P126 employed (was employed in): E57 Material

declared property

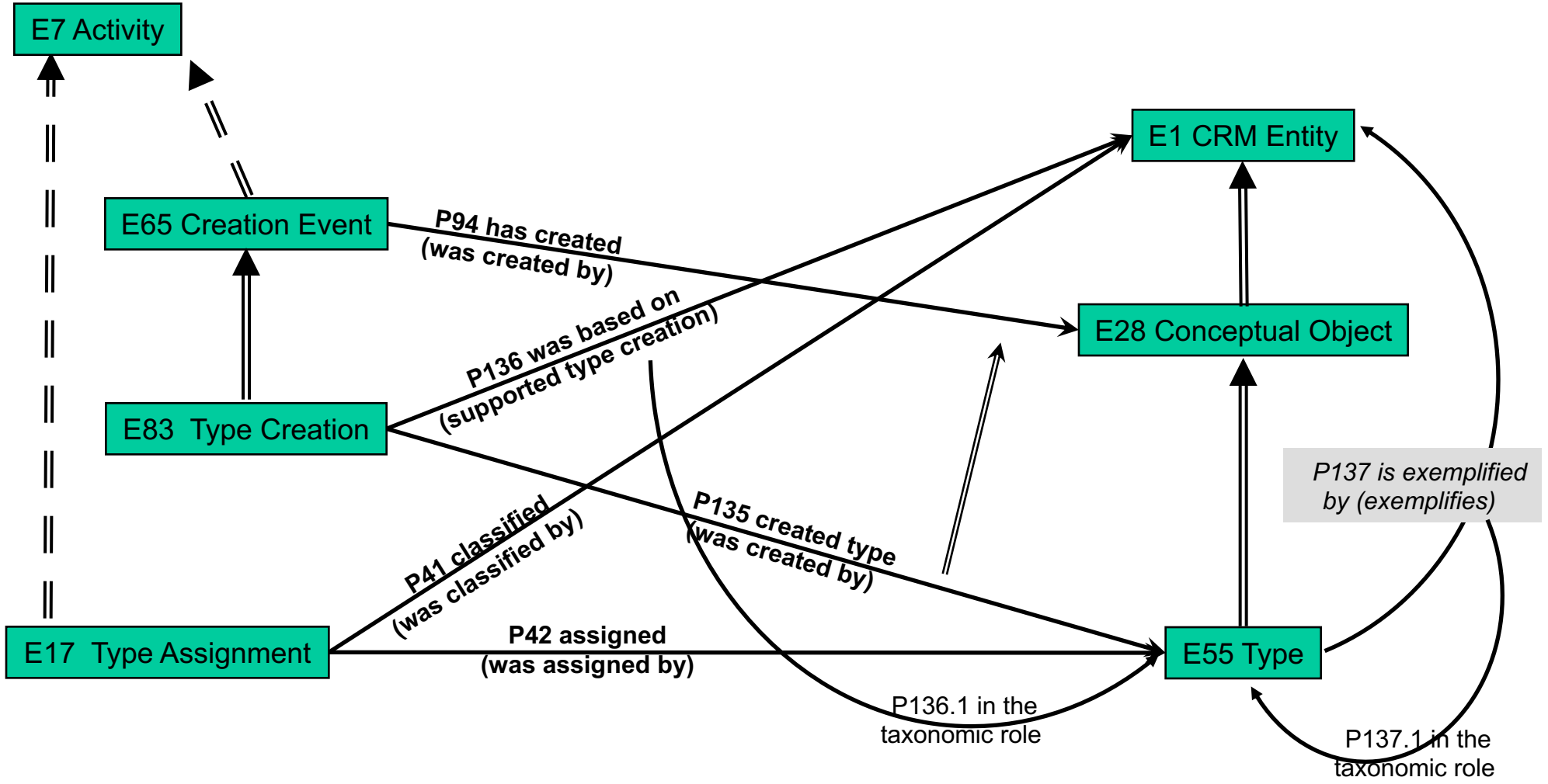
The CIDOC CRM

Ways of Changing Things



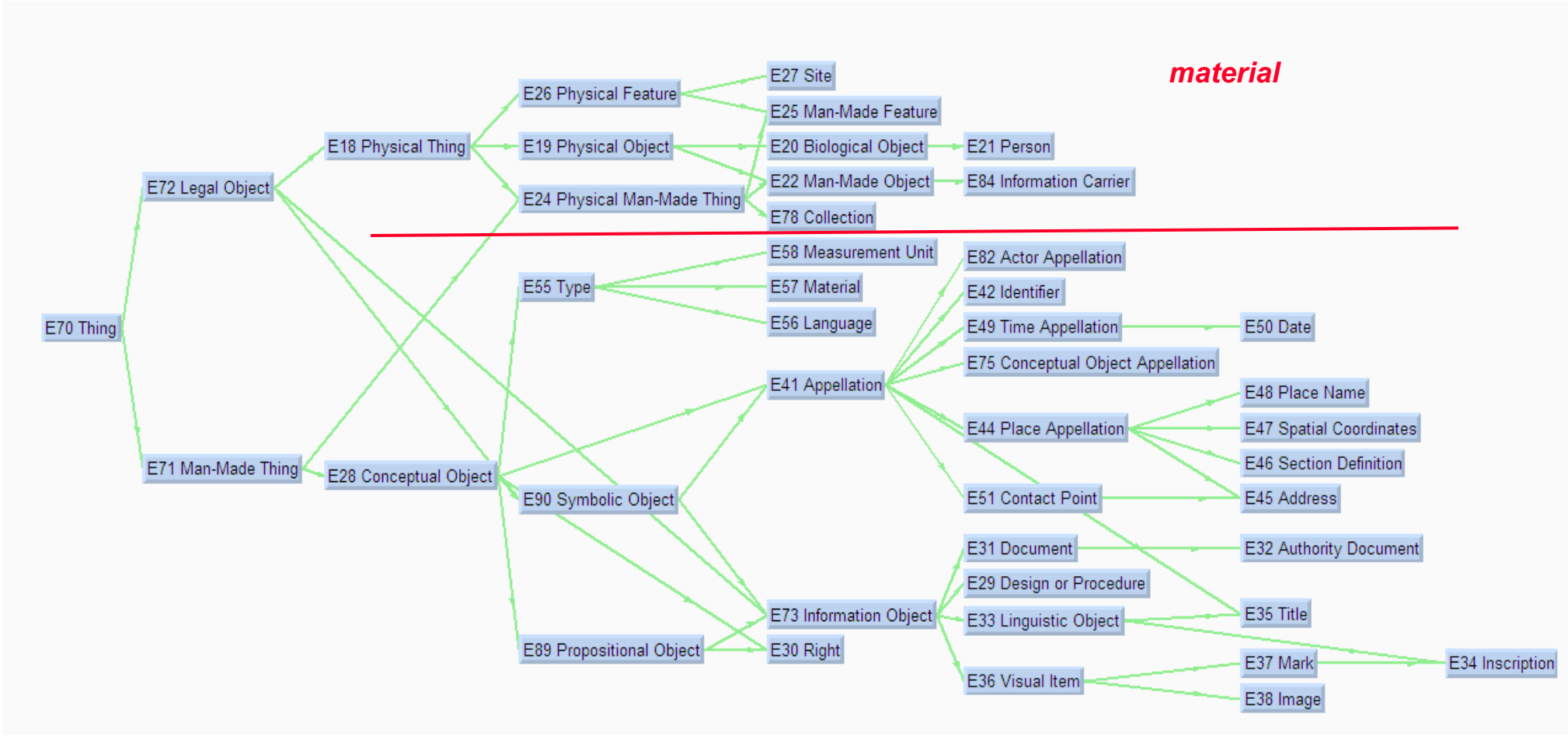
The CIDOC CRM

Taxonomic discourse



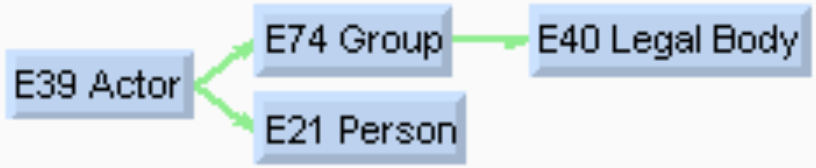
The CIDOC CRM

E70 Thing



The CIDOC CRM

E39 Actor



The CIDOC CRM
E53 Place

E53 Place

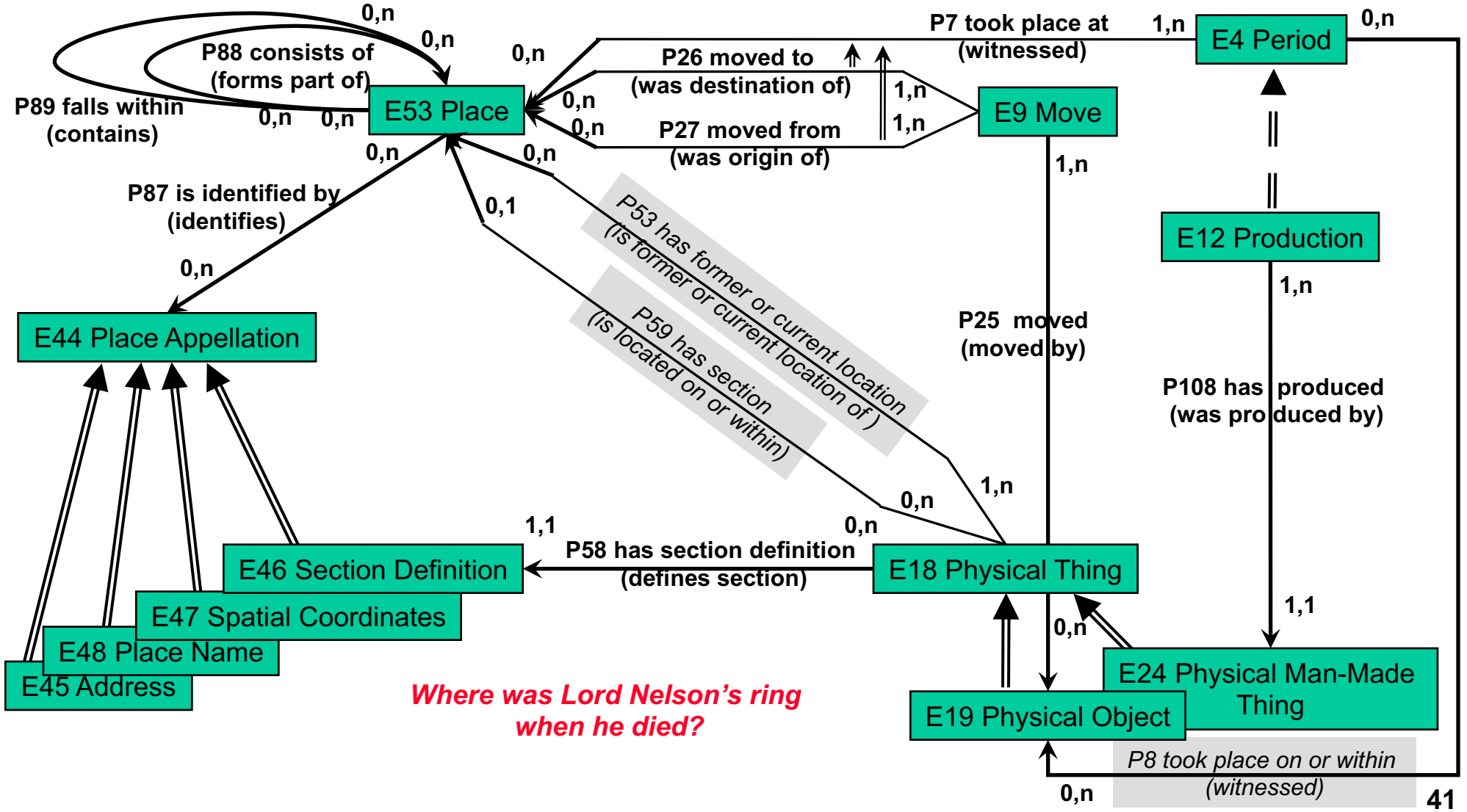
- **A place is an extent in space, determined diachronically with regard to a larger, persistent constellation of matter, often continents -**

  by coordinates, geophysical features, artefacts, communities, political systems, objects - but not identical to

- **A "CRM Place" is not a landscape, not a seat - it is an abstraction from temporal changes - "the place where…"**
- **A means to reason about the "where" in multiple reference systems.**
- **Examples:**
  - figures from the bow of a ship
  - African dinosaur foot-prints in Portugal
  - where Nelson died

The CIDOC CRM

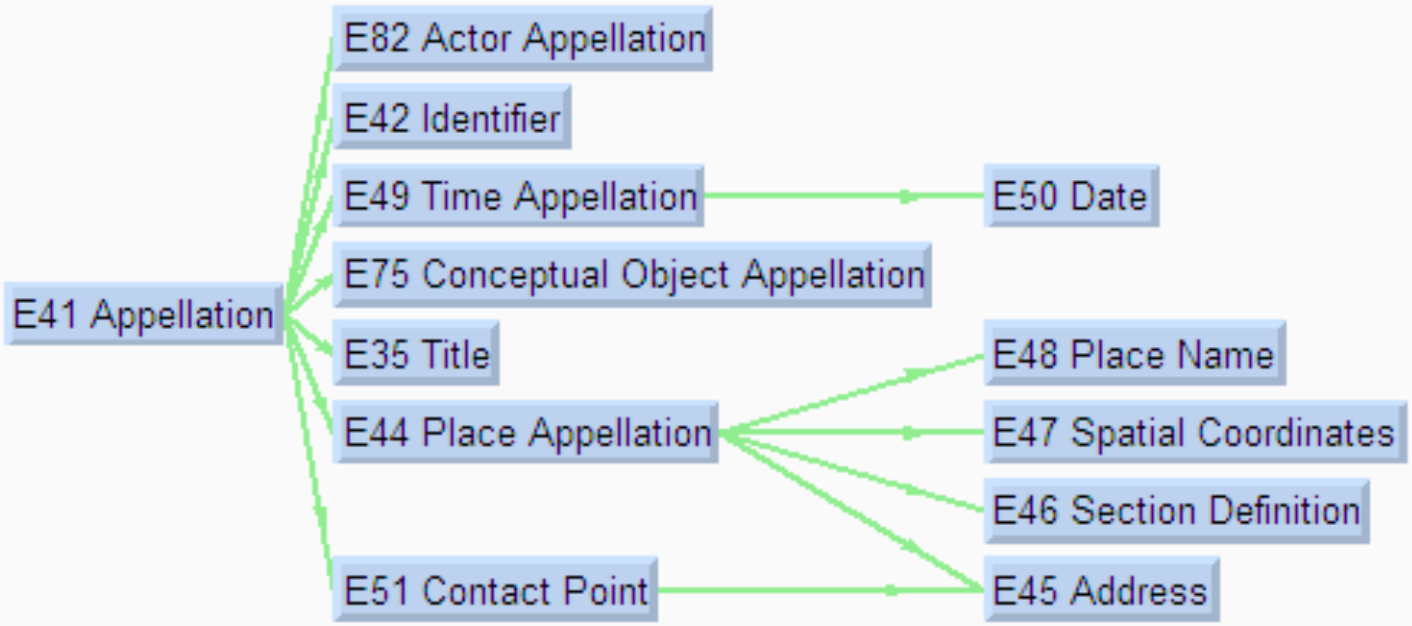
Properties of E53 Place



Where was Lord Nelson's ring when he died?

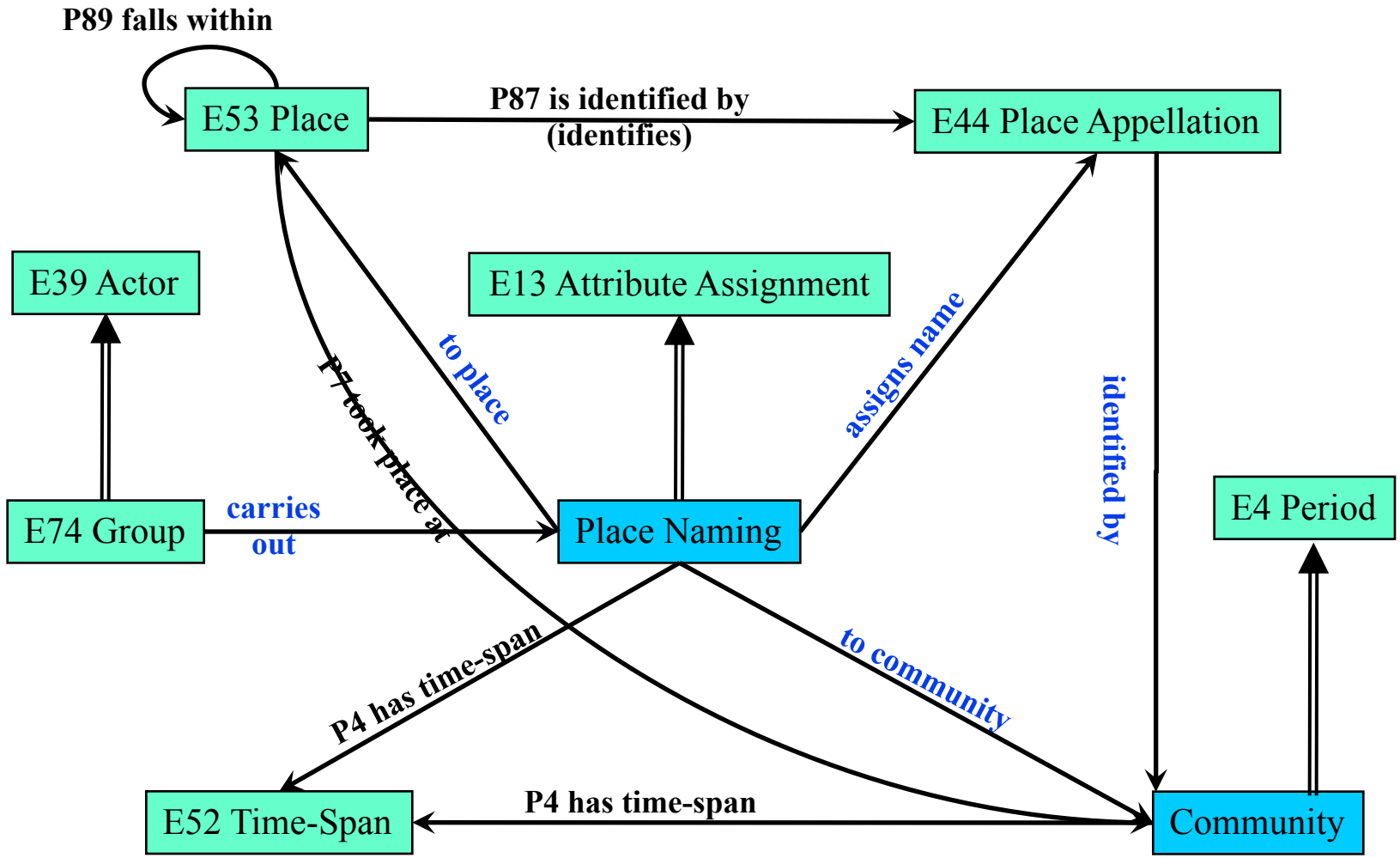
The CIDOC CRM

E41 Appellation



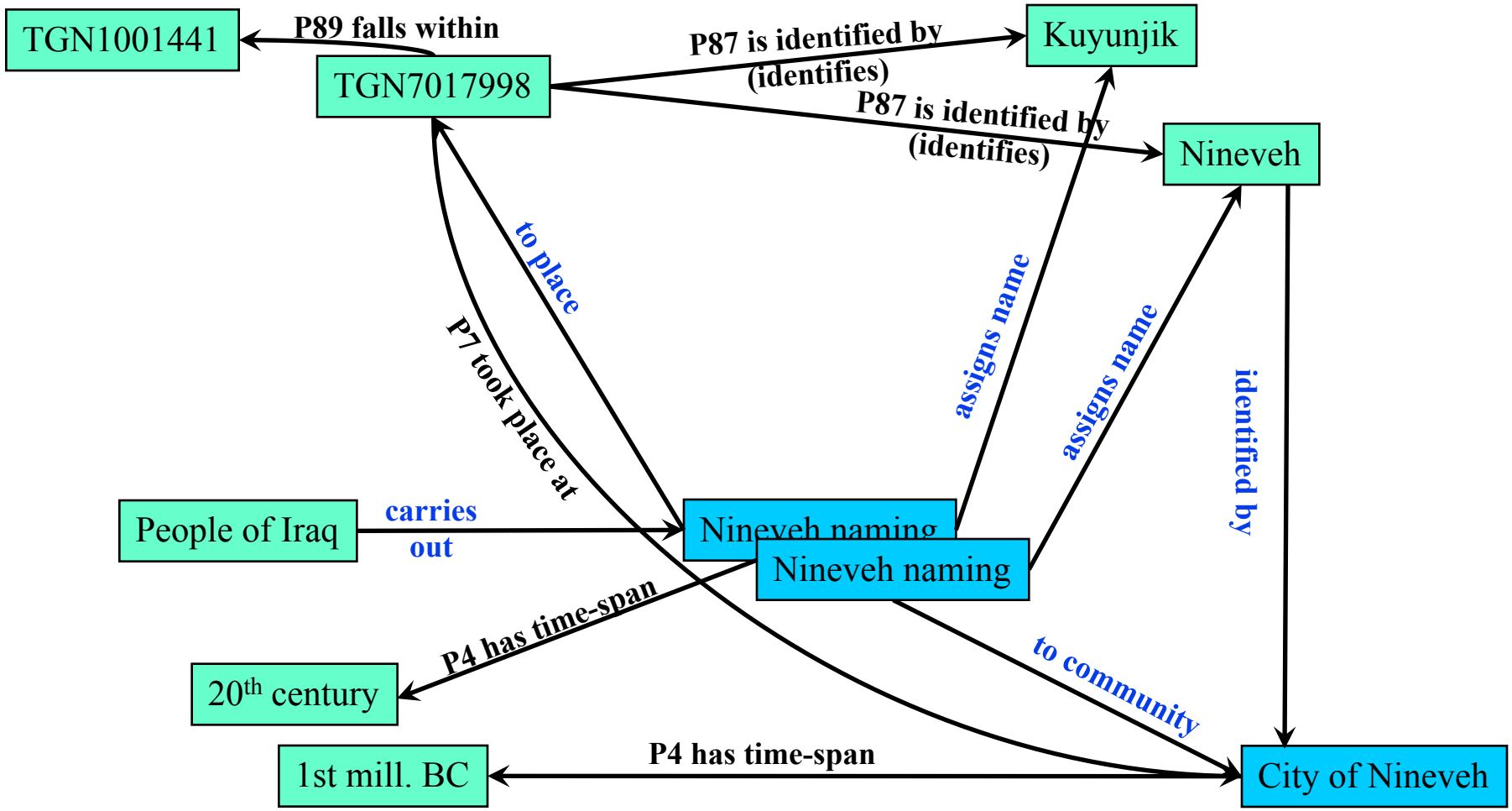
The CIDOC CRM

Extension Example: Getty's TGN



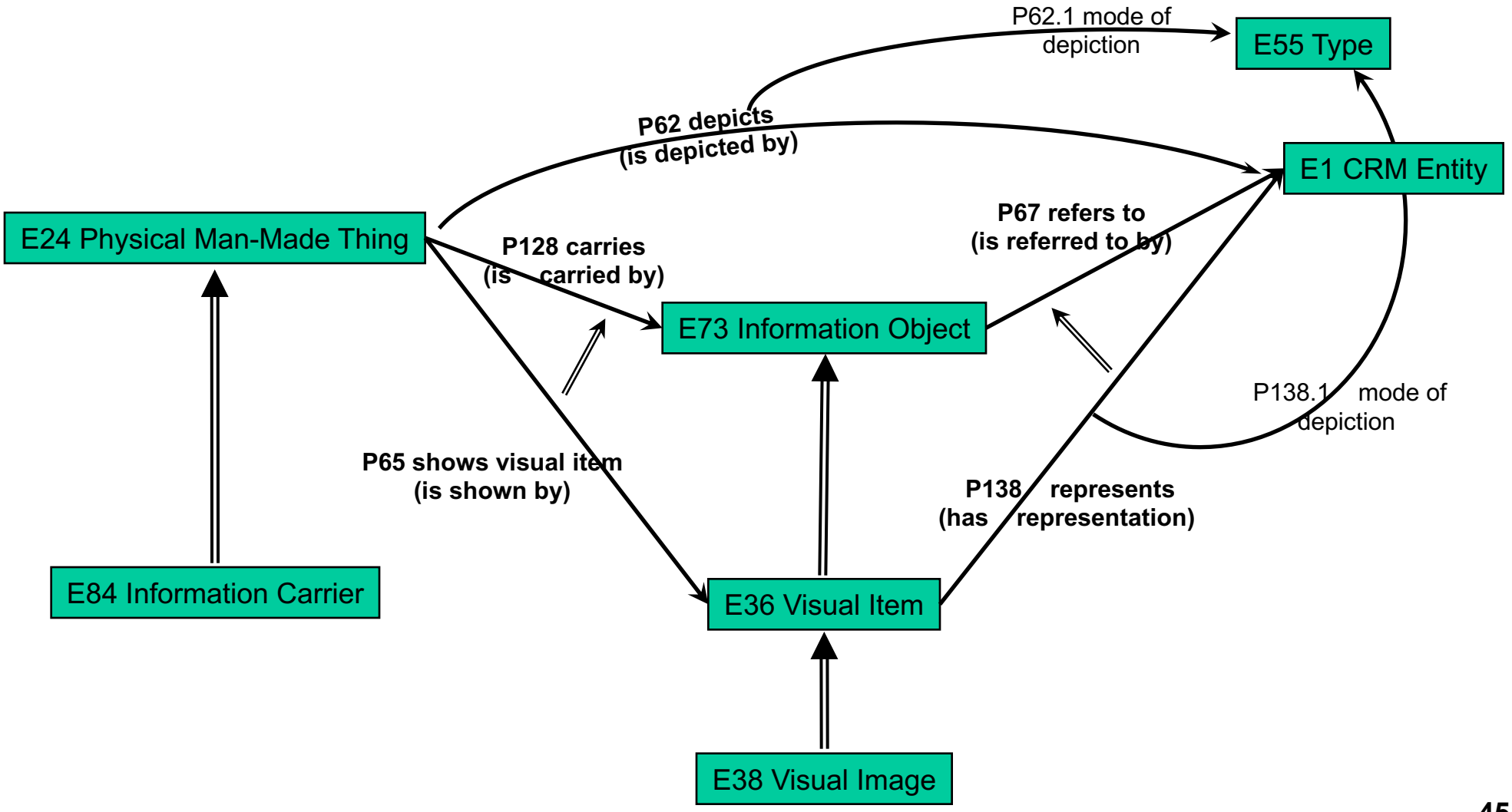
The CIDOC CRM

Sample of the TGN extension



The CIDOC CRM

Visual Content and Subject

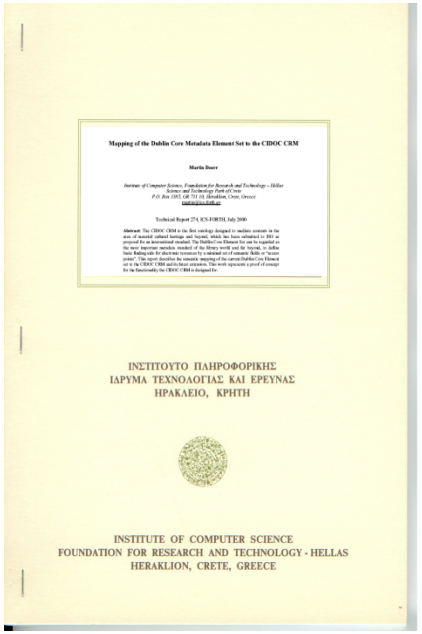


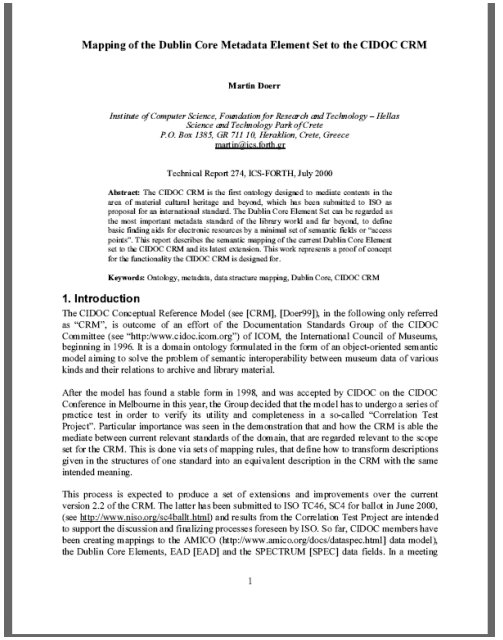
The CIDOC CRM

Application: Mapping DC to the CRM (1)

Example: DC Record about a Technical Report

Type: text
Title: Mapping of the Dublin Core Metadata Element Set to the CIDOC CRM
Creator: Martin Doerr
Publisher: ICS-FORTH
Identifier: FORTH-ICS / TR 274 July 2000
Language: English





Mapping of the Dublin Core Metadata Element Set to the CIDOC CRM

Martin Doerr

Institute of Computer Science, Foundation for Research and Technology – Hellas
Science and Technology Park of Crete
P.O. Box 1385, GR 711 10, Heraklion, Crete, Greece
martin@ics.forth.gr

Technical Report 274, ICS-FORTH, July 2000

Abstract: The CIDOC CRM is the first ontology designed to mediate contents in the area of material cultural heritage and beyond, which has been submitted to ISO as proposal for an international standard. The Dublin Core Element Set can be regarded as the most important metadata standard of the library world and far beyond, to define basic finding aids for electronic resources by a minimal set of semantic fields or "access points". This report describes the semantic mapping of the current Dublin Core Element set to the CIDOC CRM and its latest extension. This work represents a proof of concept for the functionality the CIDOC CRM is designed for.

Keywords: Ontology, metadata, data structure mapping, Dublin Core, CIDOC CRM

1. Introduction

The CIDOC Conceptual Reference Model (see [CRM], [Doer99]), in the following only referred as "CRM", is outcome of an effort of the Documentation Standards Group of the CIDOC Committee (see "http:/www.cidoc.icom.org") of ICOM, the International Council of Museums, beginning in 1996. It is a domain ontology formulated in the form of an object-oriented semantic model aiming to solve the problem of semantic interoperability between museum data of various kinds and their relations to archive and library material.

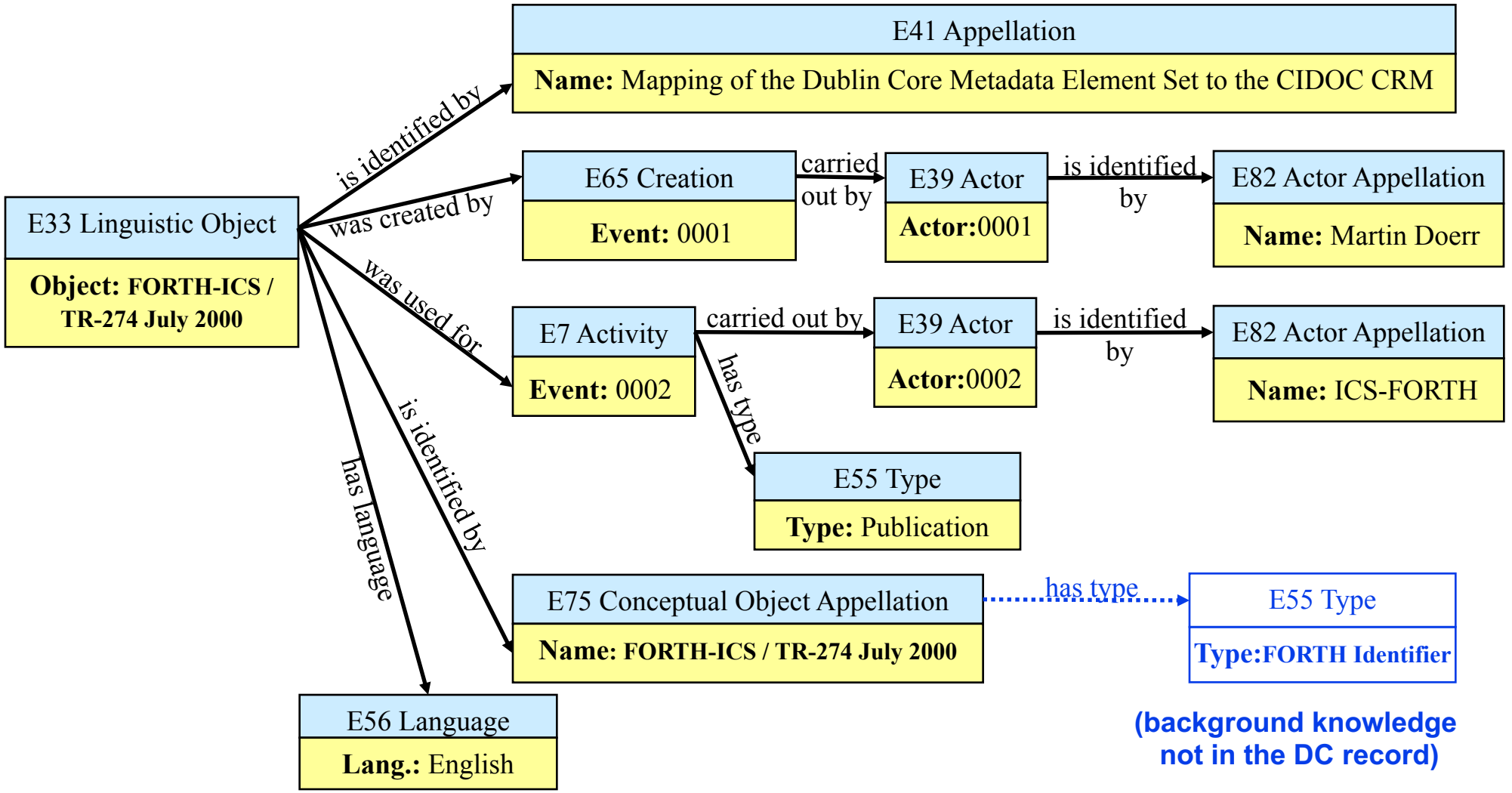
After the model has found a stable form in 1998, and was accepted by CIDOC on the CIDOC Conference in Melbourne in this year, the Group decided that the model has to undergo a series of practice test in order to verify its utility and completeness in a so-called "Correlation Test Project". Particular importance was seen in the demonstration that and how the CRM is able the mediate between current relevant standards of the domain, that are regarded relevant to the scope set for the CRM. This is done via sets of mapping rules, that define how to transform descriptions given in the structures of one standard into an equivalent description in the CRM with the same intended meaning.

This process is expected to produce a set of extensions and improvements over the current version 2.2 of the CRM. The latter has been submitted to ISO TC46, SC4 for ballot in June 2000, (see http://www.niso.org/sc4ballt.html) and results from the Correlation Test Project are intended to support the discussion and finalizing processes foreseen by ISO. So far, CIDOC members have been creating mappings to the AMICO (http://www.amico.org/docs/dataspec.html] data model), the Dublin Core Elements, EAD [EAD] and the SPECTRUM [SPEC] data fields. In a meeting

1

The CIDOC CRM

Application: Mapping DC to the CRM (2)

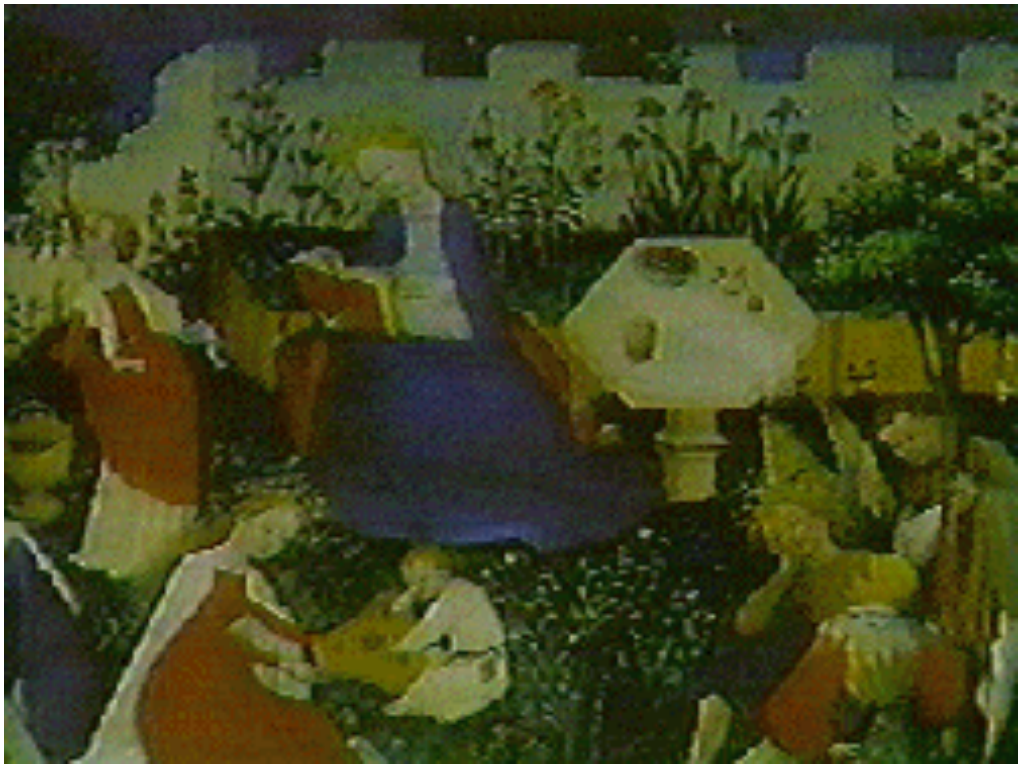


The CIDOC CRM

Application: Mapping DC to the CRM (3)

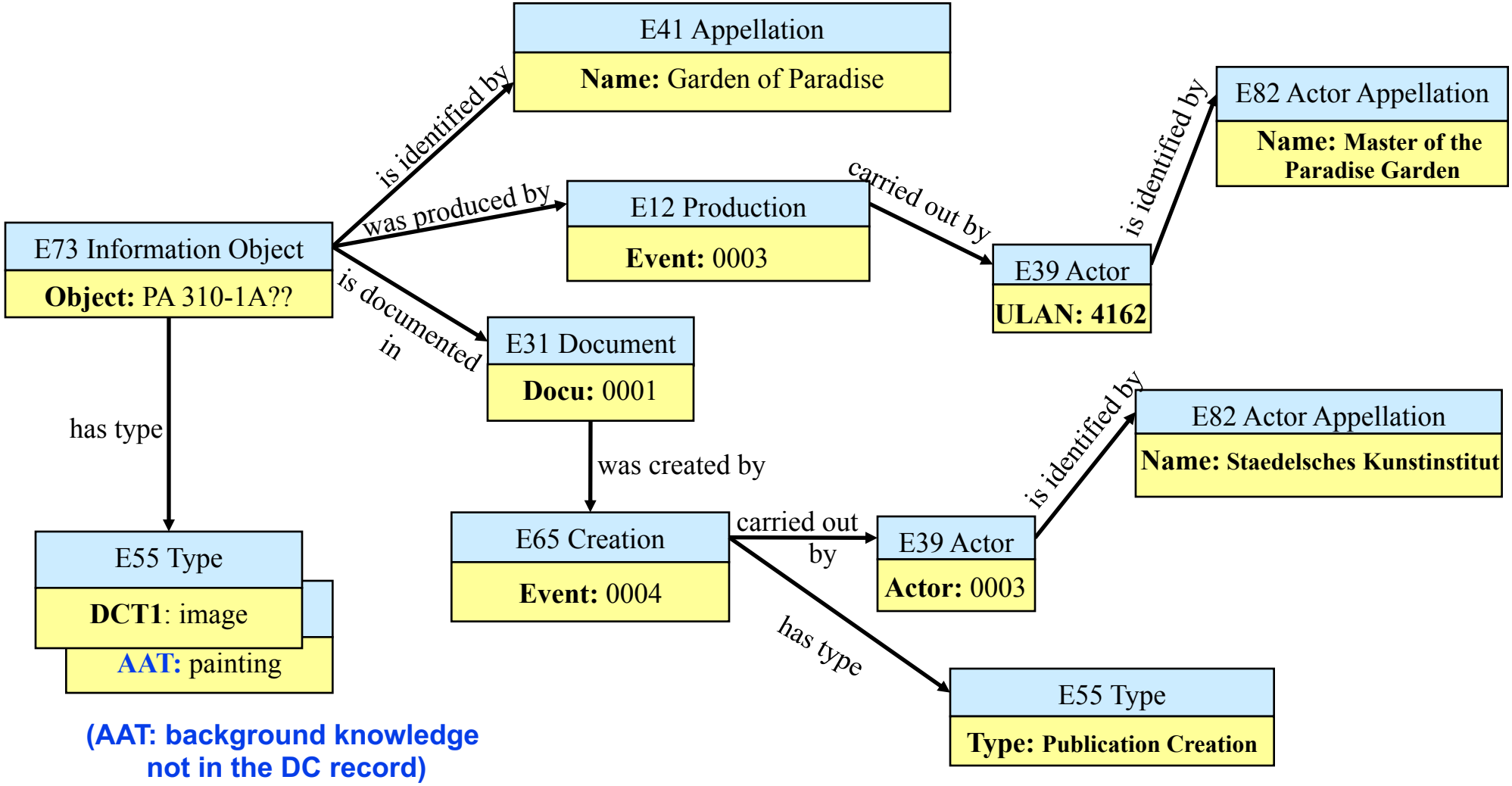
Example: DC Record about a painting

Type.DCT1: image
Type: painting
Title: Garden of Paradise
Creator: Master of the Paradise Garden
Publisher: Staedelsches Kunstinstitut



The CIDOC CRM

Application: Mapping DC to the CRM (4)



(AAT: background knowledge not in the DC record)

The CIDOC CRM
Lessons from mapping experiences

- **Semantic Interoperability can defined by the capability of mapping**
- **Mapping for epistemic networks is relatively simple:**
  - *Specialist / primary information databases frequently employ a flat schema, reducing complex relationships into simple fields*
  - *Source fields frequently map to composite paths under the CRM, making semantics explicit using a small set of primitives*
  - *Intermediate nodes are postulated or deduced (e.g., "birth" from "person"). They are the hooks for integration with complementary sources*
  - *Cardinality constraints must not be enforced= Alternative or incomplete knowledge*
- **Domain experts easily learn schema mapping**
  - *IT experts may not understand meaning, underestimate it or are bored by it!*
- **Intuitive tools for domain experts needed:**
  - **Separate identifier matching from schema mapping**
  - **Separate terminology mediation from schema mapping**

The CIDOC CRM

Differences to other ontologies

- Generally: Many ontologies:-
  - lack an empirical base
  - have a functionally insufficient system of relationships (terminology driven)
  - Have a lack of functional specifications
- The CRM misses concepts not in the empirical base (e.g., contracts), but detects concepts that are not lexicalized (e.g."Persistent Item"), because they are functionally required
- **DOLCE**: Lexical base, intuition. Very good theoretically motivated logical description. Foundational relationships. Over specified relationships (e.g. modes of participation). Bad model of space-time. Strong overlap with CRM
- **BFO**: Philosophically motivated. Poor model of relationships. Notion of a precise, deterministic underlying reality. Empirical verification difficult. Strong overlap with CRM
- **IndeCs, ABC Harmony**: Small ontologies, event centric, strong overlap with CRM (harmonized!)
- **SUMO**: Large aggregation of concepts without functional specifications

The CIDOC CRM

Applications: Integration with CRM Core (1)

E84 Information Carrier
The "Monument to Balzac"(S1296)

P62 depicts

E21 Person
Honoré de Balzac

P62 depicts

E84 Information Carrier
The "Monument to Balzac" (plaster)

P16B was used for

P108B was produced by

E12 Production
Bronze casting"Monument to Balzac" in 1925

P134 continued

P108B was produced by

P2 has type

E55 Type
plaster

P2 has type

E55 Type
bronze

P4 has time-span

E52 Time-Span
1925

P120B occurs after

P14 carried out by

E12 Production
Rodin making "Monument to Balzac" in 1898

P7 took place at

E52 Time-Span
1917

P4 has time-span

E40 Legal Body
Rudier (Vve Alexis) et Fils

P14 carried out by

E69 Death
Rodin's death

P4 has time-span

E52 Time-Span
1898

P2 has type

E55 Type
companies

E55 Type
sculptors

P2 has type

E53 Place
France (nation)

P100B died in

E21 Person
Rodin Auguste

P98B was born

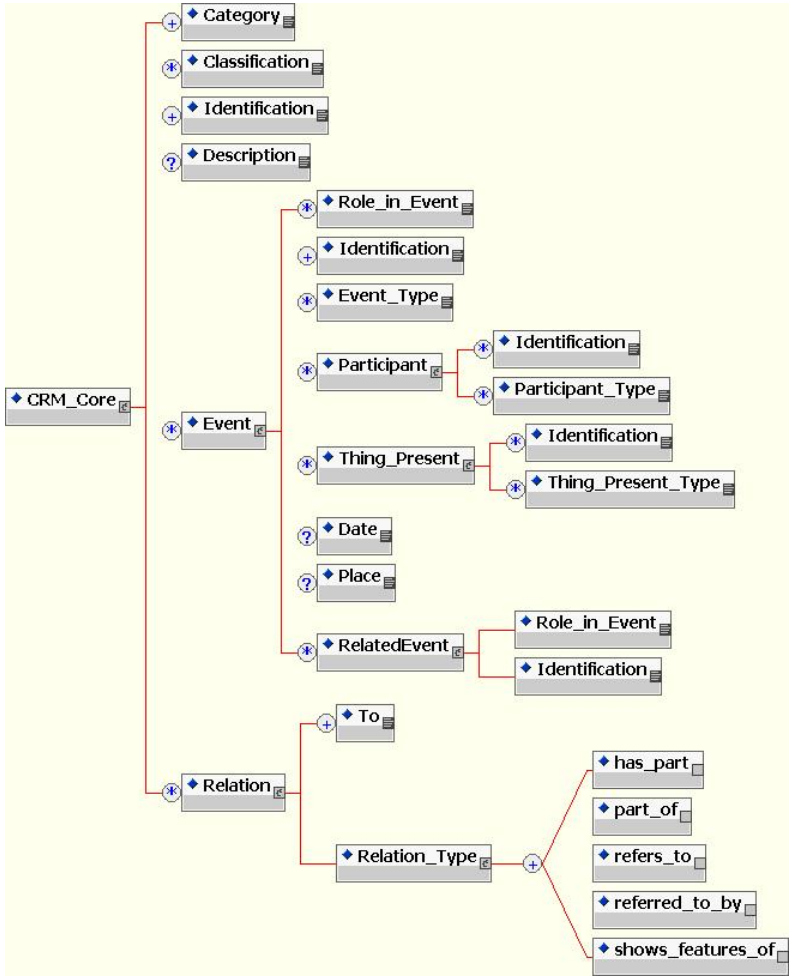
E67 Birth
Rodin's birth

P4 has time-span

E52 Time-Span
1840

The CIDOC CRM

Applications: Integration with CRM Core (2)



CRM Core

A minimal metadata element set

Artist (CRM Core).

Category = E21 Person

Classification = artists

Classification = sculptors

Identification =Rodin, Auguste

Identification =ID: 500016619

Event

Role in Event =P98B was born
Identification= Rodin‘s birth
Event Type = E67 Birth
Date = 1840

Event

Role in Event =P100B died in
Identification= Rodin‘s death
Event Type = E69_Death
Date = 1917
Related event
Role in Event =P120 occurs before
Identification= Bronze casting Monument to Balzac in 1925

Work (CRM Core).

Category = E84 Information Carrier
Classification =sculpture (visual work)
Classification =plaster
Identification =The Monument to Balzac (plaster)
Description =Commissioned to honor one of France's greatest novelists, Rodin spent seven years preparing for Monument to Balzac. When the plaster original was exhibited in Paris in 1898, it was widely attacked. Rodin retired the plaster model to his home in the Paris suburbs. It was not cast in bronze until years after his death.

Event
Role in Event =P108B was produced by
Identification= Rodin making Monument to Balzac in 1898
Event Type = E12 Production
Participant
Identification =Rodin, Auguste
Identification =ID: 500016619
Participant Type = artists
Participant Type = sculptors
Date = 1898
Place = France (nation)
Related event
Role in Event =P134B was continued by
Identification= Bronze casting Monument to Balzac in 1925

Event
Role in Event =P16B was used for
Identification= Bronze casting Monument to Balzac in 1925
Event Type = E12 Production
Participant
Identification =Rudier (Vve Alexis) et Fils
Participant Type = companies
Thing Present
Identification =The Monument to Balzac (S.1296)
Thing Present Type =bronze
Thing Present Type =sculpture (visual work)
Date = 1925
Related event
Role in Event =P120B occurs after
Identification= Rodin's death

Relation
To = Honore de Balzac
Relation type
refers to

The CIDOC CRM

Methodological aspects

- The CRM aims at semantic integration beyond context.
- It aims at pulling together all relevant sources and data to evaluate a scientific or scholarly question not answered by an individual document
- Based on the CRM, effective integration schemata can be defined, such as "CRM Core", the full CRM or extensions of the CRM
- The CRM can fit rich and poor models under one common logical frame-work . For instance Dublin Core (DC) maps to the CRM
- Idea: Not being prescriptive creates lots of flexibility
  - It does not propose what to describe. It allows interpretation of what museums and archives actually describe

The CIDOC CRM
Documents and Knowledge

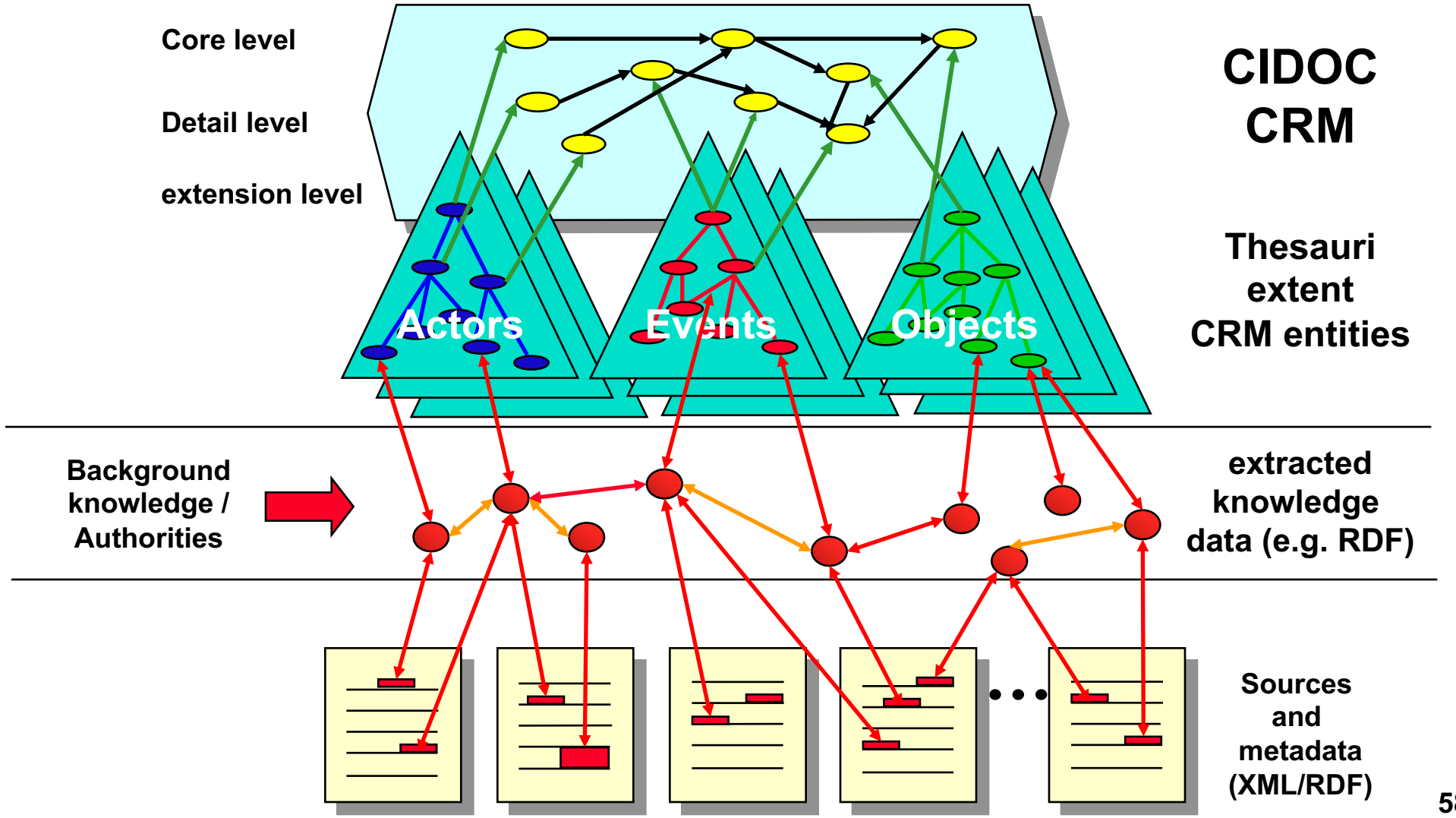
- Scientific and scholarly work produces knowledge by argumentation
  - This comes in closed units, "documents"
  - They have a history of evolution, "versions"
  - The knowledge is "directed"
  - It can only be evaluated in context
    - document about Mona Lisa
    - theory about the origin of the Minoan people
- It should be possible to map primary document structures to the CRM. This is easy:
  - E.g. good is: "creator - creation place - creation date"
    bad is : "provenance", "place associations - life-cycle dates" etc.
  - Good document structures map easily
  - No completeness requirements

The CIDOC CRM
Knowledge management

- **Three-level knowledge management:**
  - **Acquisition (can be motivated by the CRM):**
    - **sequence and order, completeness, constraints to guide and control data entry.**
    - **ergonomic, case-specific language, optimized to specialist needs**
    - **often working on series of analogous items**
    - Low interoperability needs (**capability to be mapped!**)
  - **Integration / comprehension (target of the CRM):**
    - break up document boundaries, relate facts to wider context
    - match **shared identifiers** of items, **aggregate alternatives**
    - no preference direction of search, no cardinality constraints
    - High interoperability needs (**mapping to a global schema**)
  - **Presentation, story-telling (can be based on CRM)**
    - **explore context, paths, analogies orthogonal to data acquisition**
    - **present in order, allow for digestion**
    - deduction and induction

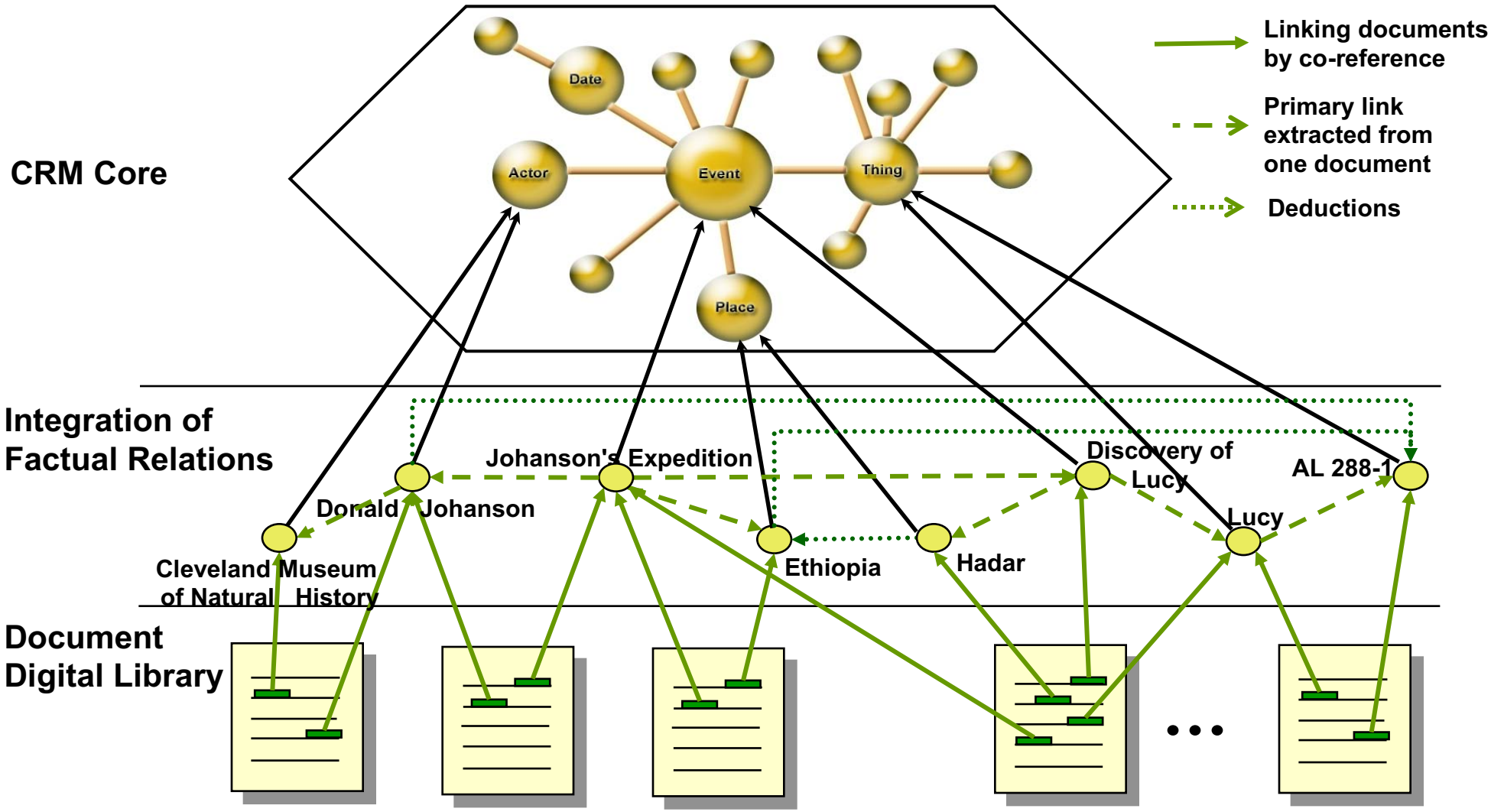
The CIDOC CRM -Application

Repository Indexing



The CIDOC CRM

Documents and Factual Knowledge



The CIDOC CRM

Benefits of the CRM (From Tony Gill)

- Elegant and simple compared to comparable Entity-Relationship models
- Coherently integrates information at varying degrees of detail
- Readily extensible through O-O class typing and specializations
- Richer semantic content; allows inferences to be made from underspecified data elements
- Designed for mediation of cultural heritage information

The CIDOC CRM
State of Development

- Publication as ISO 21127:2006 in October 2006
- Work on extension covering FRBR, FRAD and CRM resulted in "$FRBR_{OO}$", accepted by IFLA and CIDOC
- Ongoing work on TEI – CRM harmonization
- Application models (CRM Core, good and rich data exchange formats, extensions)
- OWL version being finalized

The CIDOC CRM
Conclusions

- Doing all that, we encounter a surprise compared with common preconceptions:
  - Nearly no domain specificity (e.g.“current permanent location”), generic concepts appear in medicine, biodiversity etc.
  - Rather a notion of scientific method emerges, such as “retrospective analysis”, “taxonomic discourse” etc.
  - Extraordinary small set of concepts
  - Extraordinary convergence: adding dozens of new formats hardly introduces any new concept
- This approach is economic, investment pays off
  - The CRM should become our language for semantic interoperability,