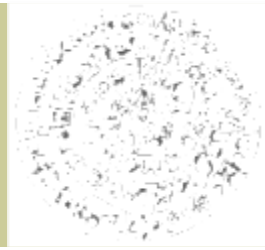


Reasoning based on property propagation on CIDOC-CRM and CRMdig based repositories

K. Tzompanaki, M. Doerr, M. Theodoridou, I. Fundulaki

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

TPDL 2013
Practical Experiences with CIDOC CRM and its Extensions

CRMEX2013, Valletta, Malta, September 26, 2013

Outline

- The problem
- Related work
- Our approach
- Examples
- Conclusion

The problem

Semantic repositories follow the "Open World Assumption"

- *incomplete knowledge*
- *metadata may be created by different people who state different facts about the same artifact*
- *use the schema in different albeit correct ways*
- *users cannot know precisely what has been documented and how.*

Semantic repositories contain implicit knowledge:

- *characteristics of artifacts that have been recorded somewhere in the semantic network but are not directly associated to the object of interest*
- *characteristics that have multiple modeling alternatives.*
- *characteristics that are generalizations of sets of more specific properties.*

Related work on data provenance

Main areas of research:

- *Workflow systems – coarse-grained provenance*
- *Database systems – fine-grained provenance*

deal mainly with:

- *Data quality & reliability*
- *Audit trail*
- *Replication recipes*
- *Attribution*

Workflow systems

Provenance of a workflow: *the entire history of the derivation of the process result*

- different versions of the software and the hardware used
- the agents that were involved in the workflow chain (processes, human agents)
- the "things" (e.g. data) employed by the processes.

Workflow provenance models:

+ *Querying allows to explore and better understand results and enables knowledge re-use*

− *computational processes only on digital artifacts*

OPM, Provenir Ontology

− *no notion of acquisition (measurement, observation), place*

− *Confuse agentive role with substance of actors, machines, S/W, context*

− *No notion of temporal indeterminacy*

W3C Recommendation Provenance Ontology (PROV-O)

− *precondition: No use of a larger reference ontology*

Data provenance *provides a detailed trace of how a piece of data has been obtained from a transformation process (i.e. query)*

Data provenance may indicate:

- *the tuples involved in the computation of a result tuple (why-provenance)*
- *where these tuples reside (where-provenance)*
- *the query operators used to obtain the result tuple (how-provenance)*

+ *extensively studied for relational databases*

– *only recently for Linked Data*

Our approach

Need of a mechanism to reason upon complex structured metadata

British Museum website:

the object "Horsemen from the west frieze of the Parthenon" is part of the Parthenon

Acropolis Museum: *Parthenon was created by Pheidias.*

Integrated metadata described in CIDOC-CRM:

- *"Horsemen from the west frieze of the Parthenon"* ***crm: forms part of*** *"Parthenon"*
- *"Parthenon"* ***crm: was produced by*** *"Construction of Parthenon"* ***crm: carried out by*** *"Pheidias"*

- *There is a likelihood that Pheidias was involved in the making of the Horsemen*
- *Inferences will increase recall and reduce precision*
- *Combined with application dependent relevance criteria can improve query results*

Our approach

***The Idea**:*

Exploit the transitive part-of and derivation chains and reason combining metadata of real world objects with metadata of digital objects and data acquisition/measurement processes

- *Generalization and inferring of metadata from related objects by using the propagation of some object properties along the transitive part-of and derivation chains of information.*
- *Simple yet powerful reasoning mechanism based on inference and completion of metadata.*

In order to:

- *help scientists query a semantic repository in order to trace and understand the source of their results*
- *reproduce results*
- *ease quality control of results and processes*

Application environment

Semantic repository based on:

- CIDOC CRM v5.0.4
  http://www.cidoc-crm.org/rdfs/cidoc-crm

- CRMdig 3.0
  extension of CIDOC CRM appropriate for representing provenance
  http://www.ics.forth.gr/isl/CRMext/CRMdig.rdfs

Used platform:

OpenRDF-Sesame http://www.openrdf.org/

Ontotext BigOWLIM reasoner (now OWLIM-SE http://www.ontotext.com/owlim/editions)

Application environment

Applied in 3D-COFORM (IP FP7/2007-2013, no 231809, http://www.3d-coform.eu/)

The semantic repository (1M RDF triples) includes:

- metadata describing the digital provenance for empirical 3D modeling and digitization processes
- metadata about the physical objects
- metadata about the digitized objects
- deep chains of events connected by input-output
- up to tens of thousands of intermediary products that "inherit" many properties along the processing chains

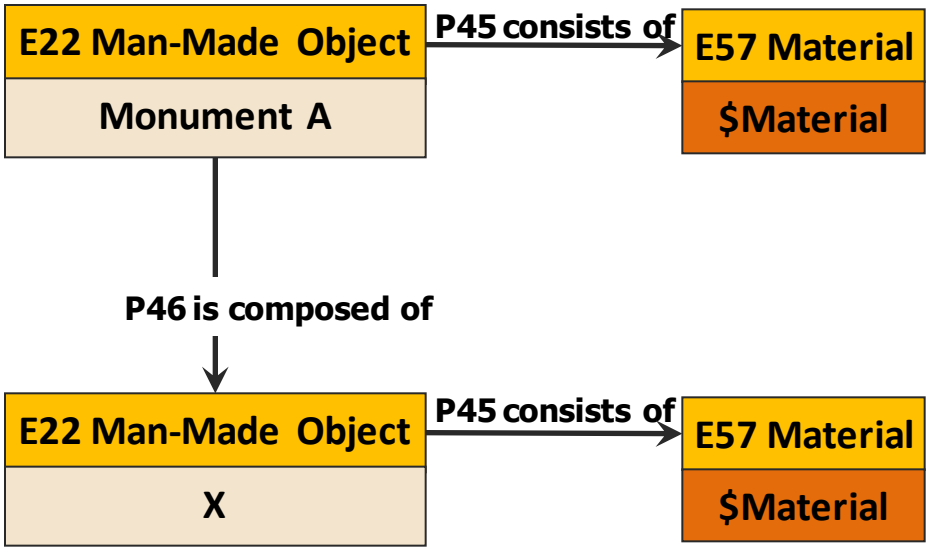
The Research Space project has also implemented this approach following our model.

Part-of chain

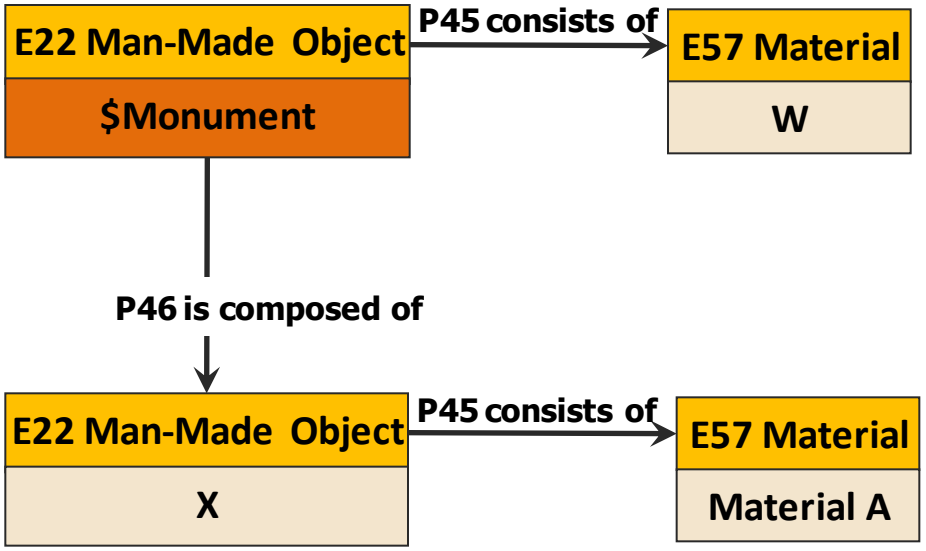
Rule 1

The property of an object is the aggregation of the explicitly defined property in the object itself and the respective properties of all its subparts.

Part-of chain



1. Forward traversal of the part of chain: Find the Material of Monument A



2. Backwards traversal of the part of chain: Find Monuments constructed from Material A

Queen Victoria Example

E22 Man-Made Object
Queen Victoria

P2 has type

E55 Type
Statue

P46 is composed of

E22 Man-Made Object
Stepped (2) base

E22 Man-Made Object
Base of plinth

E22 Man-Made Object
Plinth

E22 Man-Made Object
Reliefs

P45 consists of

E57 Material
Grey granite

P45 consists of

E57 Material
Grey marble

P45 consists of

E57 Material
Bronze

Queen Victoria Example

***Forward traversal query**:*

Find the material(s) from which the Queen Victoria Statue is made of

Answer set:

Grey granite, Grey marble, Bronze

***Backward traversal query**:*

Find statues made of Bronze

Answer set:

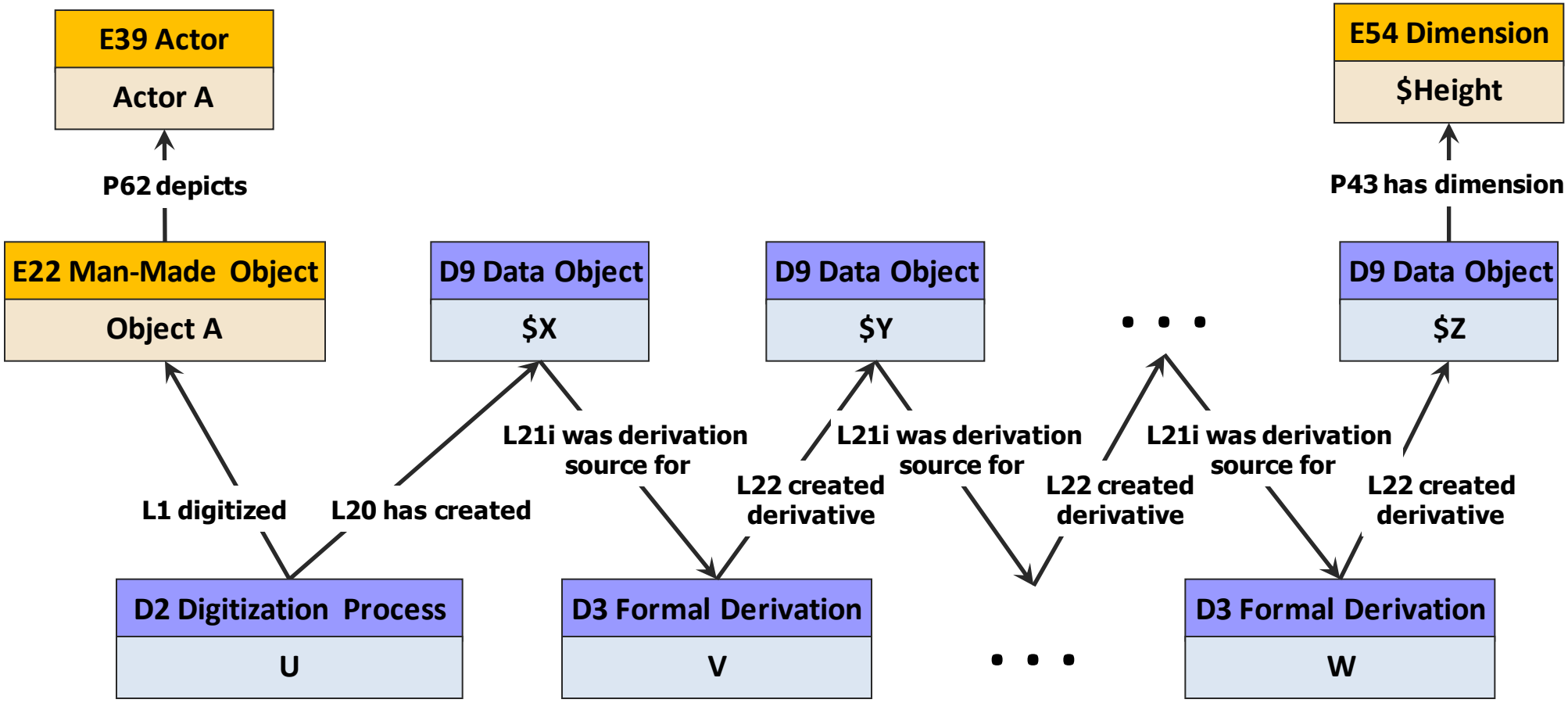
Queen Victoria

Derivation chain

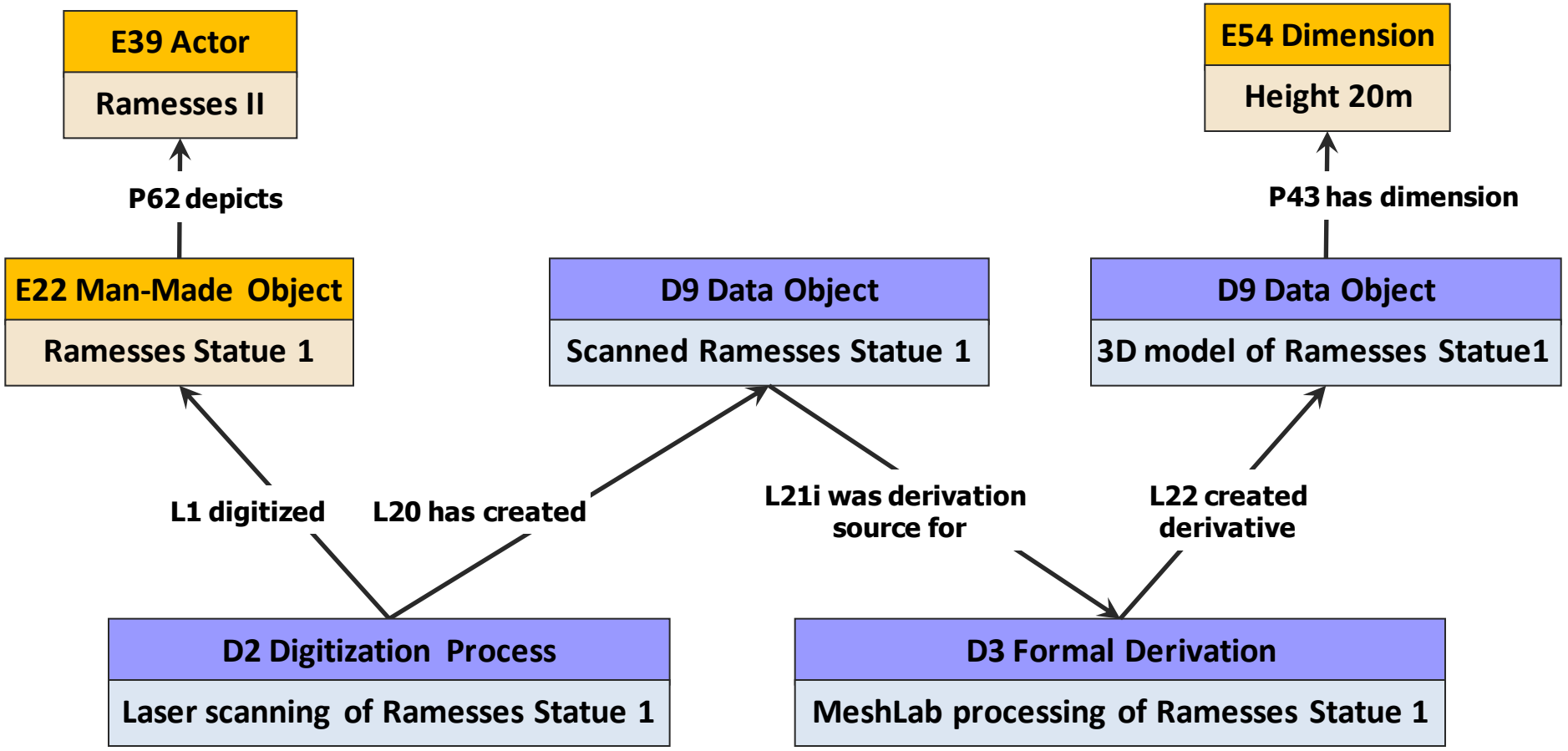
Rule 2

Physical objects may share properties with their digital representations and their derivatives.

Derivation chain



Ramesses Statue Example



Ramesses Statue Example

***Forward traversal query**:*

Find objects that depict Ramesses II

Answer set:

Ramesses Statue 1, Scanned Ramesses Statue 1, 3D model of Ramesses Statue1

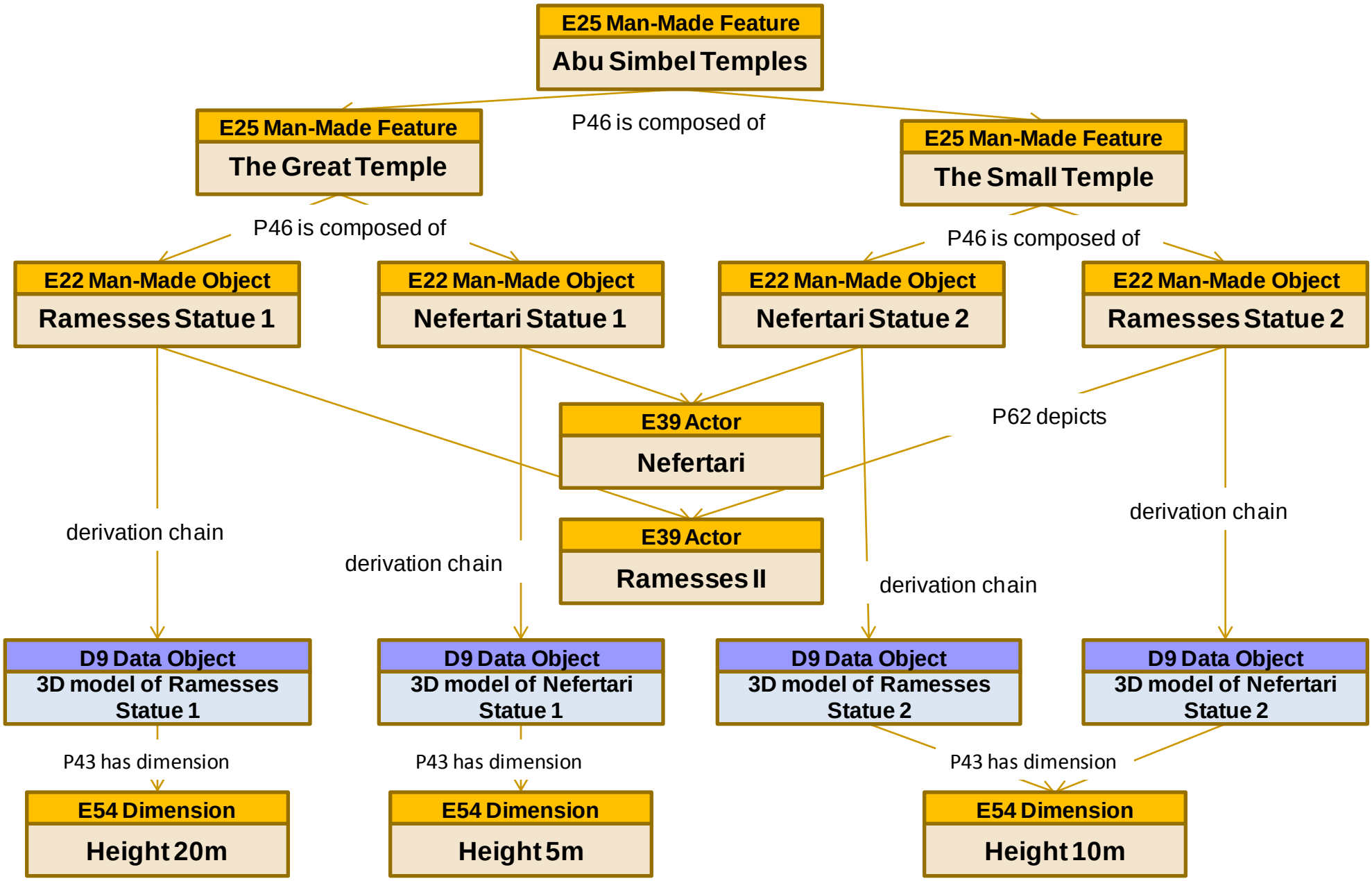
***Backward traversal query**:*

Find the size of the Ramesses Statue 1 object

Answer set:

Height 20m

Combining Part-of & Derivation chains



Abu Simbel Example

Query:

Find Temples where Ramesses II and his wife Nefertari are represented with same size statues

Answer set:

Abu Simbel Temples, The Small Temple

Example use cases

- *maintenance of repositories of digitization products*
- *garbage collection on reproducible intermediate files*
- *trace dependencies of products on tools and algorithms that should not become obsolete for long time preservation*
- *(re)production of valid, complete metadata at a loss of intermediate files*
- *completion of metadata by implicit knowledge, when production chains comprise thousands of intermediates and dozens of final products without need to manage this redundancy in the repository explicitly.*

Conclusions

Exploiting property propagation along transitive derivation and part-of chains helps:

- *to combine metadata of physical objects with metadata of digital objects*
- *to complete metadata*
- *to derive useful inferences*
- *to improve query recall*

Simple yet powerful

Open issue: Precision

- *resolve relevance issues with additional application specific constraints*

Thank you!