



# *The CIDOC CRM*

## ***Making Your Data Fit for The Semantic Web***

***Martin Doerr***

***Center for Cultural Informatics,  
Institute of Computer Science  
Foundation for Research and Technology - Hellas***

**Washington, October 25, 2009**



# The CIDOC CRM:

## *Outline*

### *Outline:*

- Making your data fit
- Historical discourse core semantics – the CRM
- Form and function of the CRM
- Content of the CRM
- A data modeling exercise
- Lessons learned



# The CIDOC CRM:

*Making Your Data Fit!*

*Two Problems: Form (Encoding) and Meaning (“Semantics”)*

## *About Form:*

- Documents (text, MS Word, PDF, HTML, XML) cannot mechanically be merged!
- Linking Documents is not even linking to what they mean!
- The meaning of documents can be encoded as sets of atomic statements:
  - *RDF/ Web technology does the job!*
  - *The internal encoding can be kept invisible to the user*
- Once we have the right form, we can make
  - *better systems,*
  - *transform and export data from legacy systems*
  - *migrate from system to system*
  - *import and merge data*
  - *link meaning across the Web*



# The CIDOC CRM:

*Making Your Data Fit!*

*About common meaning:*

- find a comprehensive core model of primitives sufficient to built more complex meaning as we encounter in cultural information system.

*But the 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

*and how to make a “global model”?*

- Data overlap, but do **not fit in one** schema
- What are the **relevant relationships** capable to **connect** the meaning in our documents?



# The CIDOC CRM:

## *Historical Archives...*

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

is about...

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

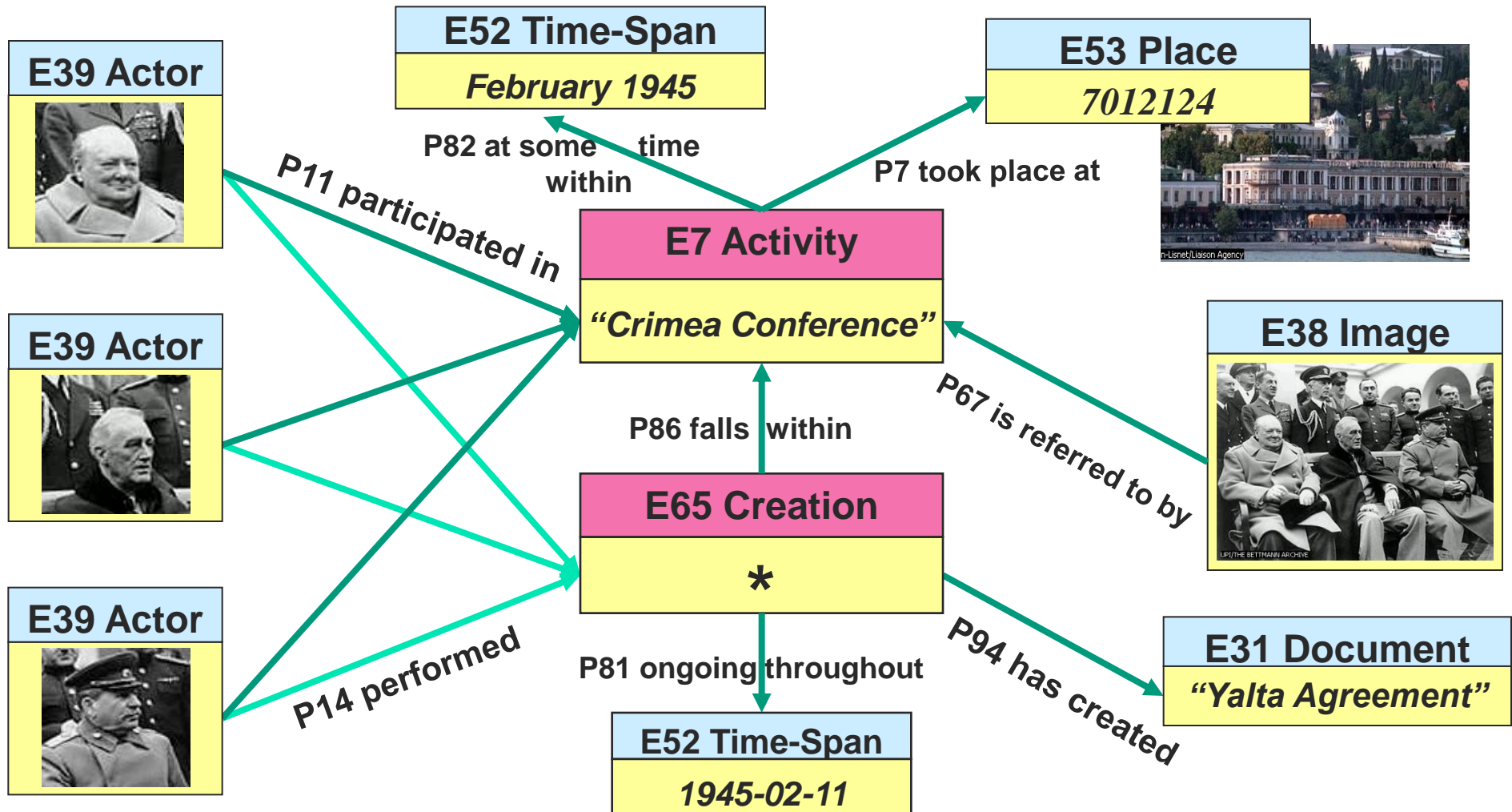






# The CIDOC CRM:

*Explicit Events, Object Identity, Symmetry*







# The CIDOC CRM:

## *Functions*

### *The CIDOC CRM...*

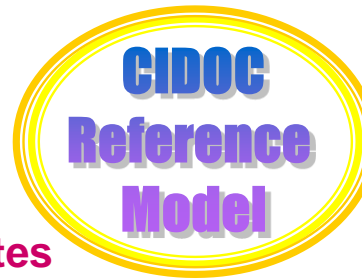
- ...captures the underlying semantics of relevant documentation structures in a **formal ontology**
- Ontologies are **formalized knowledge**: clearly defined concepts and relationships about **possible states of affairs** in a domain
- They can be understood by people and processed by machines to enable data exchange, data integration, query mediation etc.
- An “**extensible** ontology of **relationships**” providing **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:

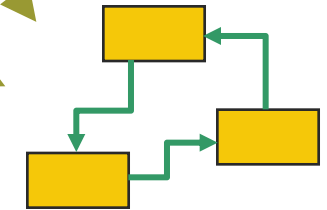
## *The Intellectual Role of the CRM*

### Conceptualization



approximates

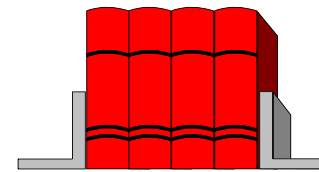
explains,  
motivates



Data structures &  
Presentation models

organize

Metadata



Data in various forms

refer to



World Phenomena



# 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,
- Currently, the **most natural** encoding is **RDF/ RDFS**
- 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:*

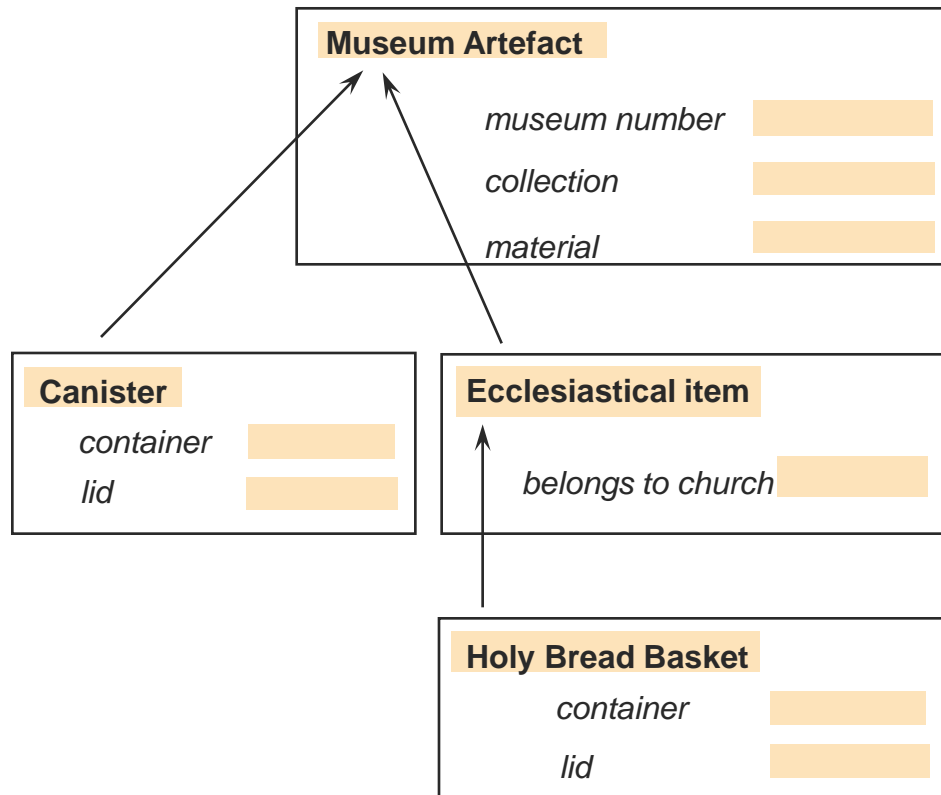
- Find **relevant relationships** – classes are only **anchors** of properties
- Classes without relationships are regarded as “terminology” (as described by **SKOS**!)
- **Events** are always **explicit** nodes (may be “**short-cut**” by frequently used **deductions**)



# The CIDOC CRM:

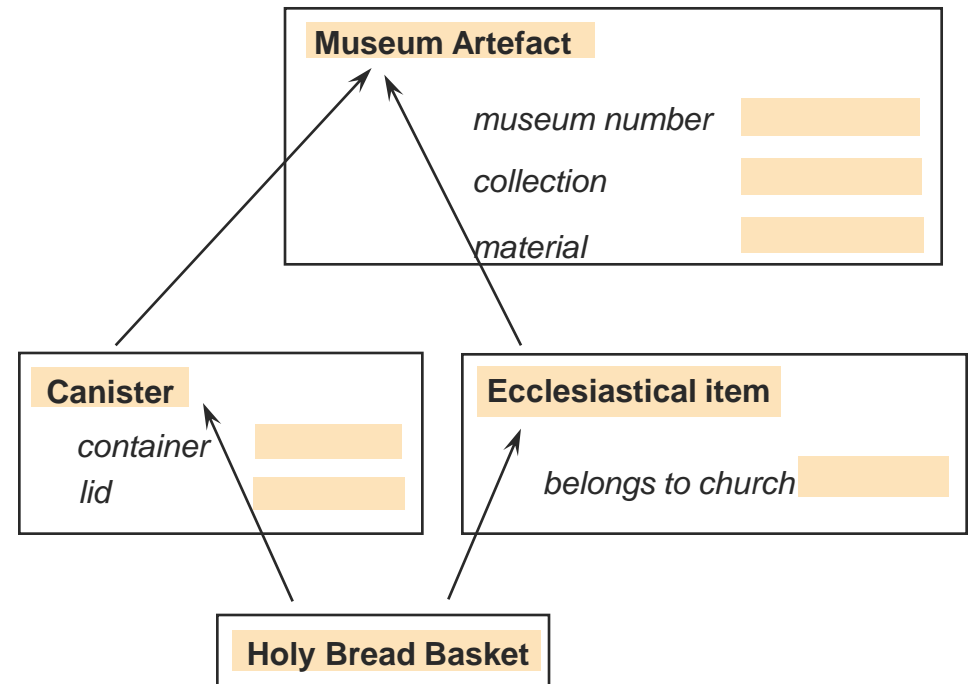
## *Justifying Multiple Inheritance*

### Single Inheritance form:



*Repetition of properties*

### Multiple Inheritance form:



*Unique identity of properties*



# The CIDOC CRM:

## *Data example (e.g from data 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**

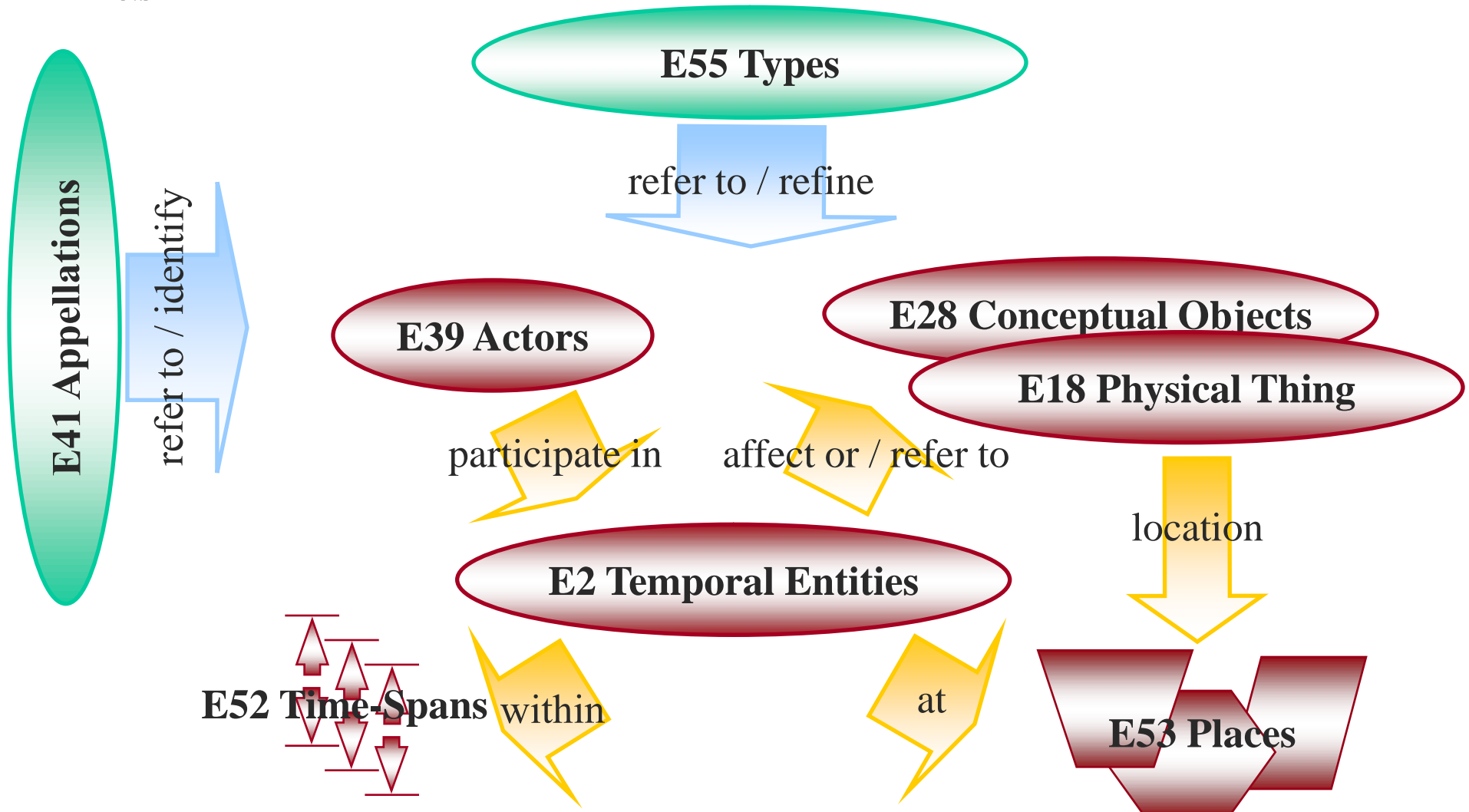
**continent** (entity **E55 Type**)

TGN data



# The CIDOC CRM:

*Top-level classes useful for integration*





# The CIDOC CRM:

## *Subjects of CRM Properties*

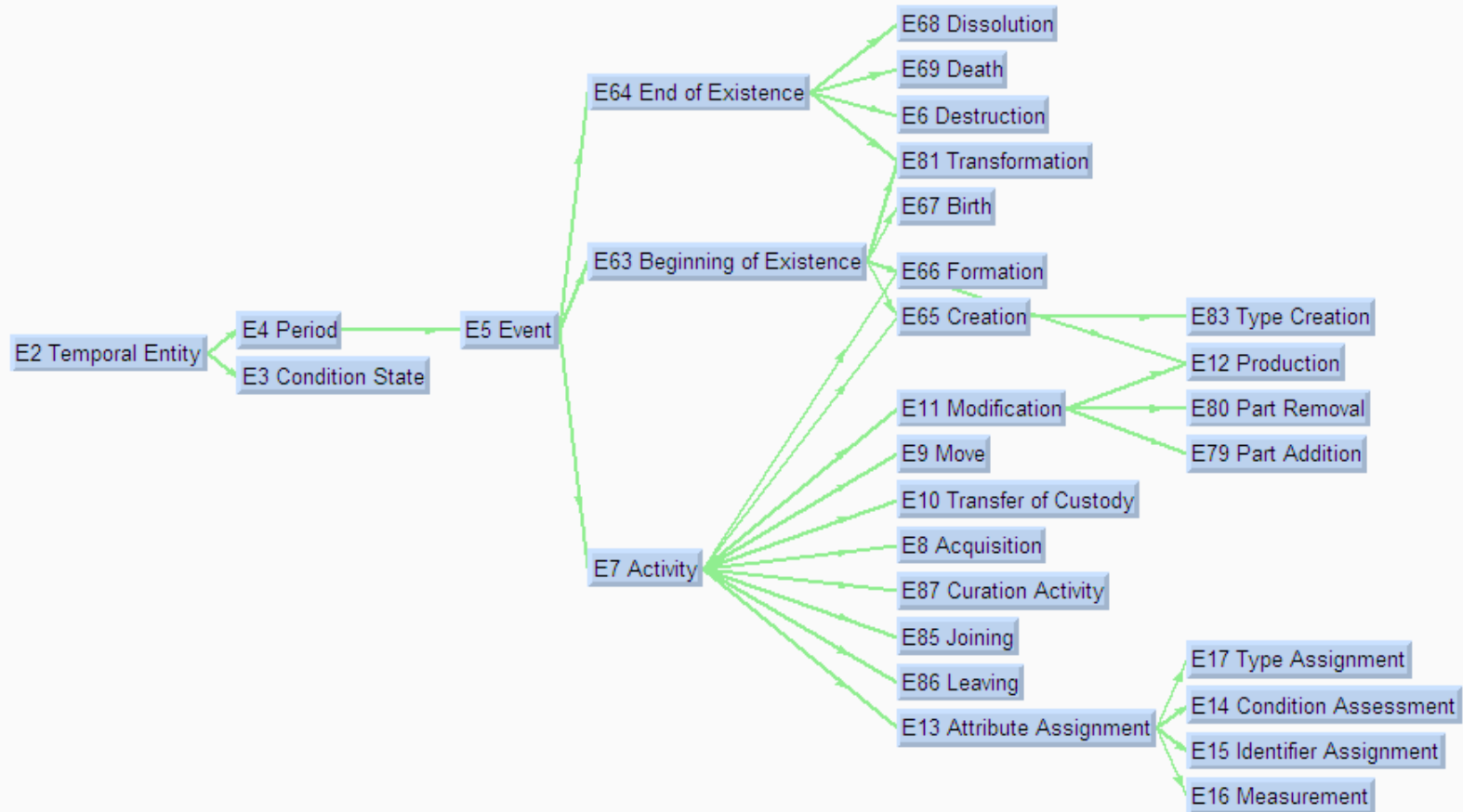
- Identification of real world items by real world names
- Observation and Classification of real world items
- Part-decomposition and structural properties of Conceptual & Physical Objects, Periods, Actors, Places and Times
- Participation of persistent items in temporal entities
  - *creates a notion 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:

## *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*

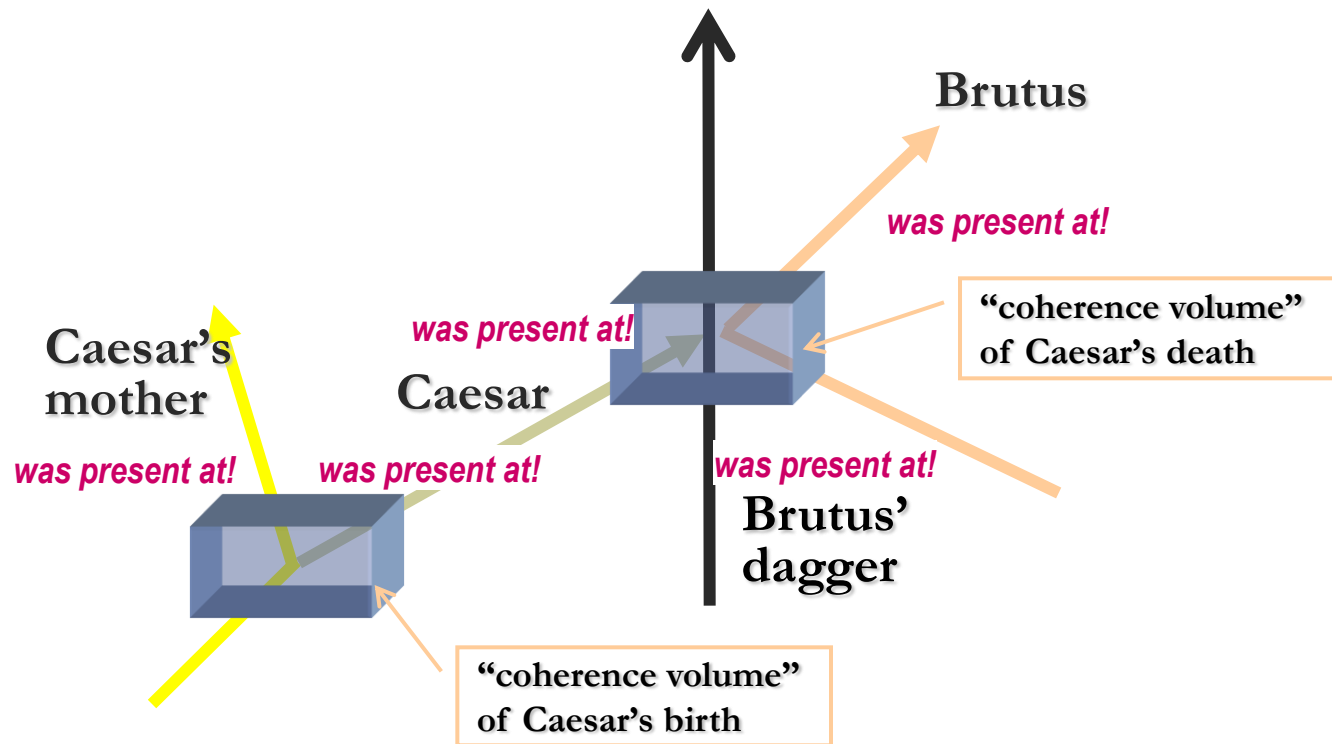




# The CIDOC CRM:

## *Historical Events as Meetings*

t



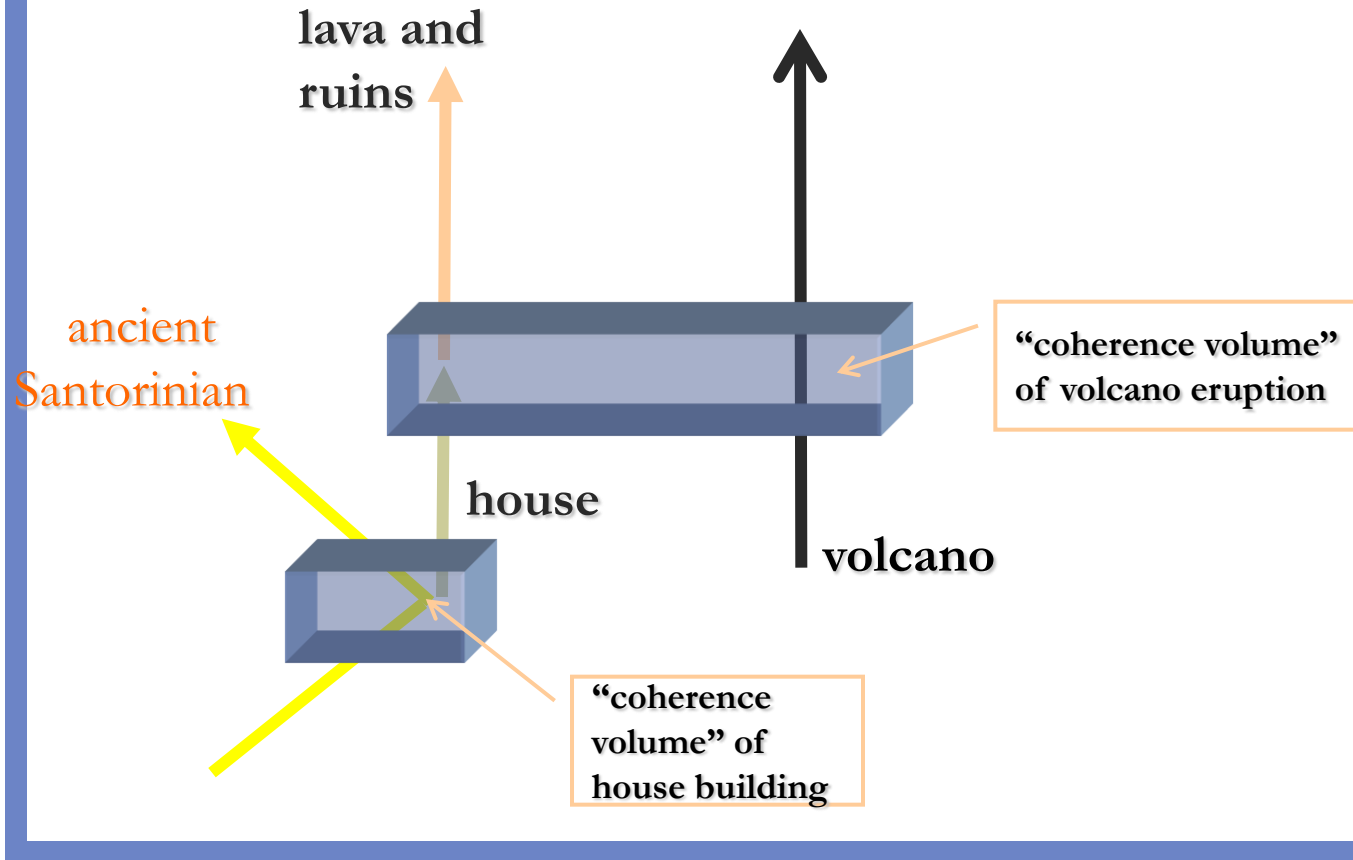
S



# The CIDOC CRM:

*Depositional events as meetings*

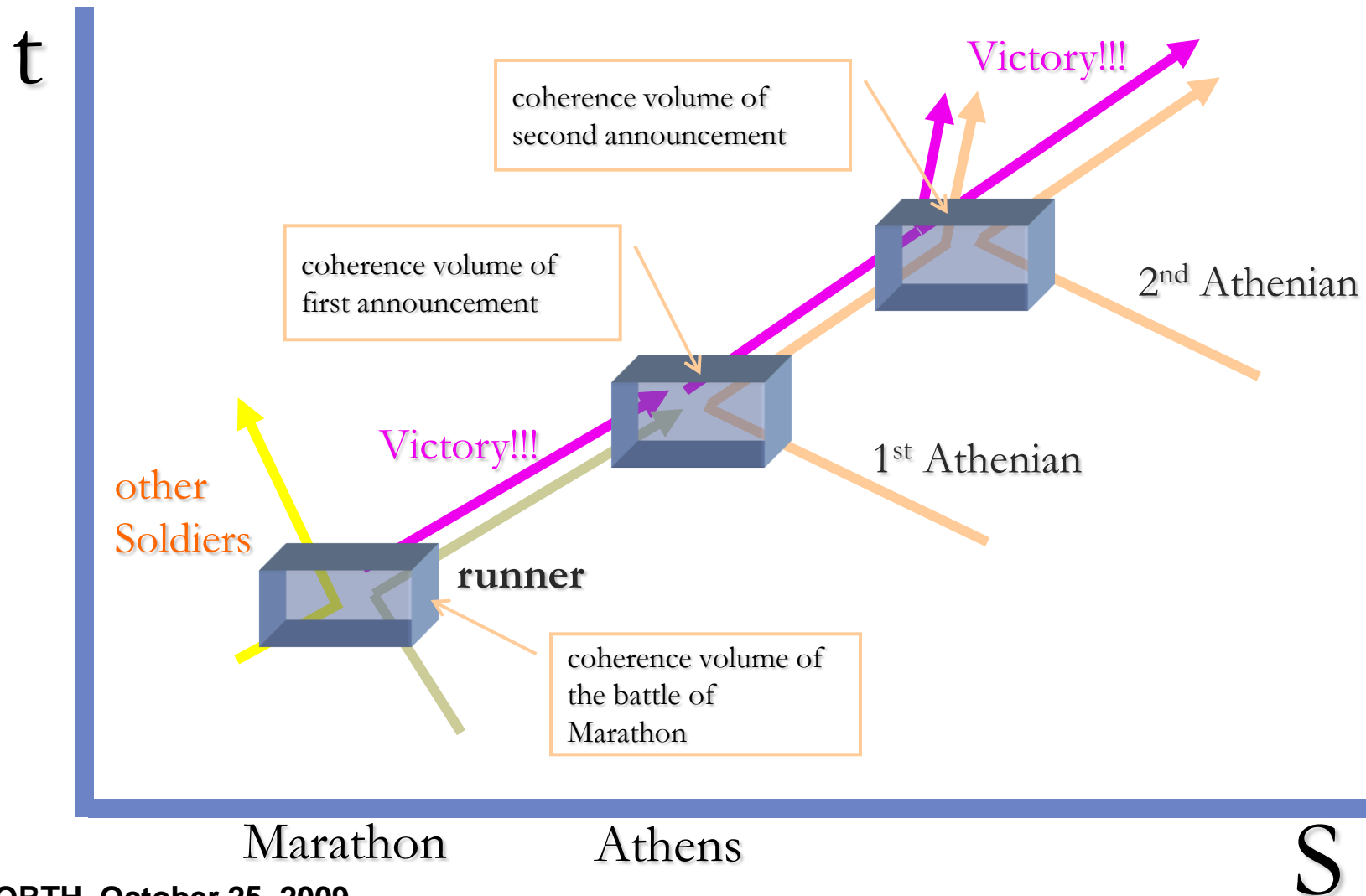
t





# The CIDOC CRM:

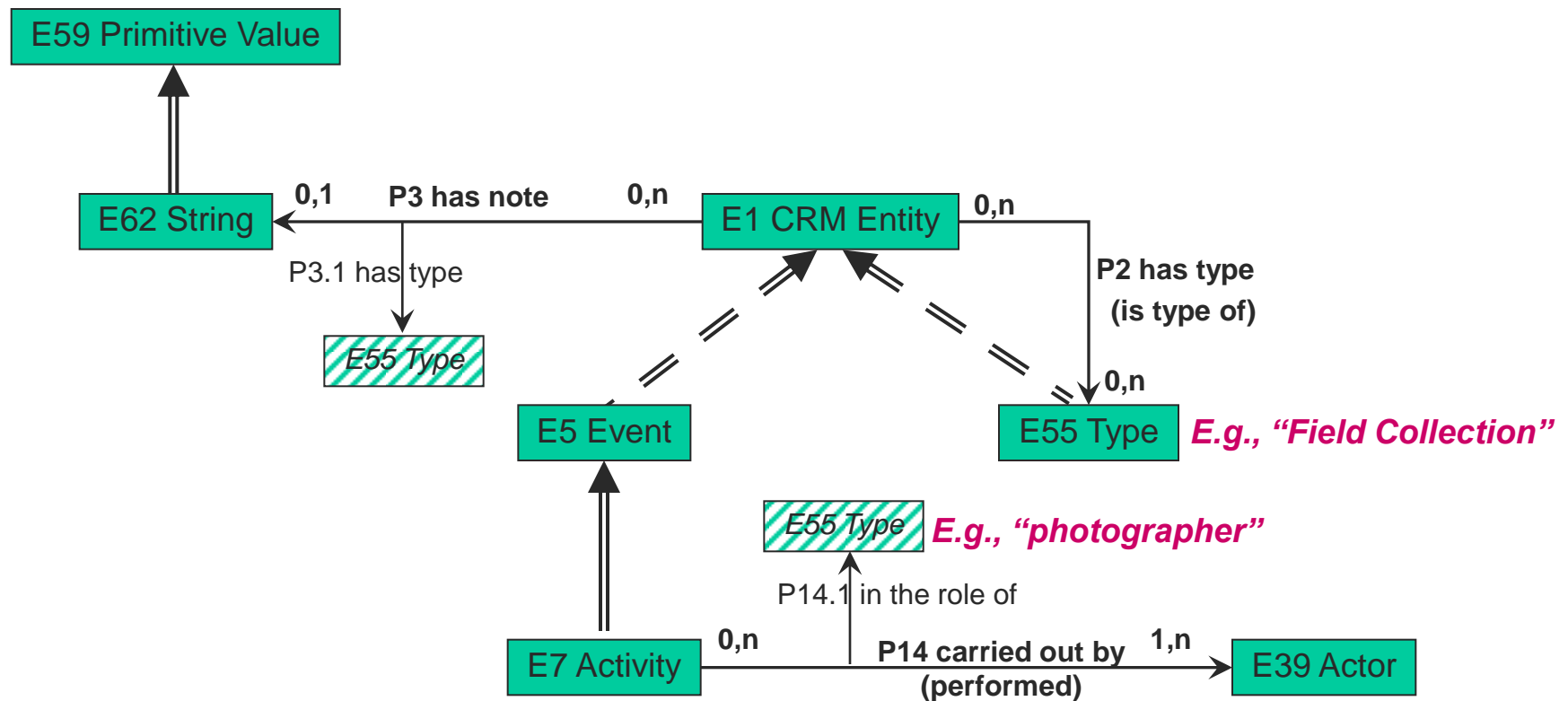
*Exchanges of information as meetings*





# The CIDOC CRM:

## *E7 Activity and inherited properties*

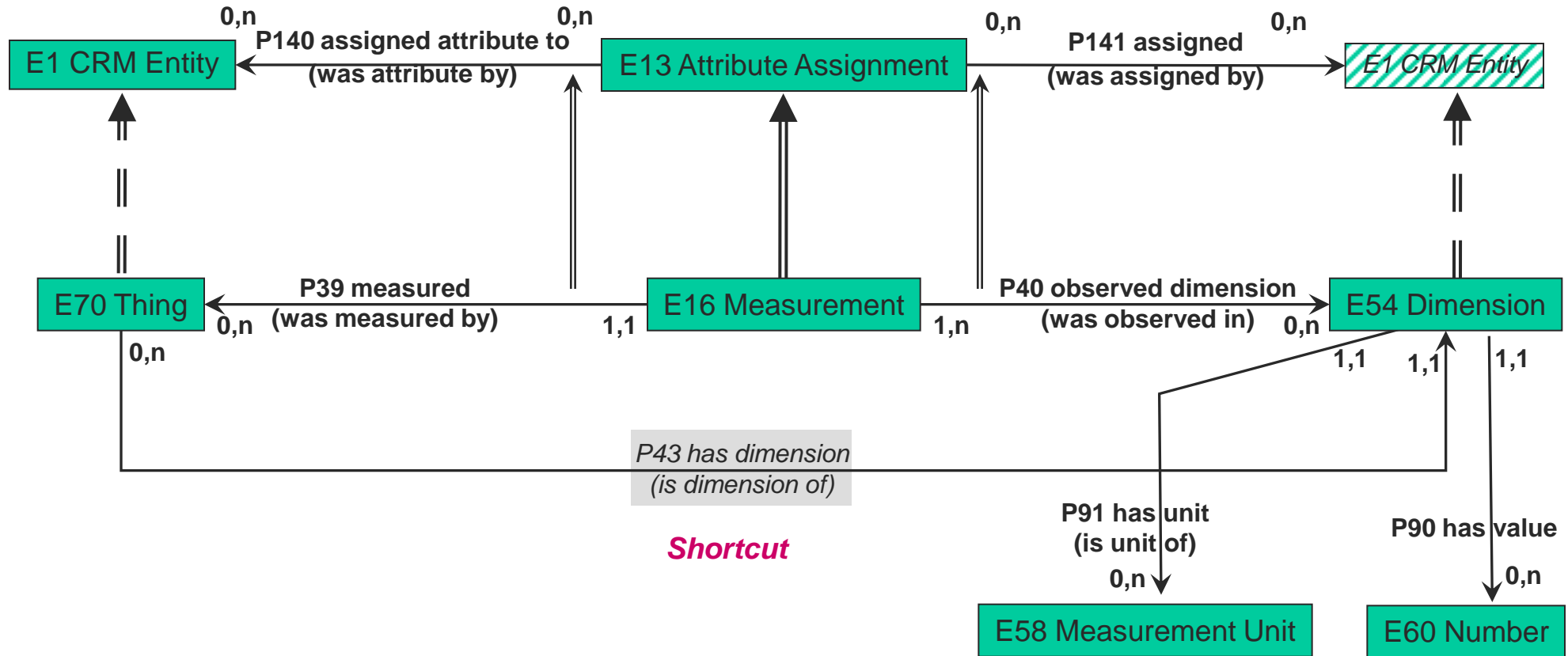






# The CIDOC CRM:

## Activities: E16 Measurement

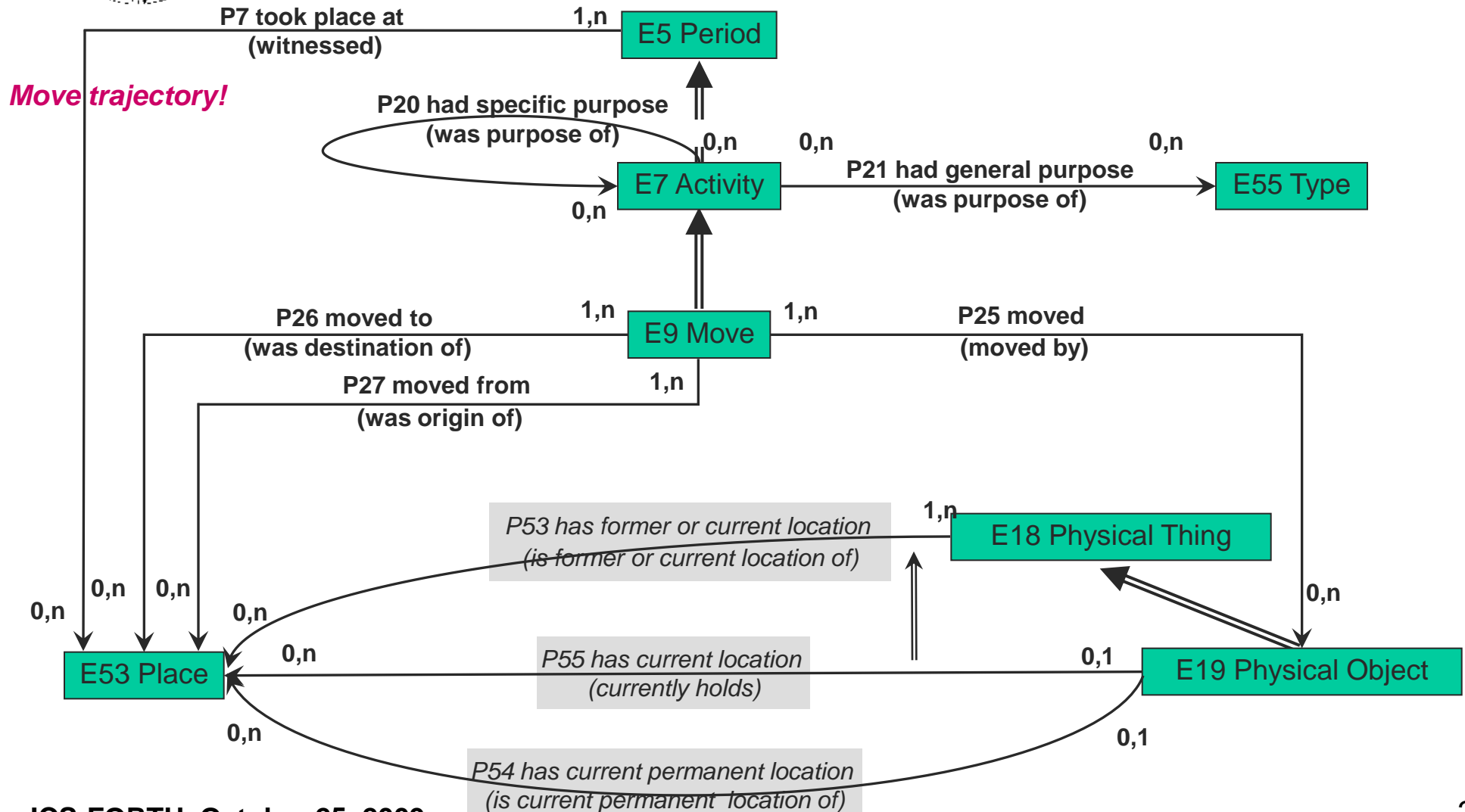






# The CIDOC CRM:

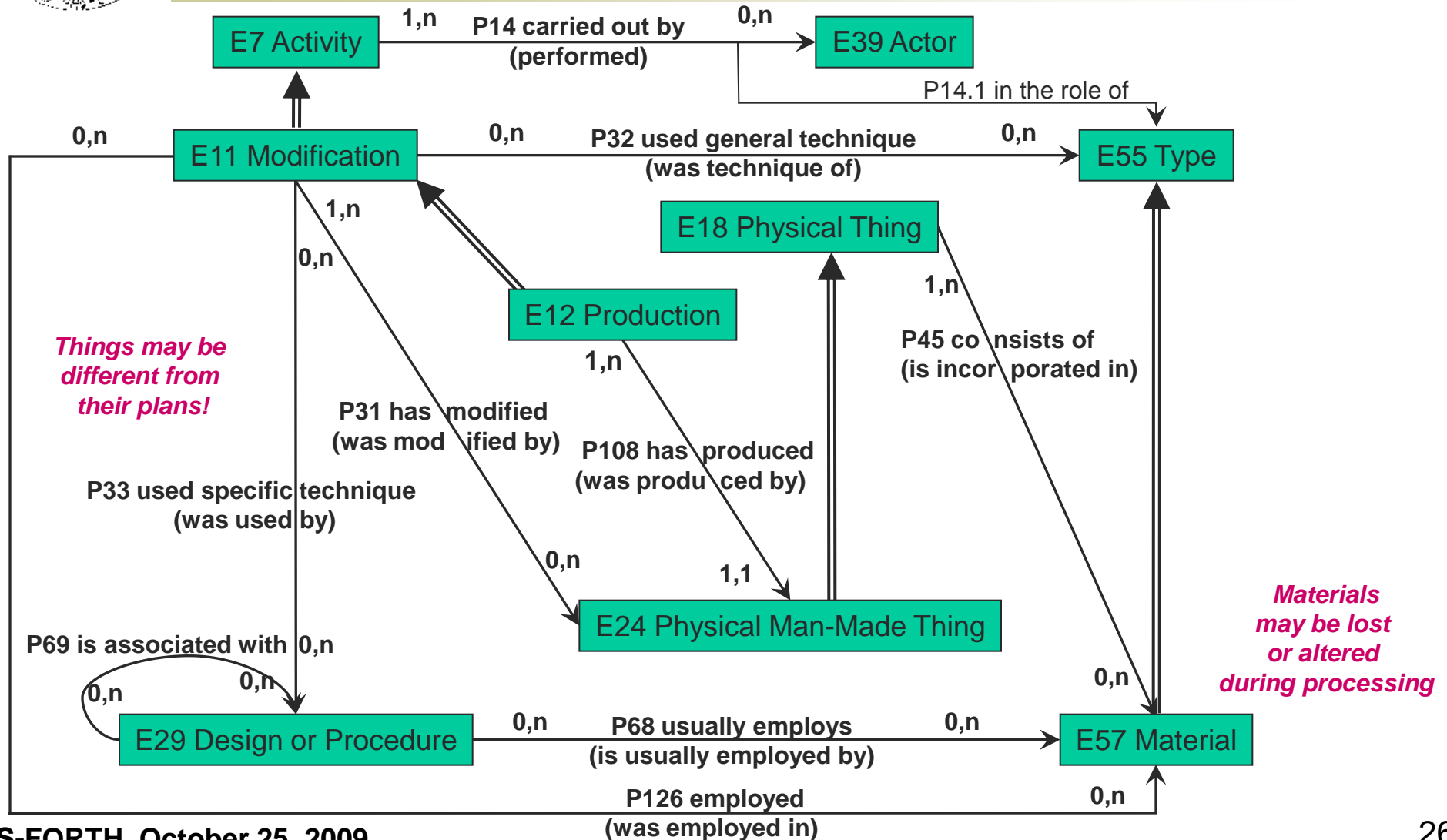
## Activities: *E9 Move*





# The CIDOC CRM:

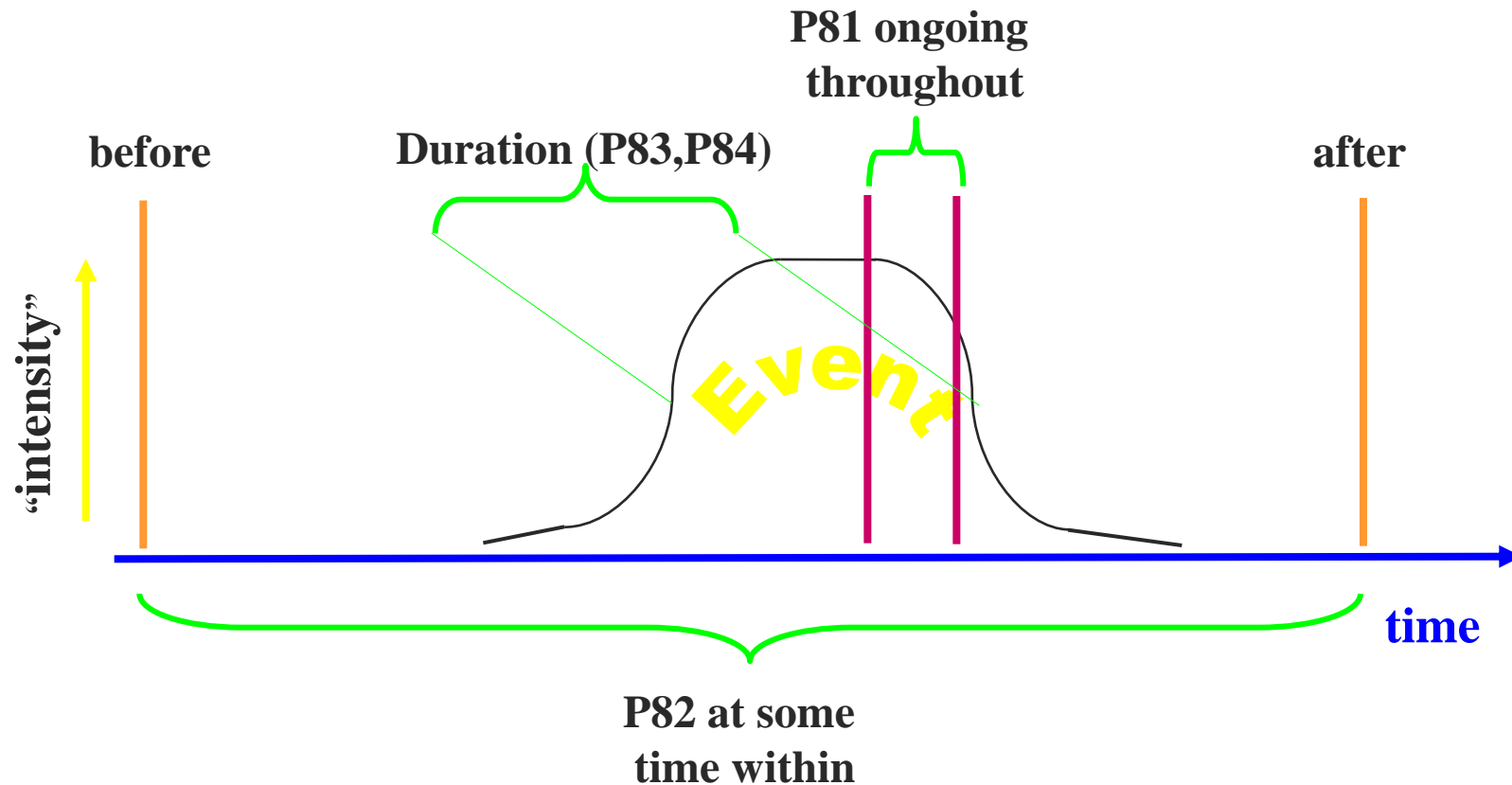
## Activities: E11 Modification/E12 Production





# The CIDOC CRM:

## *Time Uncertainty, Certainty and Duration*





# The CIDOC CRM:

## E53 Place

### E53 Place

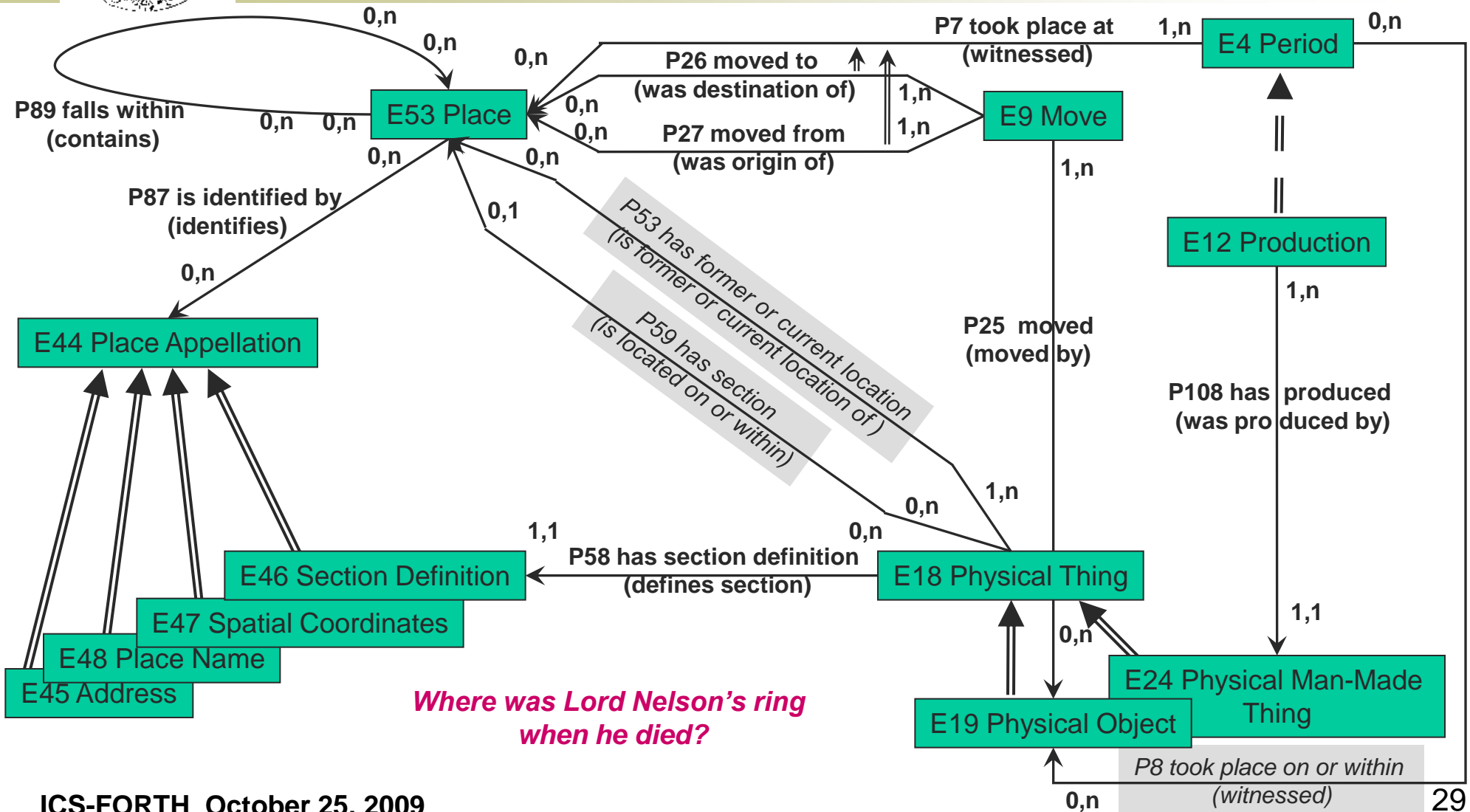
- A place is an **extent** in space, **determined diachronically** with regard to a larger, persistent constellation of matter, often continents -  
*by coordinates, geophysical features, artefacts, communities, political systems, objects*  
- but **not identical** to
- A **means** to reason about the “where” in multiple reference systems.
- Examples:
  - *the **bow** of a **ship***
  - ***African** dinosaur foot-prints in **Portugal***
  - ***where Nelson died***
- A landscape = E26 Physical Feature
- A geopolitical unit = E4 Period  
*....Has a place !*





# The CIDOC CRM:

## Properties of E53 Place







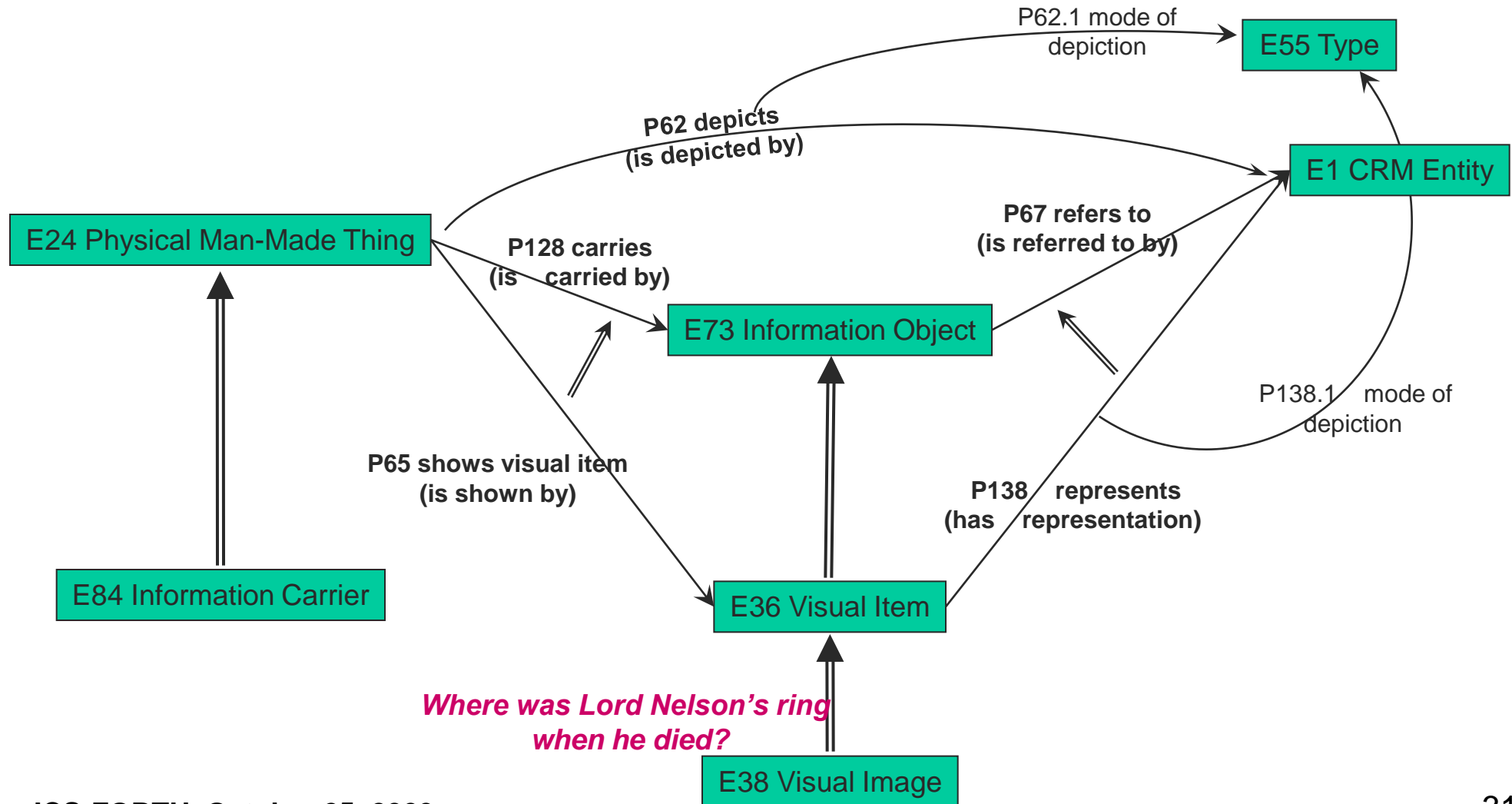
***archaeologists and biologists do the same,  
but use different language!***





# The CIDOC CRM:

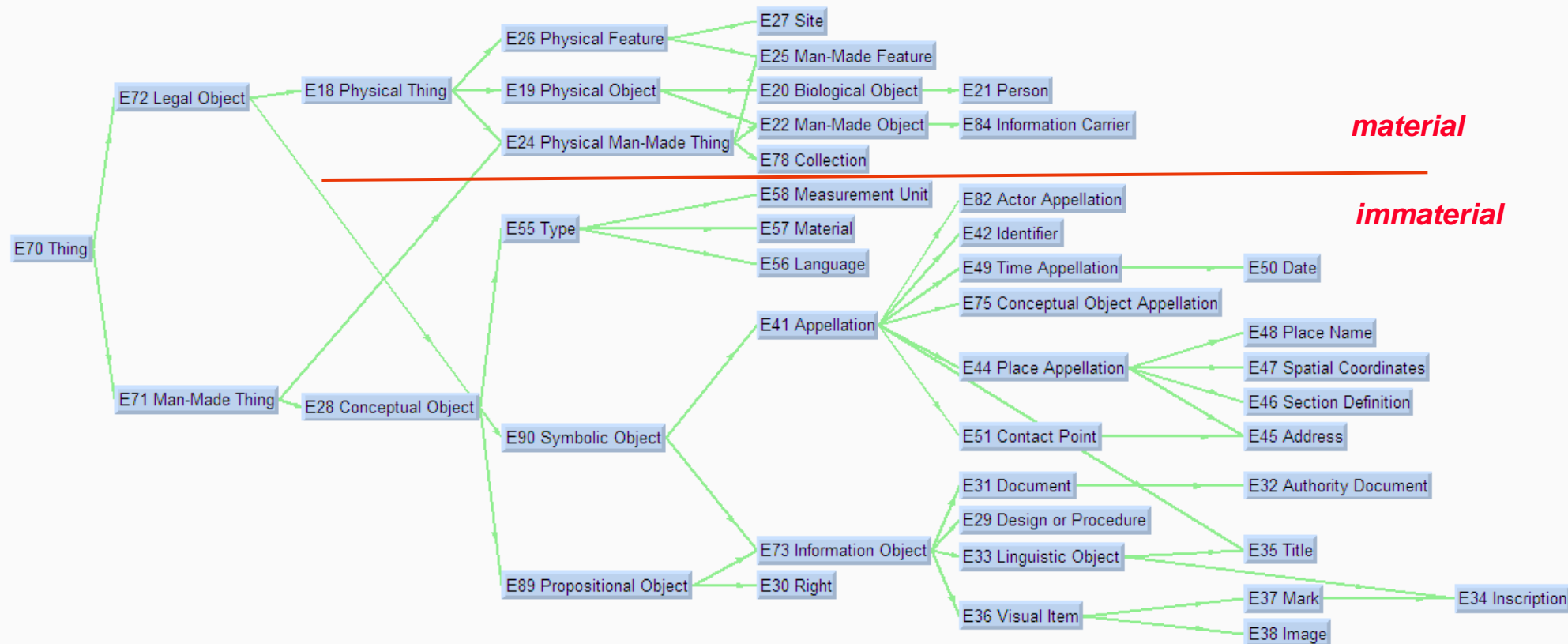
## *Visual Content and Subject*





# The CIDOC CRM:

## *E70 Thing*





# The CIDOC CRM:

## *How to use the CRM*

*It always comes to schema matching:*

- Prepare data export of a legacy of system
- New system: Make a naïve information model, match it with the CRM

### *1. Understand the CRM*

- Learn the **core classes** and **properties**
- Look up **graphics** of functional units [http://cidoc.ics.forth.gr/comprehensive\\_intro.html](http://cidoc.ics.forth.gr/comprehensive_intro.html)
- Download cross-reference manual [http://cidoc.ics.forth.gr/official\\_release\\_cidoc.html](http://cidoc.ics.forth.gr/official_release_cidoc.html)

### *2. Interpret source field by field as paths of CRM properties*

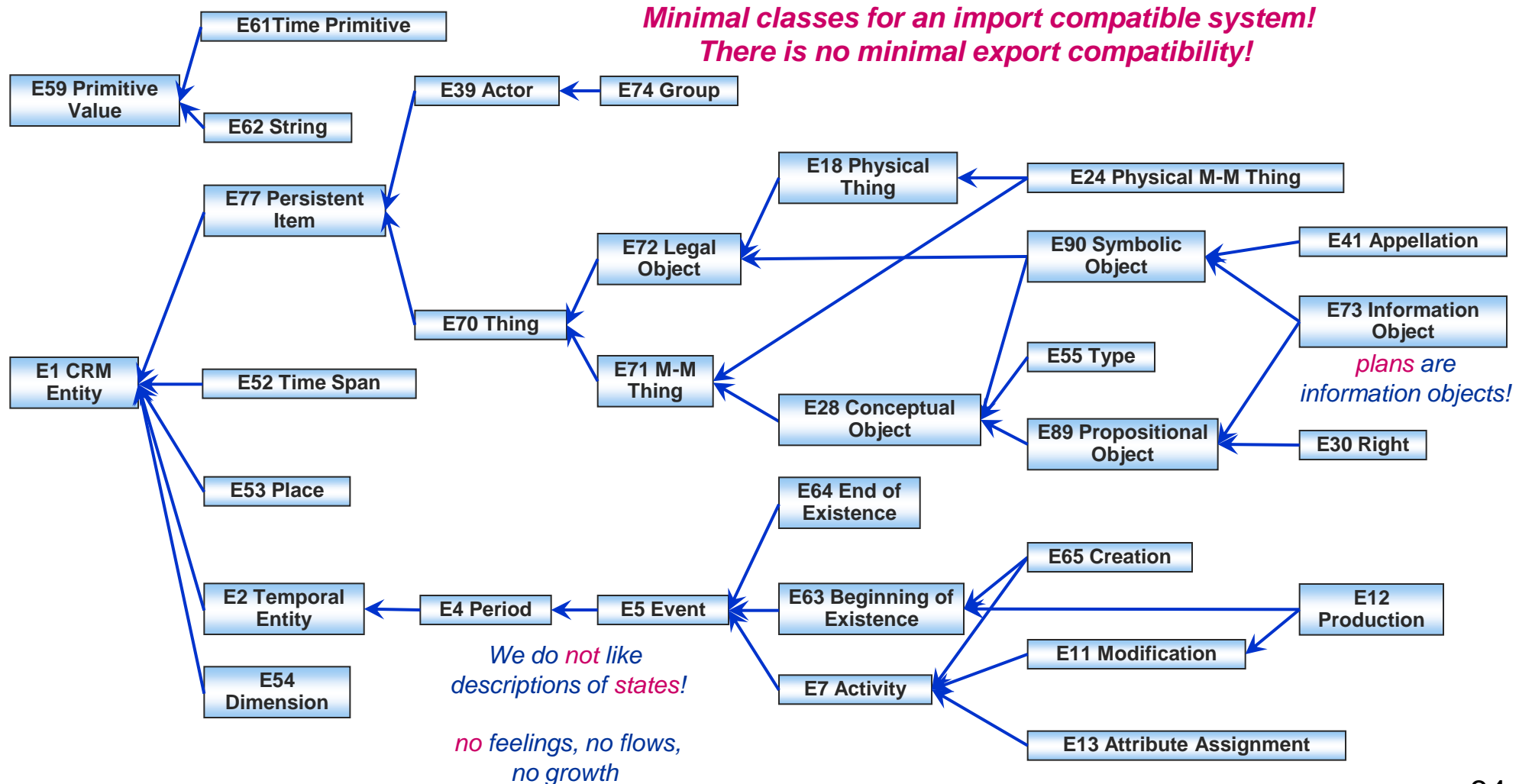
### *3. Clean up your labels and identifiers:*

- distinguish controlled, uncontrolled values, texts
- make sure all things you know can be identified



# The CIDOC CRM:

## Hierarchy of core classes





# The CIDOC CRM:

## *Example: an informal data structure*

- “AAM Recommended Procedures for Providing Information to the Public about Objects Transferred in Europe during the Nazi Era” published in May 2001
- Basis for an **Internet Portal**
- **20** data **categories** for allowing potential claimants to assess objects
- Informal (naïve): **no field** has a controlled data type, hence **cannot be** mechanically **integrated** with any other information, it is actually only a questionnaire.

*But: integrating provenance data from over the world  
is a VERY hot topic*



# The CIDOC CRM:

## *AAM Data Categories*

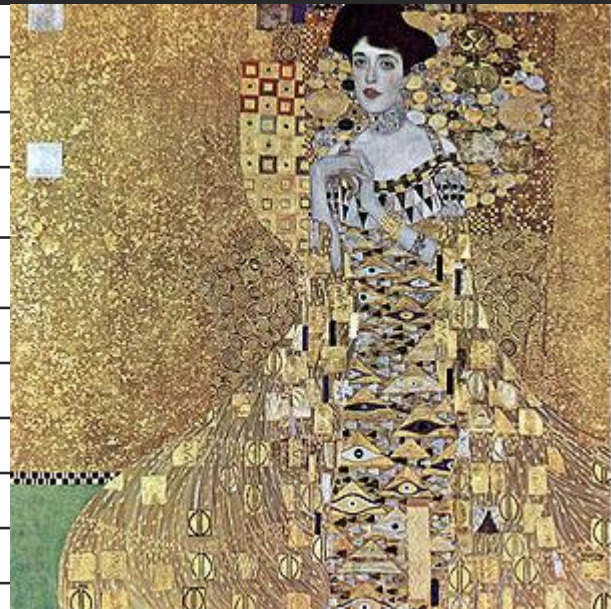
<b>Artist/Maker</b>	To include artists' names, alternate names, and previous attributions
<b>Nationality of Artist/Maker</b>	---
<b>Life Dates of Artist/Maker</b>	---
<b>Place or Culture of Object</b>	Only if artist unknown.
<b>Object Title or Name</b>	To include alternate titles.
<b>Date of Work</b>	To include approximate date, if specific date is unknown.
<b>Medium/Materials</b>	---
<b>Measurements</b>	---
<b>Date of Acquisition</b>	---
<b>Accession Number</b>	---
<b>Object Type</b>	Painting, sculpture, decorative arts, etc
<b>Subject Type</b>	Landscape, portrait, mythological subject, historical, religious, genre, etc.
<b>Signature and Marks</b> (obverse)	To include signatures, inscriptions, and marks.
<b>Labels and Marks</b> (reverse,frame, mount,..)	To describe marks and labels (prior to 1960) on the reverse of an object
<b>Description</b>	To contain description of object (its content, subject, etc.).
<b>Provenance</b>	To contain known owners,dates of ownership,places of ownership, transfer method..
<b>Exhibition History</b>	---
<b>Bibliographic History</b>	---
<b>Other Relevant Information</b>	To contain anything about the object, useful in identifying it for this purpose.
<b>Image</b>	An image is key to identifying an object.





# The CIDOC CRM:

## AAM Data Example

Artist/Maker	Klimt, Gustav	
Nationality of Artist/Maker	Austrian	
Life Dates of Artist/Maker	1862-1918	
Place or Culture of Object	Art Nouveau	
Object Title or Name	Portrait of Adele Bloch-Bauer I	
Date of Work	1907	
Medium/Materials	Oil, silver, gold	
Measurements	138X138 cm	
Date of Acquisition	1938	
Accession Number		
Object Type	Painting	
Subject Type	Portrait	
Signature and Marks (obverse)		
Labels and Marks (reverse,frame, mount,...)		
Description	...In 1903, Klimt visited the Byzantine mosaics at Ravenna, ....figure...	
Provenance	.When Nazi governments took over Austria, Bauer's property was....	
Exhibition History	In June 2006 New York's Neue Galerie is reported to have .....	
Bibliographic History	H.Czernin. <i>Die Fälschung: Der Fall Bloch-Bauer und das Werk Gustav Klimts</i> . 2006	
Other Relevant Information		
Image	<a href="http://www.neuegalerie.org/data/databases/neuegalerie_0378/widgets/neuegalerie_collection_c/00/00/00/07/full_brigit/original.jpg">http://www.neuegalerie.org/data/databases/neuegalerie_0378/widgets/neuegalerie_collection_c/00/00/00/07/full_brigit/original.jpg</a>	



# The CIDOC CRM:

## *AAM to CRM matching*

**AAM: Artist/Maker**

**CRM Path:**

**E22 Man-Made Object: Painting: Portrait of Adele Bloch-Bauer I**

**P108B was produced by**

**E12 Production:** Production of Portrait of Adele Bloch-Bauer I in 1907

**P14 carried out by**

**E21 Person: Klimt, Gustav 1862-1918**

*Possible Identifiers in RDF :*

a) E21 Person: [uuid:ee1d3ac0-b824-11de-8a39-0800200c9a66](#)  
rdfs:label **Klimt, Gustav 1862-1918**

b) E21 Person: [http://www.getty.edu/research/conducting\\_research/vocabularies/ulan/#500030531](http://www.getty.edu/research/conducting_research/vocabularies/ulan/#500030531)

c) E21 Person: [urn:aacr2:Klimt.Gustav\\_1862-1918](#) ??



# The CIDOC CRM:

## *AAM to CRM matching*

### ***AAM: Life Dates of Artist/Maker***

#### ***CRM Path:***

**E21 Person:** Klimt, Gustav 1862-1918

**P98B** was born

**E67 Birth:** Birth of Klimt, Gustav 1862-1918

**P4** has time-span

**E52 Time-Span:** July 14, 1862

**P100B** died in

**E69 Death:** Death of Klimt, Gustav 1862-1918

**P4** has time-span

**E52 Time-Span:** February 6, 1918



# The CIDOC CRM:

## *AAM to CRM matching*

### ***AAM: Nationality of Artist/Maker***

*What does that mean??*

### ***CRM Path:***

*Take as a type of manufacturer (default!):*

**E21 Person:** Klimt, Gustav 1862-1918

P2 has type

**E55 Type:** Artist  
Austrian

*Take as country of birth*

**E21 Person:** Klimt, Gustav 1862-1918

P98B was born

**E67 Birth:** Birth of Klimt, Gustav 1862-1918

P10 falls within:

**E4 Period:** State of Austria



*Same as birth date!*



# The CIDOC CRM:

## *AAM to CRM matching*

***AAM: Nationality of Artist/Maker***

***CRM Path:***

*Take as citizenship*

**E21 Person:** Klimt, Gustav 1862-1918

**P98B is current or former member of**

**E74 Group:** Austrian Nation

It is not the CRM, which is ambiguous, it is your data!



# The CIDOC CRM:

## ***AAM: Object Type***

### ***CRM Path:***

**E22 Man-Made Object:** Painting: Portrait of Adele Bloch-Bauer I

P2 has type

**E55 Type (T22):** Painting => SKOS!



# The CIDOC CRM:

## *AAM to CRM matching*

### ***AAM: Object Title or Name***

*To include alternate titles*

### ***CRM Path:***

**E22 Man-Made Object:** Painting: Portrait of Adele Bloch-Bauer I

**P102 has title**

**P102.1 has type** Primary **or** Alternate **(T35):** Primary

**E35 Title:** "Portrait of Adele Bloch-Bauer I"



# The CIDOC CRM:

## *AAM to CRM matching*

### ***AAM:Date of Work***

*Attached to the creation event!*

### ***CRM Path:***

**E12 Production:** Production of Portrait of Adele Bloch-Bauer I in 1907

**P4 has time-span**

**E52 Time-Span:** 1907





# The CIDOC CRM:

## *AAM to CRM matching*

### ***AAM:Medium/Materials***

*Assume no mention of technique*

### ***CRM Path:***

**E22 Man-Made Object:** Painting: Portrait of Adele Bloch-Bauer I

**P45 consists of**

**E57 Material:** Oil  
Silver  
Gold



# The CIDOC CRM:

## *AAM to CRM matching*

### **CRM Representation so far:**

**E22 Man-Made Object:** Painting: Portrait of Adele Bloch-Bauer I

P2 has type

E55 Type (T22): Painting

P45 consists of

E57 Material: Oil; Silver; Gold

P108B was produced by

E12 Production: Production of Portrait of Adele Bloch-Bauer I in 1907

P4 has time-span

E52 Time-Span: 1907

P14 carried out by

E21 Person: Klimt, Gustav 1862-1918

P2 has type

E55 Type: Artist; Austrian

P98B was born

E67 Birth: Birth of Klimt, Gustav 1862-1918

P4 has time-span

E52 Time-Span: July 14, 1862

P100B died in

E69 Death: Death of Klimt, Gustav 1862-1918

P4 has time-span

E52 Time-Span: February 6, 1918



# The CIDOC CRM:

## *AAM to CRM matching*

### ***AAM:Description***

### ***CRM Path:***

**E22 Man-Made Object:** Portrait of Adele Bloch-Bauer I

P3 has note

P3-1 has type Description (T62)

**E62 String:** This acquisition made available in part through the generosity of the heirs of the Estates of Ferdinand and Adele Bloch-bauer. In 1903, Klimt visited the Byzantine mosaics at Ravenna, Italy, and was impressed by these exquisite works, in which the figure of the Empress Theodora is rendered in a bejeweled setting inlaid with gold. Upon his return to Vienna, he began to work in what became known as his "Golden Style," incorporating gold elements into both his allegorical and portrait paintings. In *Adele Bloch-Bauer I*, one of Klimt's greatest achievements, gold is used in a variety of contexts, from the lustrous background to the shining fabric of...



# The CIDOC CRM:

## *AAM to CRM matching*

### ***AAM:Date of Acquisition***

*This should go to Provenance data*

### ***CRM Path:***

**E22 Man-Made Object:** Painting: Portrait of Adele Bloch-Bauer I

**P24B** changed ownership through

**E8 Acquisition:** Acquisition of Portrait of Adele Bloch-Bauer I in 2006

**P4** has time-span

**E52 Time-Span:** June 18, 2006

**P22** transferred title to

**E21 Person:** Ronald Lauder (Provider of Data)

**P23** transferred title from

**E21 Person:** Maria Altmann

**P2** has type

**E55 Type (T8):** Purchase, Sale

**P7** took place at

**E53 Place:** United States of America



# The CIDOC CRM:

## *AAM to CRM matching*

### **AAM:Provenance**

*unstructured representation*

### **CRM Path:**

**E22 Man-Made Object:** Painting: Portrait of Adele Bloch-Bauer I

**P3 has note**

**P3-1 has type:** Provenance (T62)

**E62 String:** ... .When Nazi governments took over Austria,  
Bauer's property was .....



# The CIDOC CRM:

## *AAM to CRM matching*

### **AAM:Provenance**

*Structured representation:*

#### **CRM Path:**

E22 Man-Made Object: **Painting: Portrait of Adele Bloch-Bauer I**

.....

P24B changed ownership through

E8 Acquisition: **Acquisition of Portrait of Adele Bloch-Bauer I in 1938**

P2 has type

E55 Type (T8): **Confiscation**

P21 transferred title from

E21 Person: **Husband of Adele Bloch-Bauer**

P22 transferred title to

E74 Group: **The Nazi Government**

P4 has time-span

E52 Time-Span: **1938**

P24B changed ownership through

.....



# The CIDOC CRM:

## *AAM to CRM matching*

### ***AAM:Exhibition History***

*should also be part of provenance!*

### ***CRM Path:***

**E22 Man-Made Object:** Painting: Portrait of Adele Bloch-Bauer I

**P12B** was present at

**E7 Activity:** Exhibition in Neue Galerie New York City in 2006

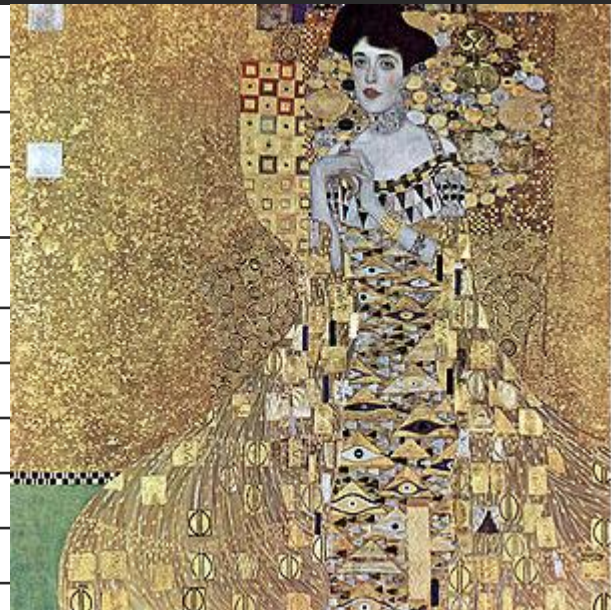
**P7** took place at

**E53 Place:** New York City/ USA



# The CIDOC CRM:

*CRM-inside: GUI can make it look as before...*

Artist/Maker	Klimt, Gustav	
Nationality of Artist/Maker	Austrian	
Life Dates of Artist/Maker	1862-1918	
Place or Culture of Object	Art Nouveau	
Object Title or Name	Portrait of Adele Bloch-Bauer I	
Date of Work	1907	
Medium/Materials	Oil, silver, gold	
Measurements	138X138 cm	
Date of Acquisition	1938	
Accession Number		
Object Type	Painting	
Subject Type	Portrait	
Signature and Marks (obverse)		
Labels and Marks (reverse, frame, mount,...)		
Description	...In 1903, Klimt visited the Byzantine mosaics at Ravenna, ....figure...	
Provenance	.When Nazi governments took over Austria, Bauer's property was....	
Exhibition History	In June 2006 New York's Neue Galerie is reported to have .....	
Bibliographic History	H.Czernin. <i>Die Fälschung: Der Fall Bloch-Bauer und das Werk Gustav Klimts</i> . 2006	
Other Relevant Information		
Image	<a href="http://www.neuegalerie.org/data/databases/neuegalerie_0378/widgets/neuegalerie_collection_c/00/00/00/07/full_brigit/original.jpg">http://www.neuegalerie.org/data/databases/neuegalerie_0378/widgets/neuegalerie_collection_c/00/00/00/07/full_brigit/original.jpg</a>	





# The CIDOC CRM:

## Lessons Learned

### Conclusions

- Modelling with the CRM teaches good practice
  - e.g., *Avoid fields with overlapping meaning: Acquisition & Provenance*
  - e.g., *Avoid ambiguous characteristics, such as “nationality”*
  - e.g., *Making data more generic: explicit events describes complex production cases*
- Compatibility with the CRM allows for machine readable identity
  - e.g., *mechanical integration of event parts (e.g., provenance and auctions!)*
- Compatibility with the CRM allows for *mechanically collecting* stories from multiple source before reading them! The CRM is adequate to cultural research.
- *Using the CRM can be very simple, even a few core concepts give much more semantics than Dublin Core and other metadata formats.*