Supporting Documentation at the Categorical Level

Martin Doerr

Center for Cultural Informatics

Institute of Computer Science
Foundation for Research and Technology - Hellas

St Petersburg,
September 2, 2003

Categorical Documentation
Outline

- Problem statement – diversity of interest
- Reality and knowledge representation
- Factual and categorical knowledge
- Metaproperties for the CRM
- Conclusion

Categorical Documentation

Problem Statement

- Current data structures are made to organize description of facts (particulars) by providing a system of classes (nodes, tables) and relationships (attributes, links).
- No difference is made between data that are particulars and those that are universals. Inheritance of properties due to instantiation or subsumption of universals appearing as data cannot be described.
- Data in manufacturing (spare parts), ethnography, natural history and others have this problem.
- Few work in knowledge representation about metamodels and their relations to simple models.
- Missing: A theory/proposal of data structures relating particulars and universals – i.e. “cross-categorical data” in a logically well-defined way.

Categorical Documentation

Problem Statement

- We may be interested in impressionistic paintings influenced by Hokusai.
- We may be interested in paintings from Monet's friends, or paintings presented together at an exhibition.
- We may track an archive to verify circumstances of scientific findings, e.g. which evidence had Linné when he determined some new species?
- We are hardly interested in a dinosaur bone's creator and subject.
- We may be interested in the kind of environment, geographic spread, kind of food of this bone's species.

Categorical Documentation

"Factual documentation"

- **E.g. Monet's "La Pié"**
  - objects unique and valuable by intrinsic properties

- such as valued art objects, aesthetic minerals, exceptional life forms, curiosities.
- Documentation focus is "Factual on the object":

  contents: form and specific references,
  context of creation to interpret the product:
  influence, social / physical factors,
  history of provenance / prove of authenticity
- Classification serves information selection by analogy, e.g. landscape painting, impressionism etc.

Categorical Documentation

"Categorical documentation"

- E.g. "Stuffed Fringilla coelebs 1965-0034"

  - objects not unique, with normal production value,

    - used as example out of specific context

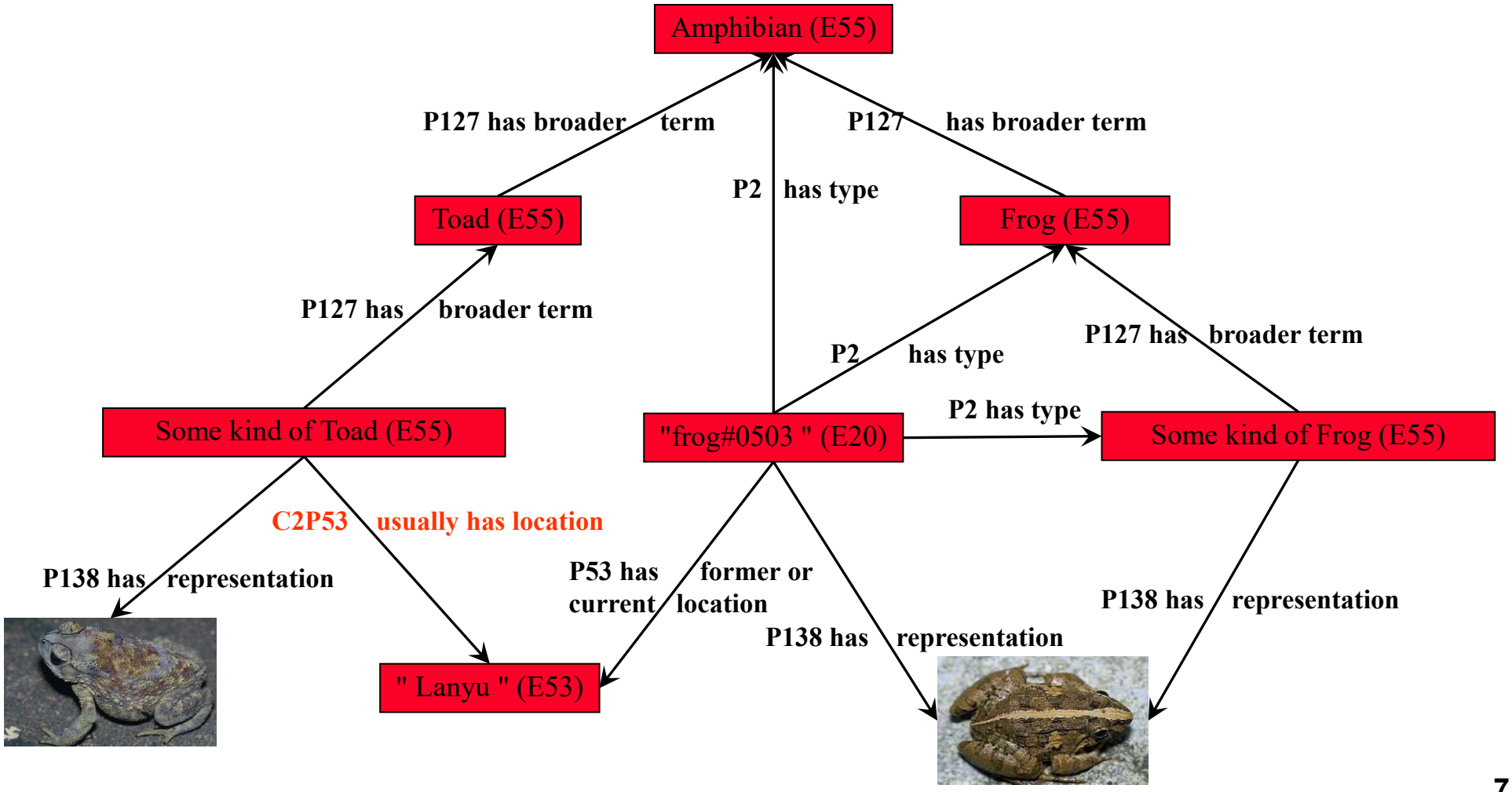
- Such as most objects in Natural History, ethnological collections, many archeological objects like frequent types of pottery etc.
- Documentation focus is "representative of its category":

  taxonomic role, deviations from prototype,

  type of context of provenance, of use;
  factual context only a statistical element for induction.
- Classification and categorical behaviour is the information, the object and its context is only an attribute.

Categorical Documentation

Biological Example (from ChiNan University)



Categorical Documentation

Ethnological Example (ChiNan University)

Technology Crop (E55)

P138 has representation

P32 used general technique

"Modif.-crop#7901" (E11)

"crop#7901" (E22)

P62 depicts

P108 has produced

P14B performed

"Da-Wu Tribe" (E74)

P14B performed

P74b is current residence of

P138 has representation

"Lanyu" (E53)

P108 has produced

"Prod.-Pottery#035" (E12)

"Pottery#035" (E22)

P62 depicts

P70 is documented in

"Map" (E31)

P138 has representation

P32 used general technique

Technology Tool-Pottery (E55)

P138 has representation

Categorical Documentation

Instantiation levels

Look at three kinds of knowledge elements:

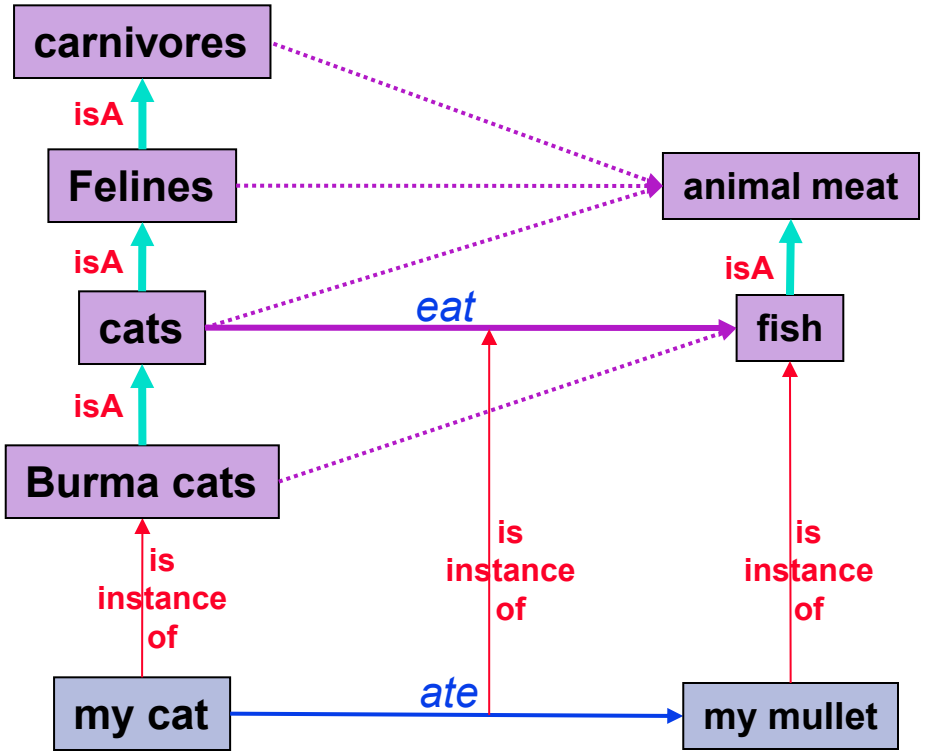
- factual: “My cat – ate – my mullet” = item- relation – item
- cross-categorical: “My cat –ate - fish” = item- relation – class
- categorical: “cats-eat-fish” = class-relationship-class

- Interpretation of factual statement is unique:

  “the predicate: ate(my cat, my mullet) holds”.

- Interpretation of categorical and cross-categorical statements is not unique.

Categorical Documentation
About categorical knowledge



There exist felines/carnivores that eat kinds of animal meat. Fish may be eaten by carnivores

= *All cats eat any fish*

All Birma cats eat any fish

Categorical Documentation

Interpretation of categorical relationships

- Interpretation of categorical relationships is not unique:

1. eat (Cat,Fish) $\Leftrightarrow$ $\exists$ x:Cat,y:Fish ($\Diamond$ ate(x,y)) = some cats can eat some fish
2. eat (Cat,Fish) $\Leftrightarrow$ $\exists$ x:Cat,y:Fish (ate(x,y)) = some cats have eaten some fish
3. eat (Cat,Fish) $\Leftrightarrow$ $\forall$ x:Cat $\exists$ y:Fish ($\Diamond$ ate(x,y)) = all cats can eat some fish
4. eat (Cat,Fish) $\Leftrightarrow$ $\forall$ x:Cat,$\forall$ y:Fish ($\Diamond$ ate(x,y)) = all cats can eat all fish
5. eat (Cat,Fish) $\Leftrightarrow$ $\forall$ x:Cat (ate(x,y) $\Rightarrow$ Fish(y)) = all cats can eat only fish

* * * * * *

7. card {x: Cat(x) $\wedge$ $\exists$ y $\wedge$ Fish(y) $\wedge$ ate(x,y)} / card {x: Cat(x)} > 0.1 = more than 10% of all cats have eaten some fish.

- Case 1. is the normal meaning of a relationship in a schema and most generic, but normally too week. Frequently, we want to register a typical behaviour, more like case 7.

- We propose : Cat “usually eats” Fish, or Cat “typically eats” Fish.

Categorical Documentation

Interpretation of cross-categorical rel.

- Interpretation of cross-categorical statements is not unique:
  1. eat kind (mycat,Fish) $\Leftrightarrow \exists$ y:Fish (ate(mycat,y)) = mycat has eaten some fish
  2. eat kind (mycat,Fish) $\Leftrightarrow \forall$ y:Fish (ate(mycat,y)) = mycat has eaten all fish

  * * * * * *

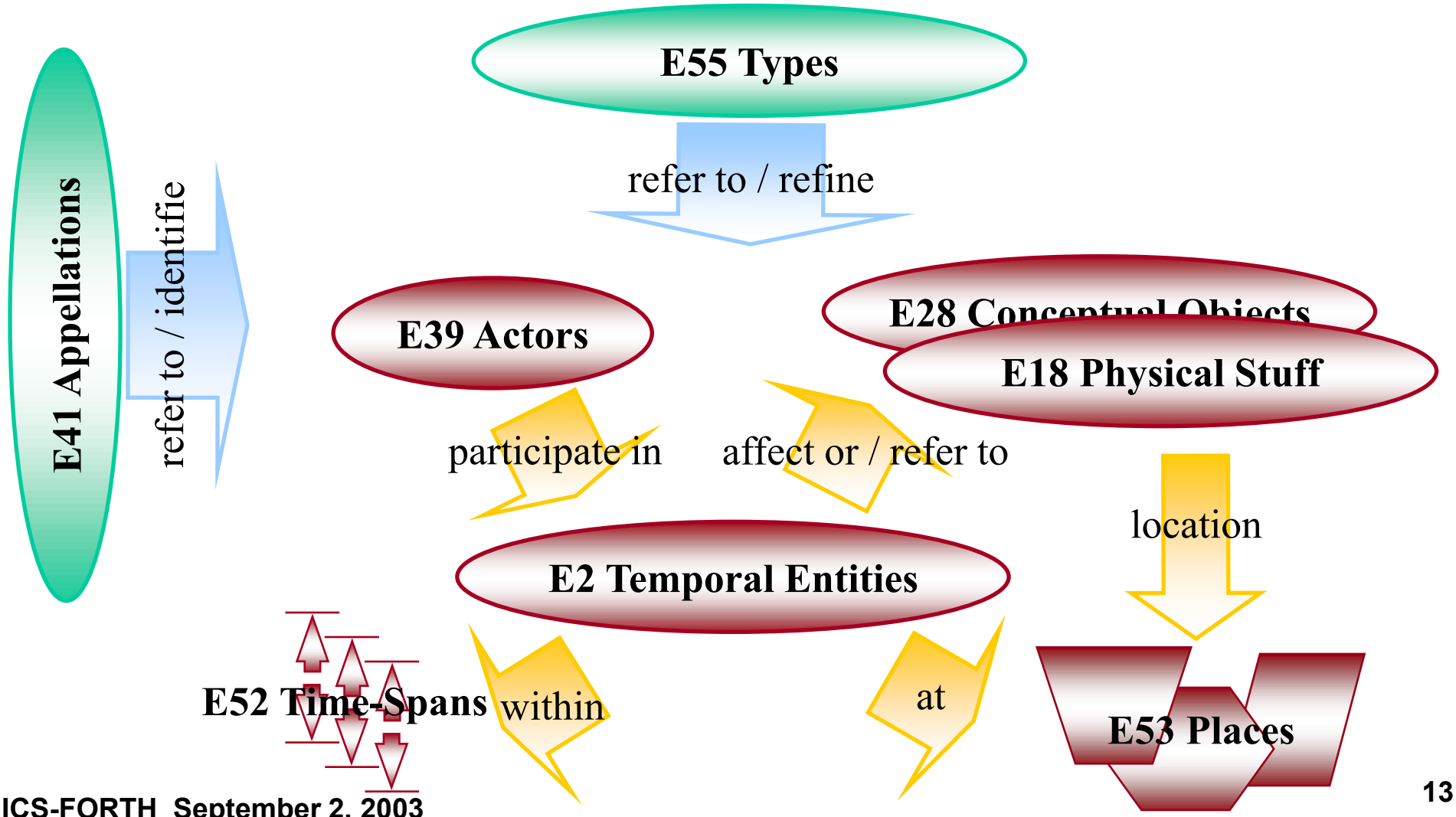
  3. card {y: Fish(y) $\wedge$ ate(mycat,y)} / card {y: ate(mycat,y)} > 0.1 = more than 10% of my cat's food is fish.

- Case 1. is the normal case to denote incomplete knowledge. Case 2. is usual for constraints, such as "Jaguars live in America".

- We propose case 1 as standard meaning. Case 2. is useful to define typologies

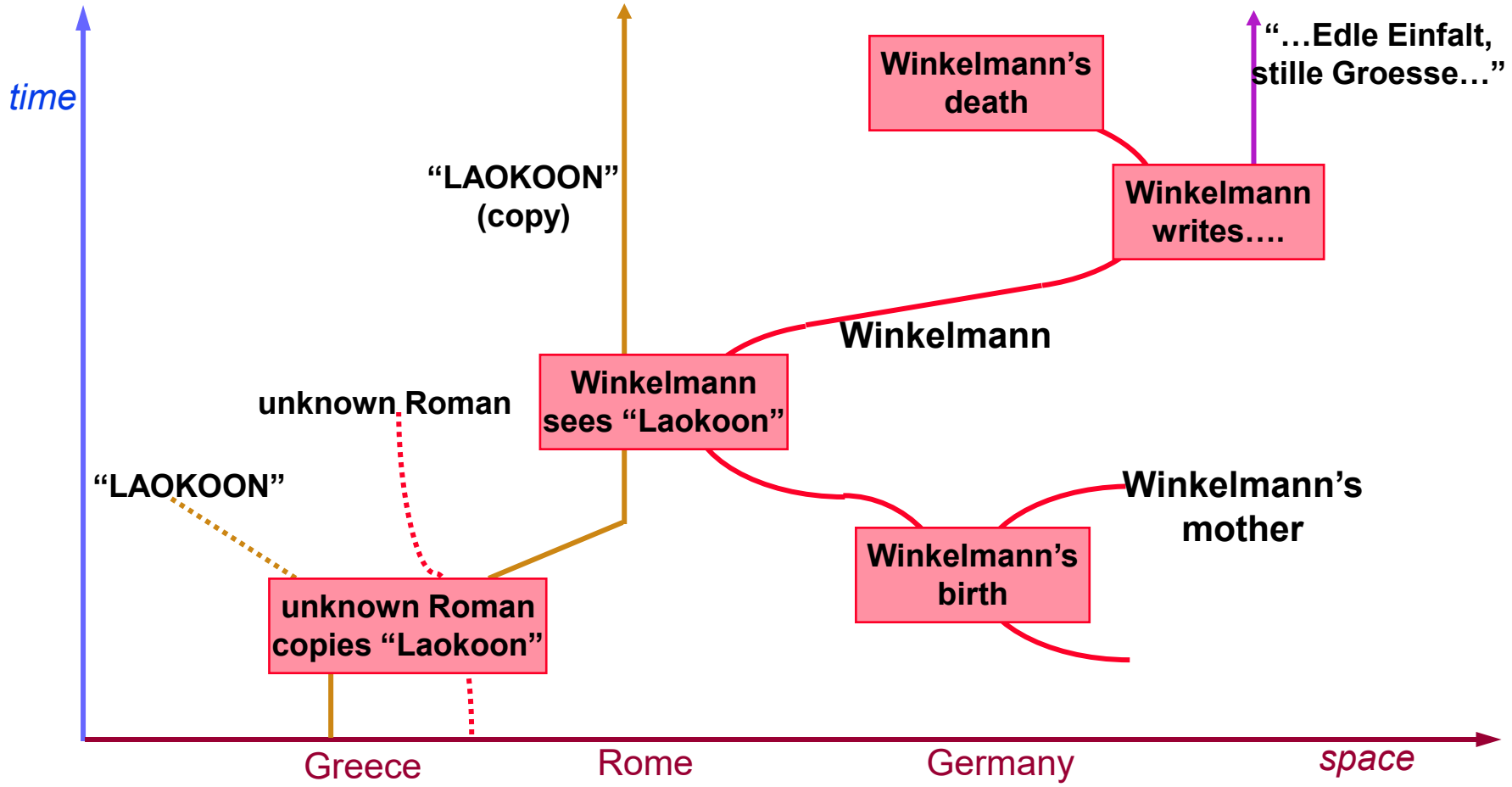
Categorical Documentation

The CIDOC CRM (ISO/CD21127) Top-Level



Categorical Documentation

Factual context as a network of "meetings"



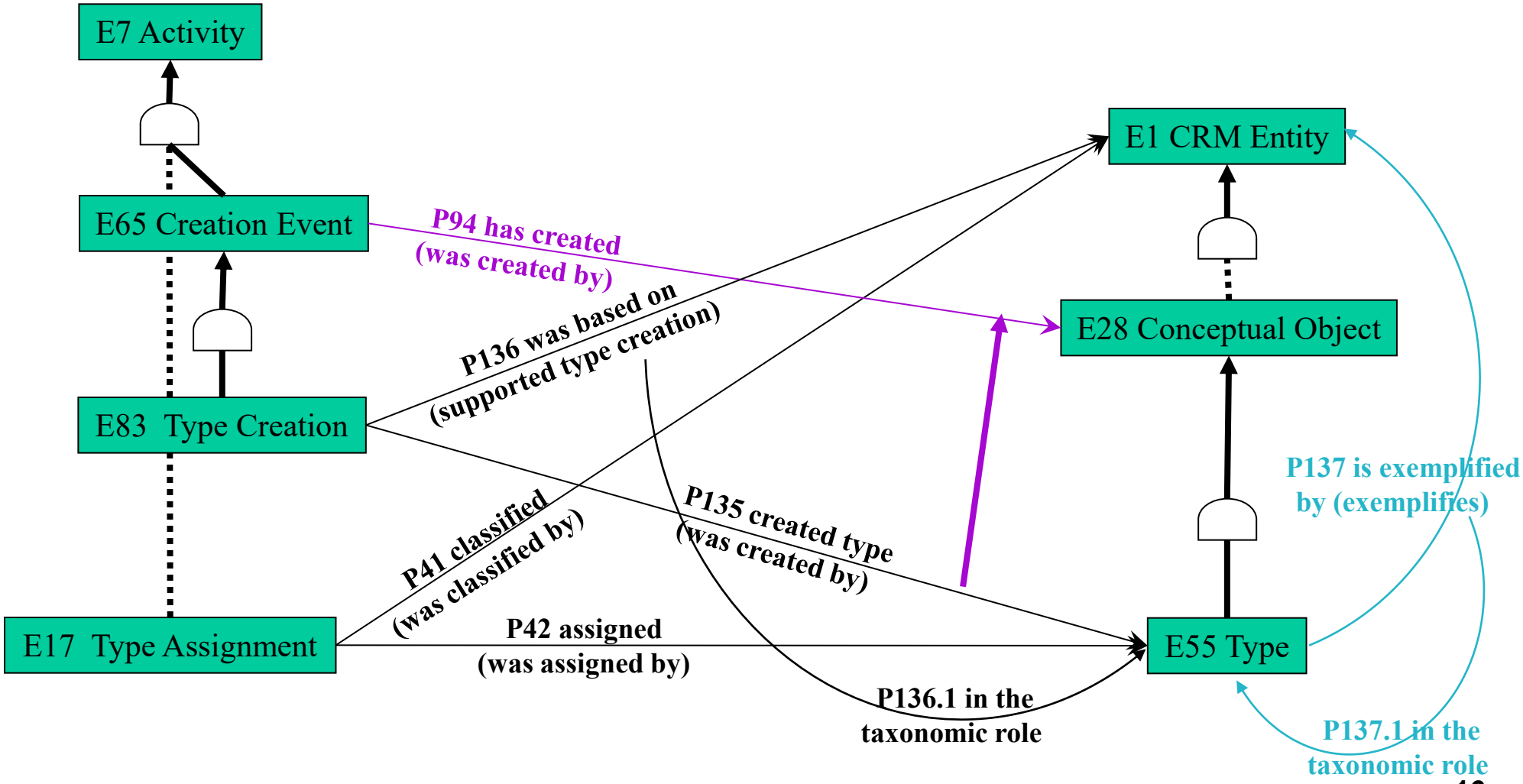
Categorical Documentation

Categorical relationships in the CRM

- *E55 Type* represents a metaclass. All CRM classes can be regarded as instances of *E55 Type*. The property *P2 has type* means *instance-of.*
- *E55 Type* is related by *P127 has broader term,* meaning *IsA*.
- Important cross-categorical relationships are defined, such as: *P125 used object of type,* together with the respective factual one: *P16 used specific object.*
- *E55 Type* is also treated as simple class in the sense of a product of the human mind.
- *There are no other categorical relationships*

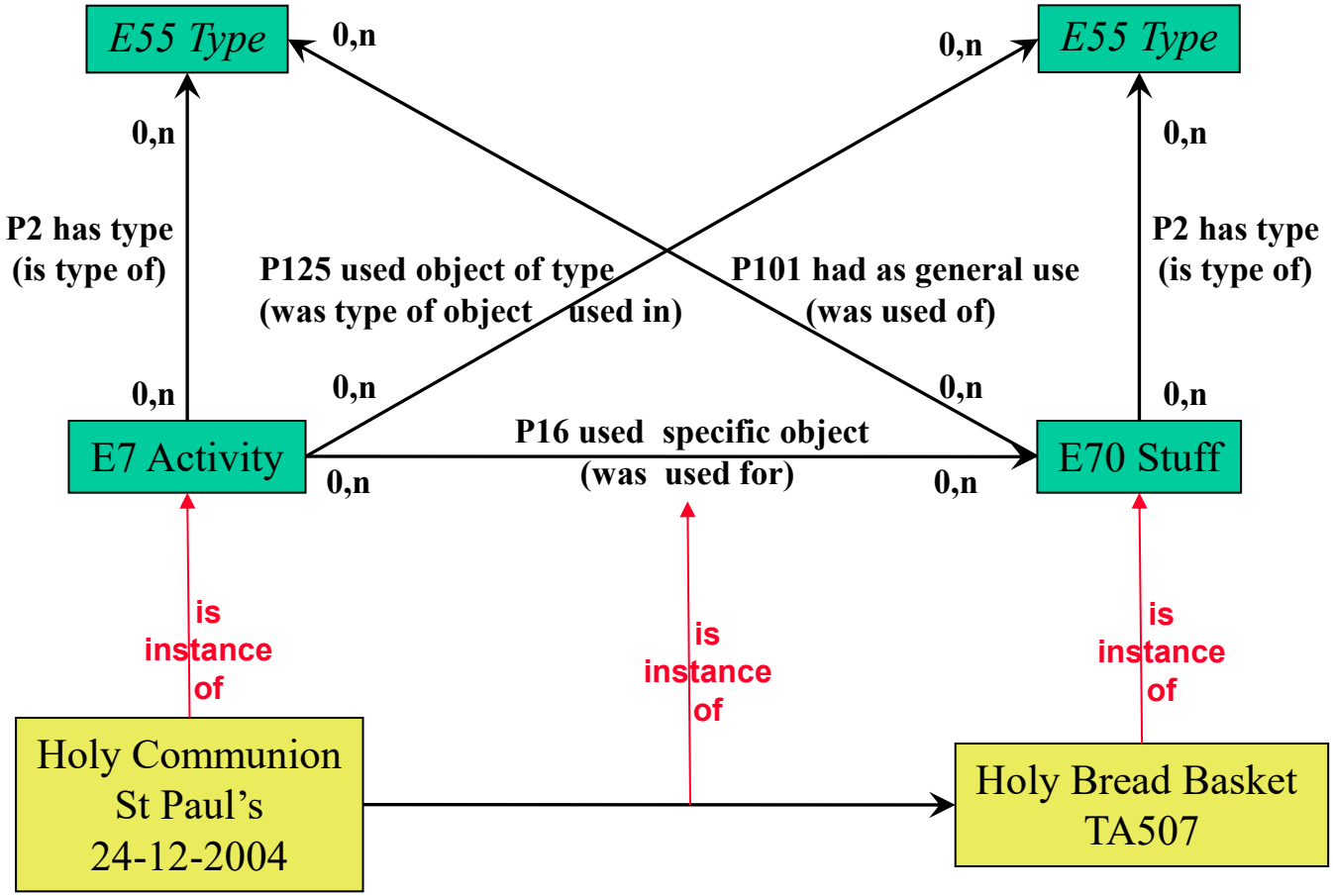
Categorical Documentation

Taxonomic discourse in the CIDOC CRM



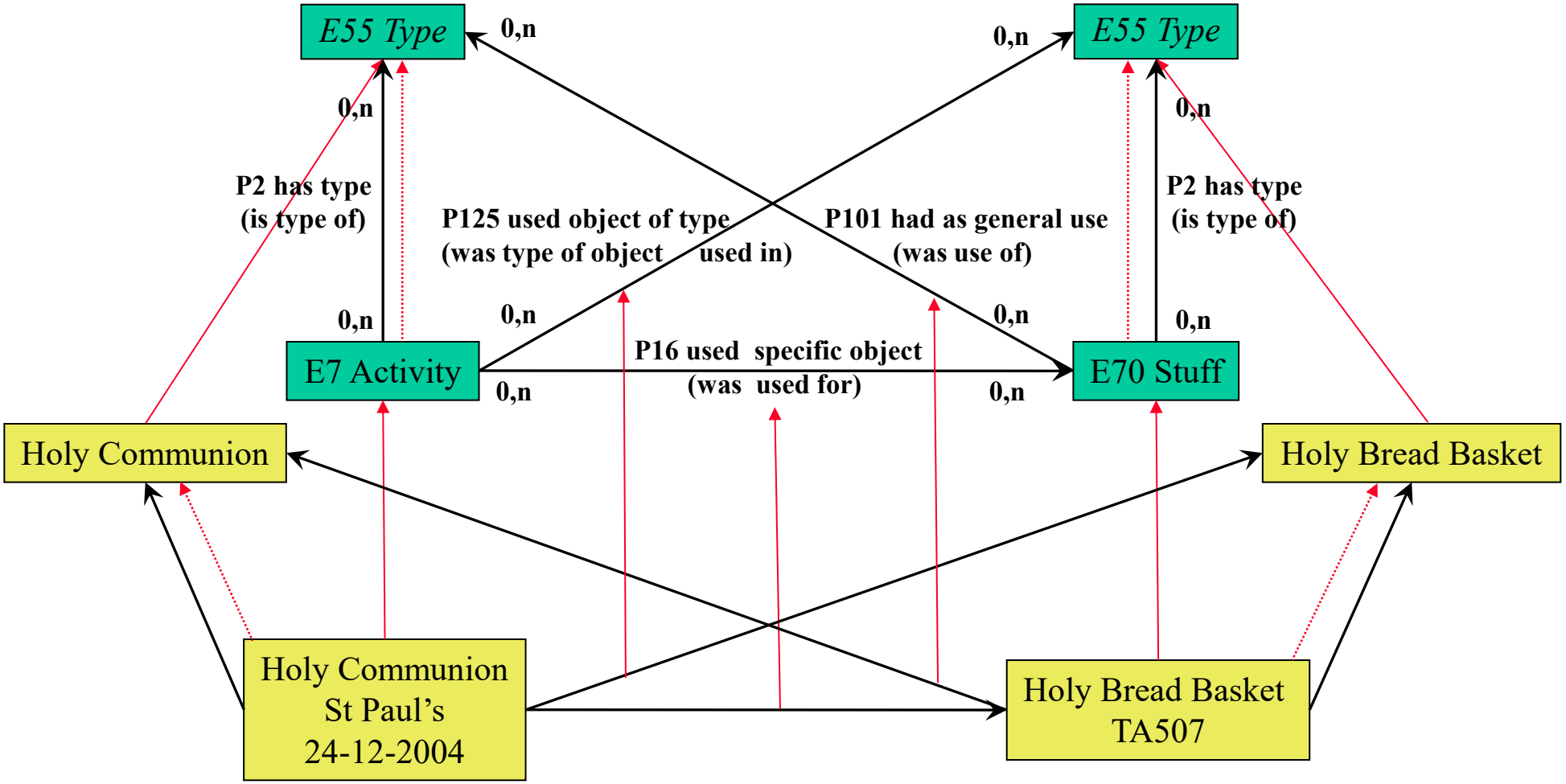
Categorical Documentation

CRM cross-categorial relationships



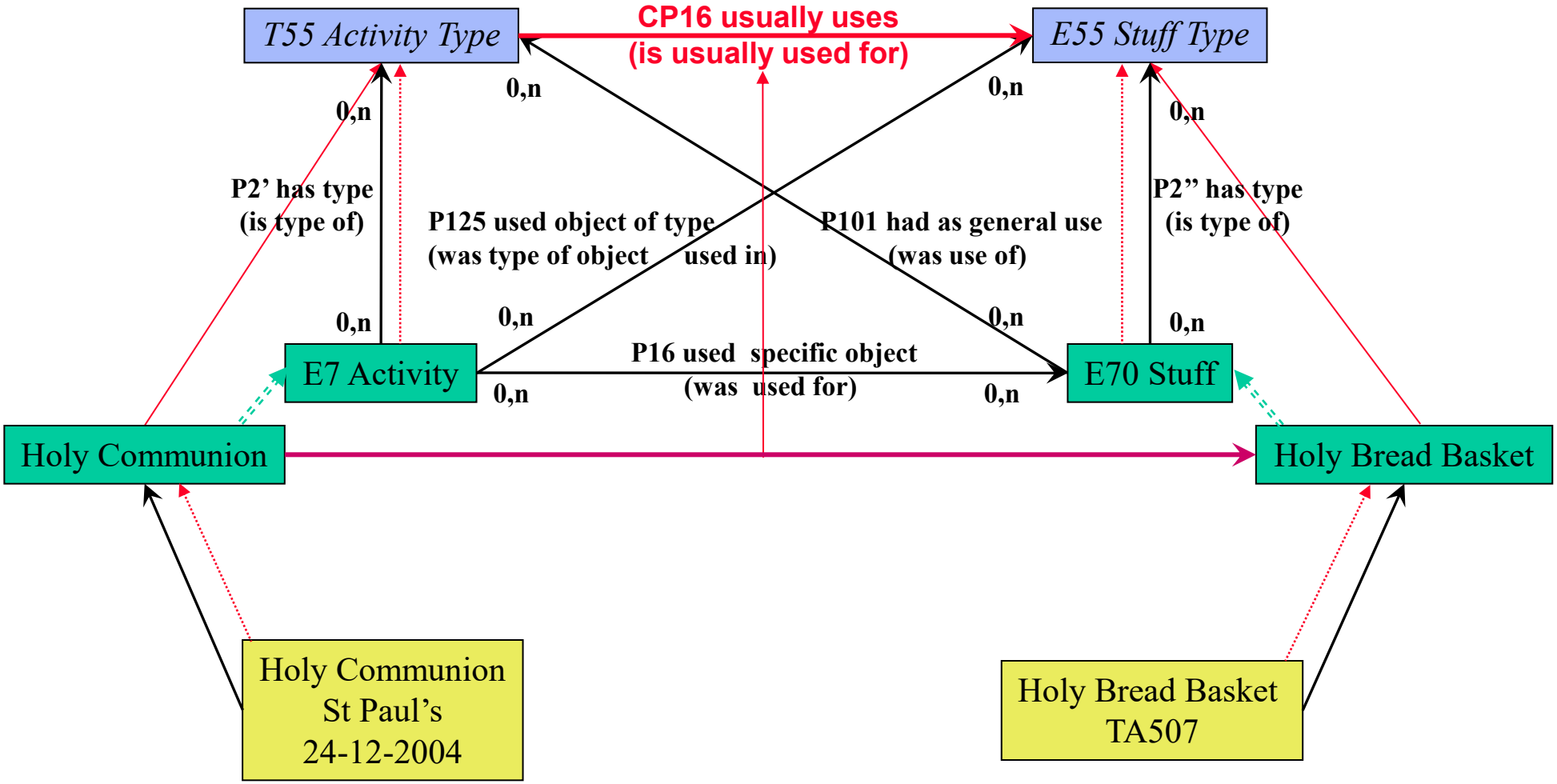
Categorical Documentation

CRM cross-categorial relationships



Categorical Documentation

New categorical relationships for the CRM



Categorical Documentation

The "MetaCRM" proposals

- There seems to be no need for new, more specialized "property verbs". Therefore we propose to generate the categorical and missing cross-categorical properties formally.
- The meaning "usually" seems to be the most important.
- cross-categorical properties should normally mean "there exists an x, instance of C, such that"
- We think of linguistic rules to create appropriate labels:
  - Activity - P16 used object - Stuff
  - Activity - C1P16 used object of type - Stuff Type
  - Activity Type – C2P16 was use of object – Stuff
  - Activity Type – C3P16 usually uses object of type – Stuff Type

Categorical Documentation

Conclusions

- A “MetaCRM” seems to be appropriate to capture typical categorical statements in museum documentation
- It can be dealt with as application of the CRM not affecting the standard
- The number of relationships becomes huge (about 400). Ideas of practical implementation must be discussed, as well as of new “short-cuts”.
- The notion of “collective events”, such as “Dai Wu pottery making” needs clarification. Is it a particular or a universal? If the latter, do we need to document it as universal?

"People Dress and Shift" (E55)

P138 has representation

P125 used object of type

P131 is identified by

"Da-Wu Tribe" (E74)

"People-T-Shaped" (E55)

P138 has representation

P14B performed

"some Owner Name" (E74)

P125 used object of type

E7 Activity

P125 used object of type

"People-Ornament" (E55)

P138 has representation

P125 used object of type

P2 has type

P2 has type

P2 has type

"People-Hat" (E55)

P138 has representation

"Fishing" (E55)

"Wearing" (E55)

"Graze" (E55)

P125 used object of type

P138 has representation

P138 has representation

"People-Others" (E55)

P138 has representation