

Semantic technologies for historical collections:

A case study from the Sloane Lab Knowledge Base

Dr Andreas Vlachidis and Dr Daniele Metilli

UCL Department of Information Studies

a.vlachidis@ucl.ac.uk

d.metilli@ucl.ac.uk

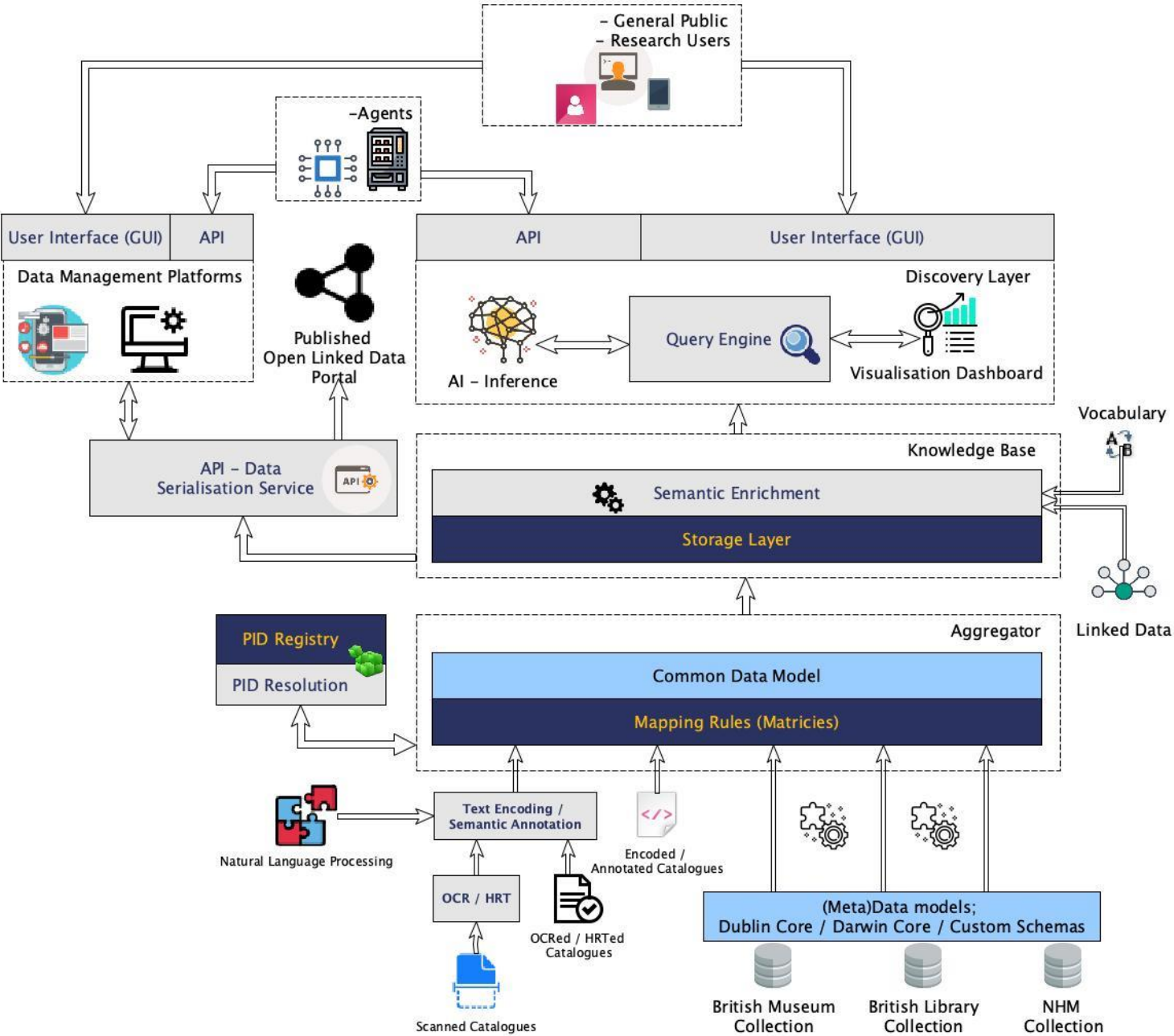
The Sloane Lab

Sloane an **exemplary case study** to unite surviving objects (current digital collections) and records across the NHM, BM and BL from the original catalogues



Reunify and extend digital collection
 Searchable and amenable to further analysis and interpretation





Infrastructure

Data Model



The **Sloane Lab data model** allows us to represent the **knowledge** collected in the project



Based on **Semantic Web** standards and built on top of the **CIDOC CRM** reference model



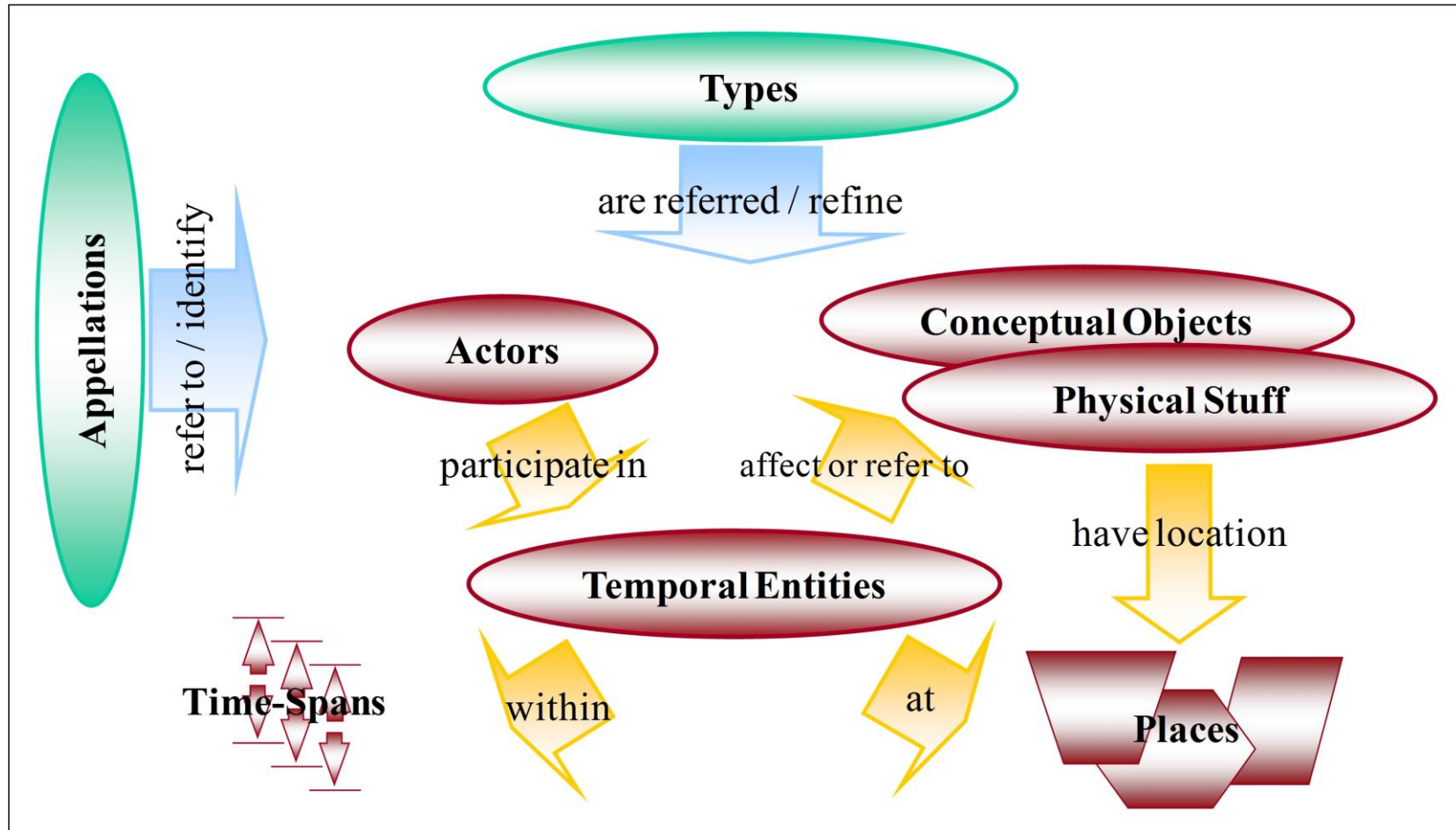
Centred on the **record** instead of the **object**, to support multiple perspectives over the same object

CIDOC - CRM

- The CIDOC CRM (Conceptual Reference Model) ontology is an **international standard (ISO 21127:2006)** and provides a real world, empirically based representation aimed at **harmonizing heterogeneous data**.
- The CRM provides a core ontology that can harmonise between **museum, archive, library**, and other specialised cultural datasets
- **81 Classes and the 160 Properties**



CIDOC - CRM



Source: <https://www.cidoc-crm.org/>

Sloane Lab Data Environment

Charting all data available to project

NATURAL HISTORY MUSEUM		
Sloane Manuscript Catalogues	IMAGED	IMAGED & MACHINE READABLE
Sloane Manuscript Catalogues	Manuscript Catalogues - Digitised	Manuscript Catalogues - Transcriptions
Fossils vol 1	Fossils vol 1 ¹	Fossils vol 1 ¹
Fossils vol 2	Fossils vol 2	Not transcribed
Fossils vol 3	Fossils vol 3	Not transcribed
Fossils vol 4	Fossils vol 4	Not transcribed
Fossils vol 5	Fossils vol 5 ²	Fossils vol 5 ²
Fossils vol 6	Fossils vol 6	Not transcribed
Insects vol 1	Insects vol 1	Not transcribed
Insects vol 2	Insects vol 2	Not transcribed
Minerals vol 1	Minerals vol 1	Not transcribed
Minerals vol 2	Minerals vol 2	Not transcribed
Minerals vol 3a	Minerals vol 3a	Not transcribed
Minerals vol 3b	Minerals vol 3b	Not transcribed
Minerals vol 4	Minerals vol 4	Not transcribed
Echinoderms	Echinoderms	Not transcribed
Supplement	Supplement	Not transcribed
Catalogus marmorum	Catalogus marmorum	Not transcribed
Collection of Vegetables 1-3000	Collection of Vegetables 1-3000	Collection of Vegetables 1-3000
Collection of Vegetables 3001-6000	Collection of Vegetables 3001-6000	Collection of Vegetables 3001-6000
Collection of Vegetables 6001-12523	Collection of Vegetables 6001-12523	Collection of Vegetables 6001-12523
Collection of Vegetables Index (part 1)	Collection of Vegetables Index (part 1)	Collection of Vegetables Index (part 1)
John Ray, Historia Plantarum Vol I (PIS copy)	John Ray, Historia Plantarum Vol I	John Ray, Historia Plantarum Vol I
John Ray, Historia Plantarum Vol II (PIS copy)	John Ray, Historia Plantarum Vol II	John Ray, Historia Plantarum Vol II
John Ray, Historia Plantarum Vol III (PIS copy)	John Ray, Historia Plantarum Vol III	John Ray, Historia Plantarum Vol III
Collection Guides	Collection Guides - Scanned	
Collection Guides - Print	Collection Guides - Scanned	
J.E. Dandy, The Sloane herbarium: an annotated list of the Heriti sicc' composing it	J.E. Dandy, The Sloane herbarium: an annotated list of the Heriti sicc' composing it	
A catalogue and historical account of the Sloane Herbarium Collection by Guy L. Wilkins, Bulletin of the British Museum (Natural History) Historical Series Vol 1, 1953-1959	A catalogue and historical account of the Sloane Herbarium Collection by Guy L. Wilkins, Bulletin of the British Museum (Natural History) Historical Series Vol 1, 1953-1959	
V. Pickering, Putting nature in a box: Hans Sloane's Vegetable Substances collection, unpublished thesis 2017	V. Pickering, Putting nature in a box: Hans Sloane's Vegetable Substances collection, unpublished thesis 2017	
Objects		
Bound Volumes	Bound Volumes - Catalogued	
Sloane Herbarium bound volumes	Sloane Herbarium bound volumes	
Folios	Folios - Scanned and/or Catalogued?	
Sloane volume: folio and specimen level cat	Sloane volume: folio and specimen level cat	
Algae, Fungi and Plants	Algae, Fungi and Plants - DOI-registered collection	
Sloane herbarium: Sloane's Jamaica material	Sloane herbarium: Sloane's Jamaica material	
Sloane herbarium: Mark Calceyo material	Sloane herbarium: Mark Calceyo material	
Sloane herbarium: Engelbert Kaempfer material	Sloane herbarium: Engelbert Kaempfer material	
Vegetable Substances	Vegetable Substances	
Insects	Invertebrates - DOI-registered collections	
Fluker insect bound volume	Fluker insect bound volume	
Petiver insect collection	Petiver insect collection	
Sloane specimens (on boxes on shelves)	Sloane specimens (on boxes on shelves)	
Invertebrates	Invertebrates - DOI-registered collections	
Molluscs	Molluscs	
Other invertebrates	Other invertebrates	
Palaeontology	Palaeontology - DOI-registered collections	
Fossils	Fossils	
Mineralogy	Mineralogy - DOI-registered collections	
Mineral specimens	Mineral specimens	
Vertebrates	Vertebrates - DOI-registered collections	
Fish	Fish	
Herps	Herps	
Mammals	Mammals	
Birds	Birds	

BRITISH LIBRARY		
Sloane Manuscript Catalogues	IMAGED	IMAGED & MACHINE READABLE
Sloane Manuscript Catalogues	Manuscript Catalogues - Digitised	Manuscript Catalogues - Digitised & Transcribed
Sloane MS 3995	not digitised?	not transcribed?
Fossils Sloane MS 3972 A	not digitised?	not transcribed?
Sloane MS 3972 B	Sloane MS 3972 B ¹	Sloane MS 3972 B ¹
Sloane MS 3972 C I	not digitised?	not transcribed?
Sloane MS 3972 C II	not digitised?	not transcribed?
Sloane MS 3972 C III	not digitised?	not transcribed?
Sloane MS 3972 C IV	not digitised?	not transcribed?
Sloane MS 3972 C V	Sloane MS 3972 C V ¹	Sloane MS 3972 C V ¹
Sloane MS 3972 C VI	not digitised?	not transcribed?
Sloane MS 3972 C VII	not digitised?	not transcribed?
Sloane MS 3972 C VIII	not digitised?	not transcribed?
Sloane MS 3972 D	not digitised?	not transcribed?
Londonia reviews	not digitised?	not transcribed?
Objects		
Collected Manuscripts	Collected Manuscripts - Catalogued	
SI MSS 1-4100 (Manuscripts collected by Sloane)	SI MSS 1-4100 (Manuscripts collected by Sloane)	
Books and Manuscripts collected by Sloane	Books and Manuscripts - Catalogued	Books and Manuscripts - Scanned
Printed Books	Printed Books	Not scanned
Sloane Renaissance Manuscripts	Sloane Renaissance Manuscripts	Sloane Renaissance Manuscripts
Other Books & Manuscripts collected by BS	Other Books & Manuscripts collected by BS	Not scanned
Other Catalogues	Other Catalogues - Digitised	
Printed Books Catalogue	Printed Books Catalogue	
Sloane Correspondence	Sloane Correspondence - Digitised	
Sloane Letters MS 4036-4069	Some letters digitised ¹	
Sloane Letters MS 4075-4078	Some letters digitised ¹	

ROYAL SOCIETY		
Draft Minutes		
Draft Minutes - Manuscripts		Draft Minutes - Fully Digitised
Meetings of the Royal Society, 1686-1691	→	Meetings of the Royal Society, 1686-1691
Meetings of the Royal Society, 1693-1698	→	Meetings of the Royal Society, 1693-1698
Meetings of the Royal Society, 1699-1702	→	Meetings of the Royal Society, 1699-1702
Meetings of the Royal Society, 1702-1707	→	Meetings of the Royal Society, 1702-1707
Meetings of the Royal Society, 1707-1711	→	Meetings of the Royal Society, 1707-1711
Enlightenment Architectures		
Manuscript Catalogues - Fully Digitised (FID)		
¹ Digital edition of MS 3972C Vol VI, Sloane's catalogue of books and printed ephemera		
² Digital edition of Sloane MS 3972B of folios listing manuscript material removed from 3972C		
³ Digital edition of Sir Hans Sloane's Catalogues of Miscellanies, Antiquities, Seals, Pictures, Mathematical Instruments, Agate		
⁴ Digital edition of Sir Hans Sloane's Catalogue of Fossils including Corals, Serpents, Echin, Crustacea, Starfishes, Humans I		
⁵ Digital edition of Sir Hans Sloane's Catalogue of Fossils including Fishes, Birds, Eggs, Quadrupeds (Volume VI)		
Sloane Letters		
Digitised Letters written to and from Sloane		
MS 620 Searchable Letters		

- Footnotes**
- 1 Also available in Enlightenment Architectures as "Digital edition of Sir Hans Sloane's Catalogue of Fossils including Corals, Serpents, Echin, Crustacea, Starfishes, Humans (Volume VI)"
 - 2 Also available in Enlightenment Architectures as "Digital edition of Sir Hans Sloane's Catalogue of Fossils including Fishes, Birds, Eggs, Quadrupeds (Volume VI)"
 - 3 Available in Sloane Letters as "MS 620 Searchable Letters"
 - 4 Also available in Enlightenment Architectures as "Digital edition of Sloane MS 3972C of folios listing manuscript material removed from 3972C"
 - 5 Also available in Enlightenment Architectures as "Digital edition of MS 3972C Vol VI, Sloane's catalogue of books and printed ephemera"
 - 6 Also available in Enlightenment Architectures as "Digital edition of Sir Hans Sloane's Catalogues of Miscellanies, Antiquities, Seals, Pictures, Mathematical Instruments, Agate"
 - 7 Also available in British Library as "Sloane MS 3972 C VI"
 - 8 Also available in British Library as "Sloane MS 3972 B"
 - 9 Also available in British Museum as "Miscellanea vol"
 - 10 Also available in Natural History Museum as Fossils vol 1
 - 11 Also available in Natural History Museum as Fossils vol 5
 - 12 Available in British Library as "Sloane Correspondence"

Colour index

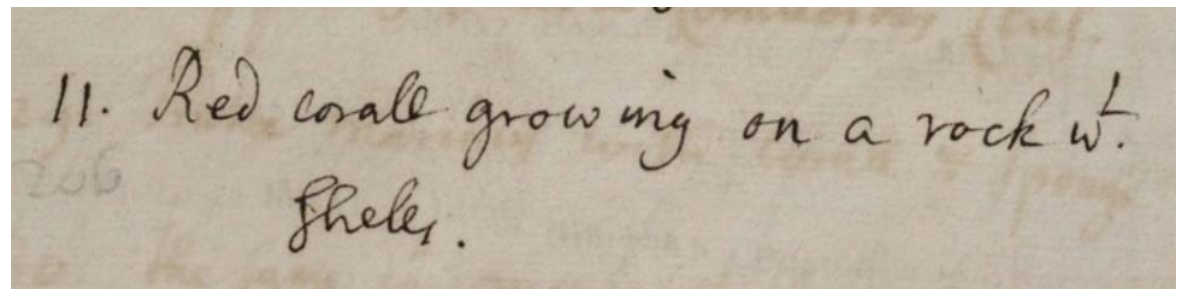
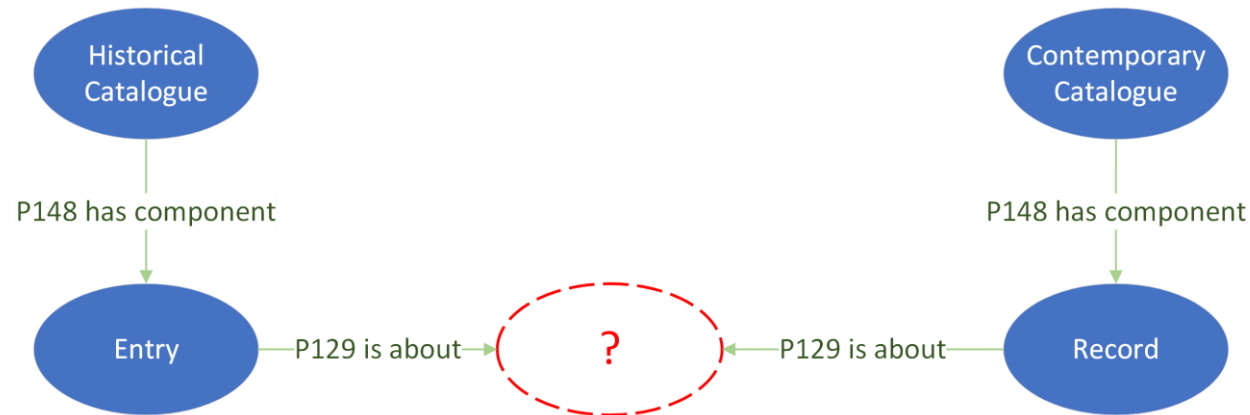
Colour-coding is only applicable to cells relating to data sources

- Digital - Available digitally and available publicly
- Digital - In the process of becoming available
- Tangible artefact
- Permanently destroyed

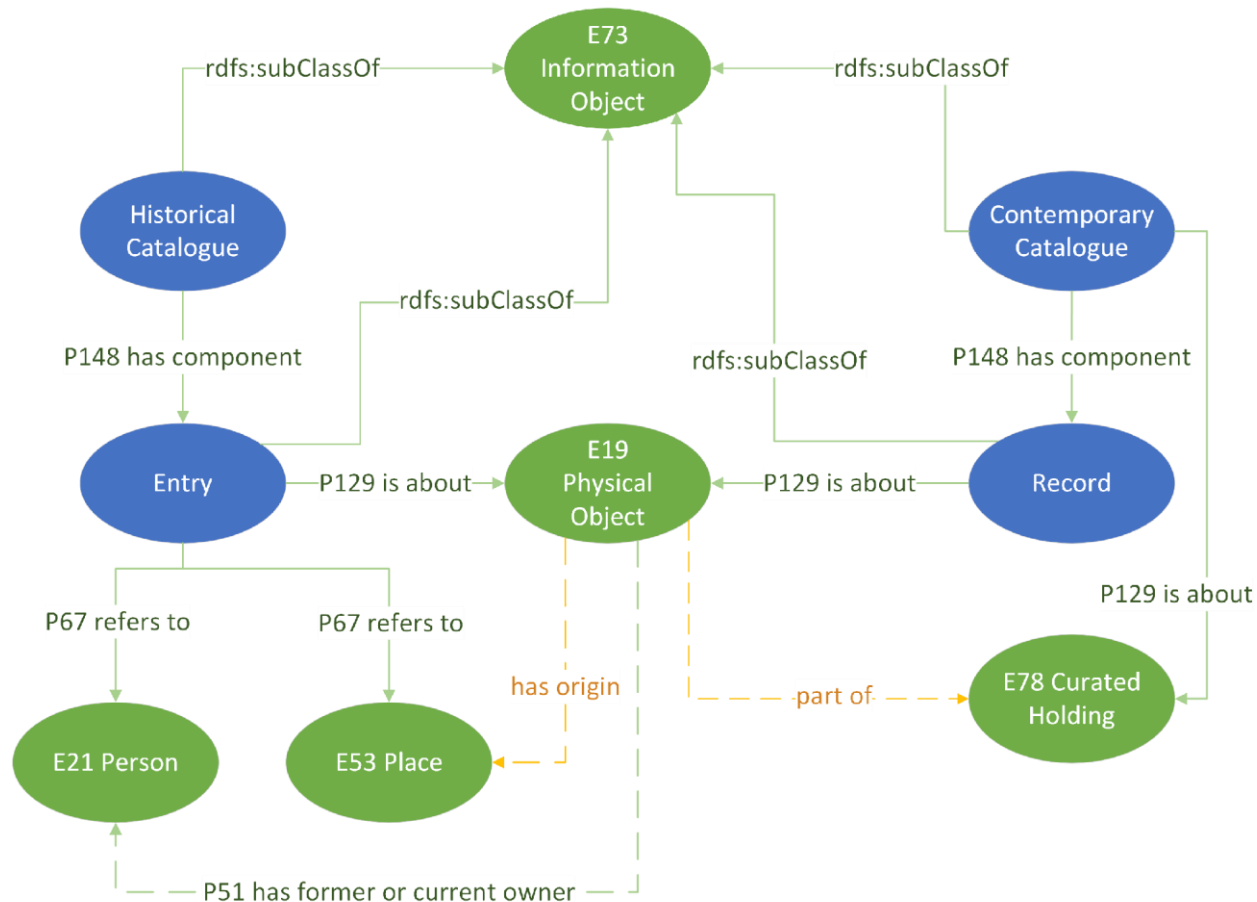
The British Museum		
Sloane Manuscript Catalogues	IMAGED	FULLY DIGITISED
Sloane Manuscript Catalogues	Manuscript Catalogues - Digitised	Manuscript Catalogues - Transcribed
Coins, Cameos, Jewels	Indices variorum	Indices variorum
Indices variorum	Not digitised	Not transcribed
Amuleta Muhametica	Miscellanea vol	Miscellanea vol
Miscellanea vol		
Objects by department	CATALOGUED (13,457)	IMAGED (11,830)
Prints & Drawings	Prints & Drawings - Catalogued	Prints & Drawings
11,663 objects	11,663 objects	How many objects imaged?
Britain, Europe & Prehistory	Britain, Europe & Prehistory	Britain, Europe & Prehistory
795 objects	795 objects	How many objects imaged?
Greece & Rome	Greece & Rome	Greece & Rome
438 objects	438 objects	How many objects imaged?
Middle East	Middle East	Middle East
262 objects	262 objects	How many objects imaged?
Asia	Asia	Asia
202 objects	202 objects	How many objects imaged?
Coins & Medals	Coins & Medals	Coins & Medals
113 objects	113 objects	How many objects imaged?
Africa, Oceania and Americas	Africa, Oceania and Americas	Africa, Oceania and Americas
97 objects	97 objects	How many objects imaged?
Egypt & Sudan	Egypt & Sudan	Egypt & Sudan
55 objects	55 objects	How many objects imaged?

The Object Problem

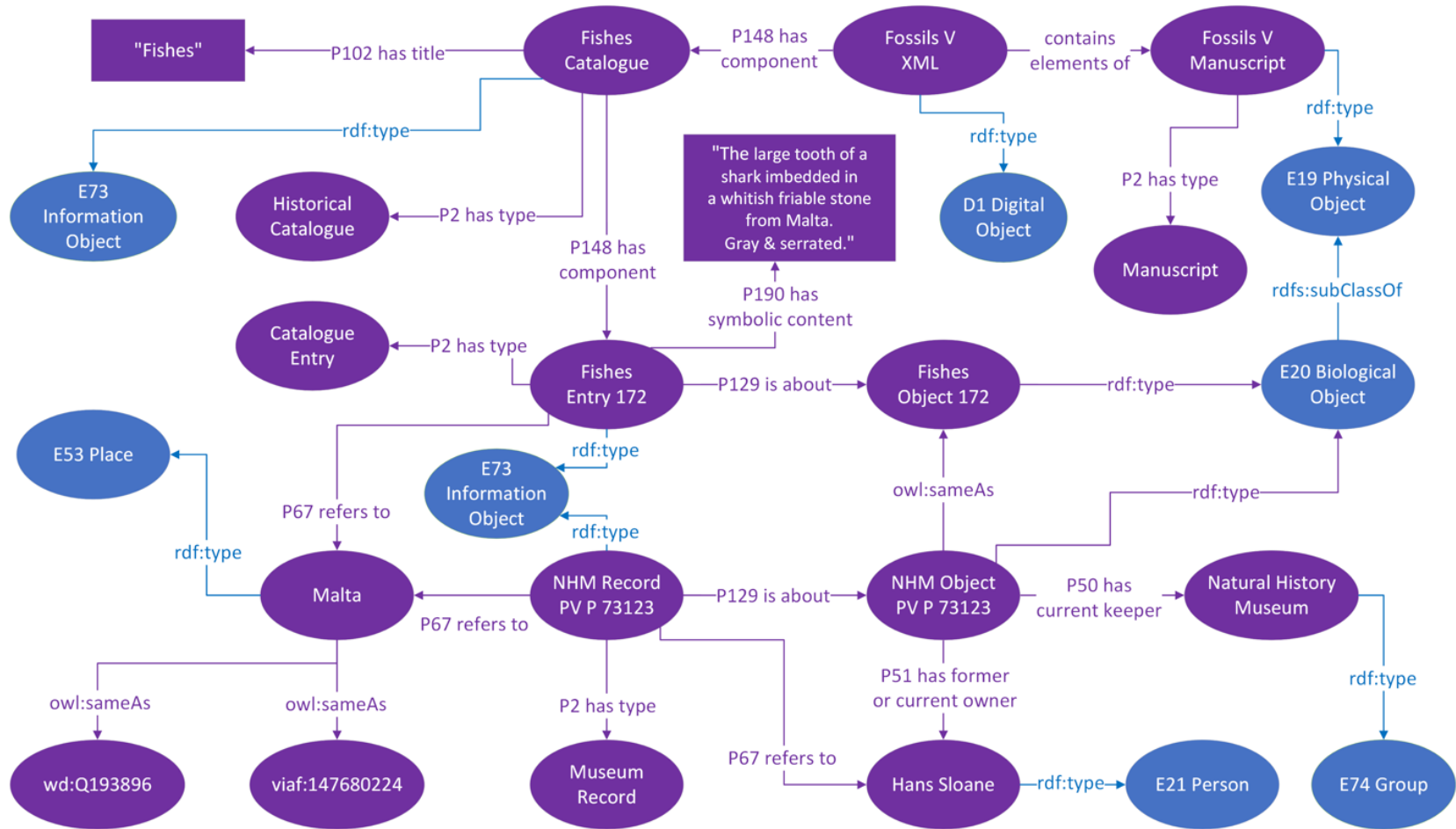
- Often we cannot rely on objects to connect historical and contemporary catalogues
- Moreover, it is challenging to “identify the boundaries between ‘object names’ and their qualifying descriptions”



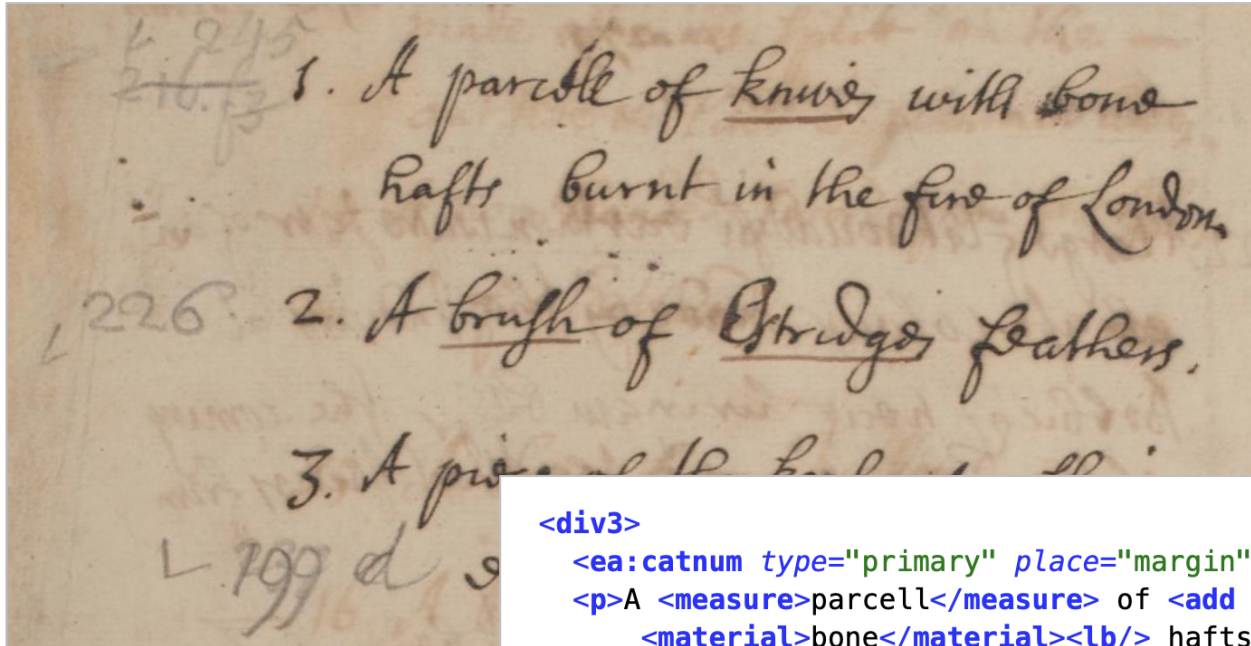
Information Object Focus



Fishes Entry 172 — “The large tooth of a shark”
object with a Natural History Museum record



Challenges - Historical Catalogue



No identifier for the catalogue entry

Identifiers for people & places

```

<div3>
  <ea:catnum type="primary" place="margin">1.</ea:catnum>
  <p>A <measure>parcell</measure> of <add rend="underline">knives</add> with
    <material>bone</material><lb/> hafts burnt in the fire of
    <placeName ref="#London">London</placeName>.</p>
  <p>
    <add rend="pencil">✓</add>
    <add rend="pencil">&lt; 245</add>
    <add rend="pencil">
      <add rend="del">216.f3</add>
    </add>
  </p>
</div3>
  
```

British Museum Data

abacus

Object Type

abacus

Museum number

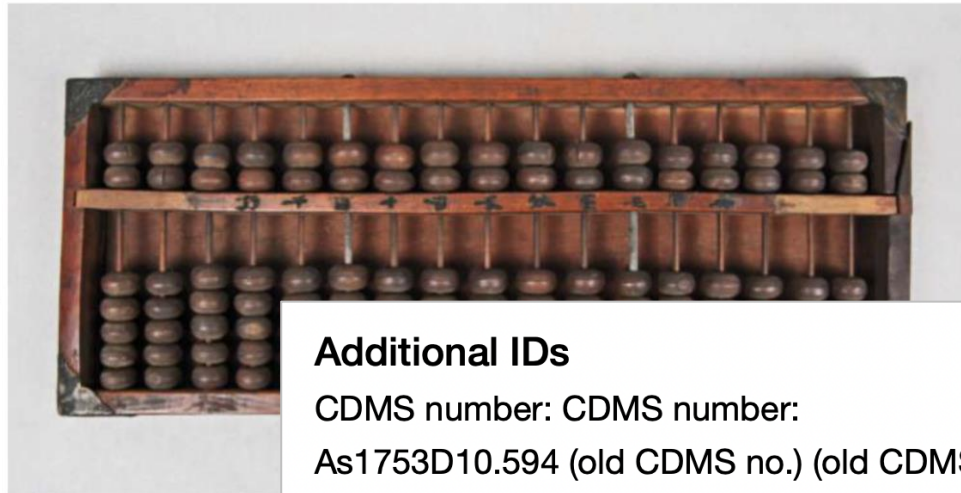
As,SLMisc.594

Description

Abacus (with 17 rows of beads, divided by cross-bar carrying Chinese characters) made of wood, metal.

Findspot

Found/Acquired: China



Additional IDs

CDMS number: CDMS number:

As1753D10.594 (old CDMS no.) (old CDMS no.)

Miscellaneous number: Miscellaneous number:
104 (paper label attached) (paper label attached)

Sloane Miscellanea Catalogue number: Sloane
Miscellanea Catalogue number: 594
(Miscellanies) (Miscellanies)

Multiple identifiers, not persistent

Not always unique

Further Challenges – Uncertainty

Aggregating records
and integrating them
into a single
knowledge base

Handling uncertainty
in records

Modelling
inconsistencies
among the records

Identifying data
absences and
attempting to remedy
them

-1100. A large reed pipe from the great Mogols country from
M^r. Dupuys.
b √ 245.

smoking-pipe

Object Type
smoking-pipe

Museum number
As,SLMisc.1100

Description
Bamboo smoking pipe with hole drilled in centre, through only one side. Bands of incised geometric patterns mainly to each end, but also a floral design to one side of the drilled hole.

Production date
18thC

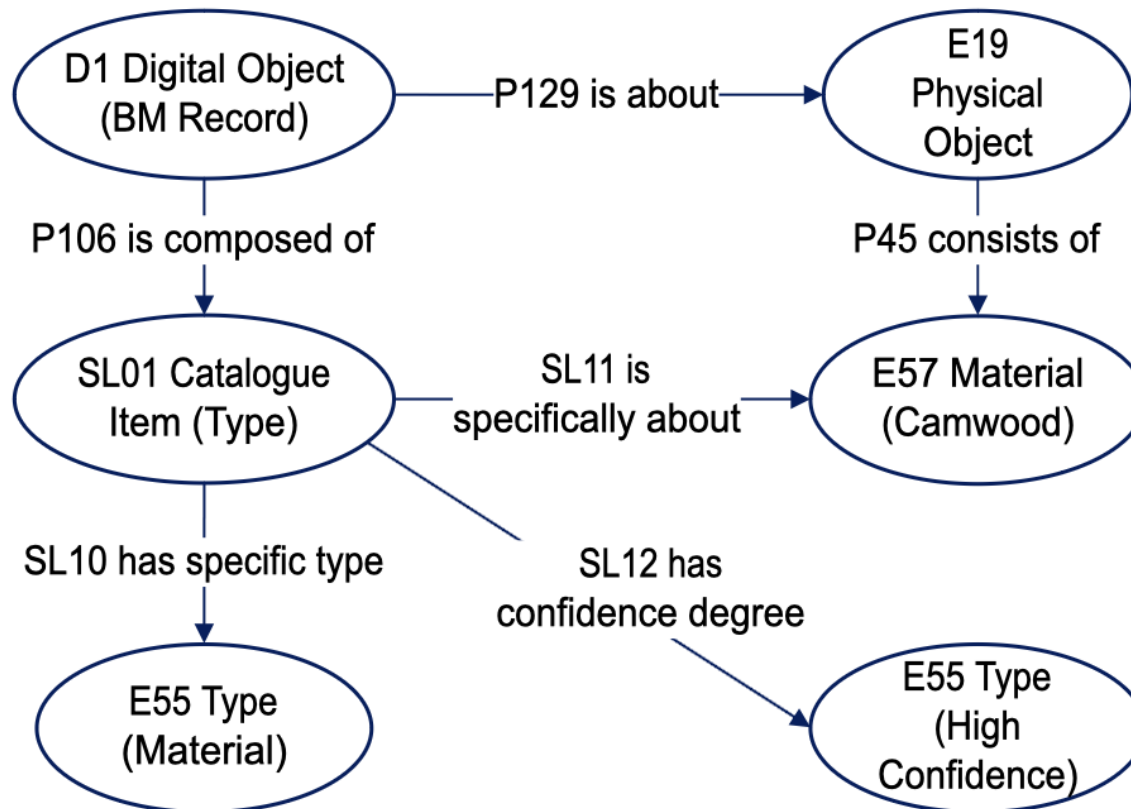
Production place
Made in: India (pre-1947) (?)
Asia: South Asia: India (pre-1947)



Extensions to CIDOC - CRM

- Sloane Lab – Catalogue Item
- Each **record** or **entry** may refer to one or more **people** (E21), **places** (E53), **materials** (E57), **techniques** (E29 Design or Procedure), or other entities that are associated to the object. We view each of these references as an individual part of the record/entry that we call **catalogue item** (Sloane Lab SL1).
- Each **catalogue item** has a specific **type** and is about one or more **entities** of this type.
- Each catalogue item may also have a **confidence level** (to represent uncertainty), and possibly a **note** attached to it.

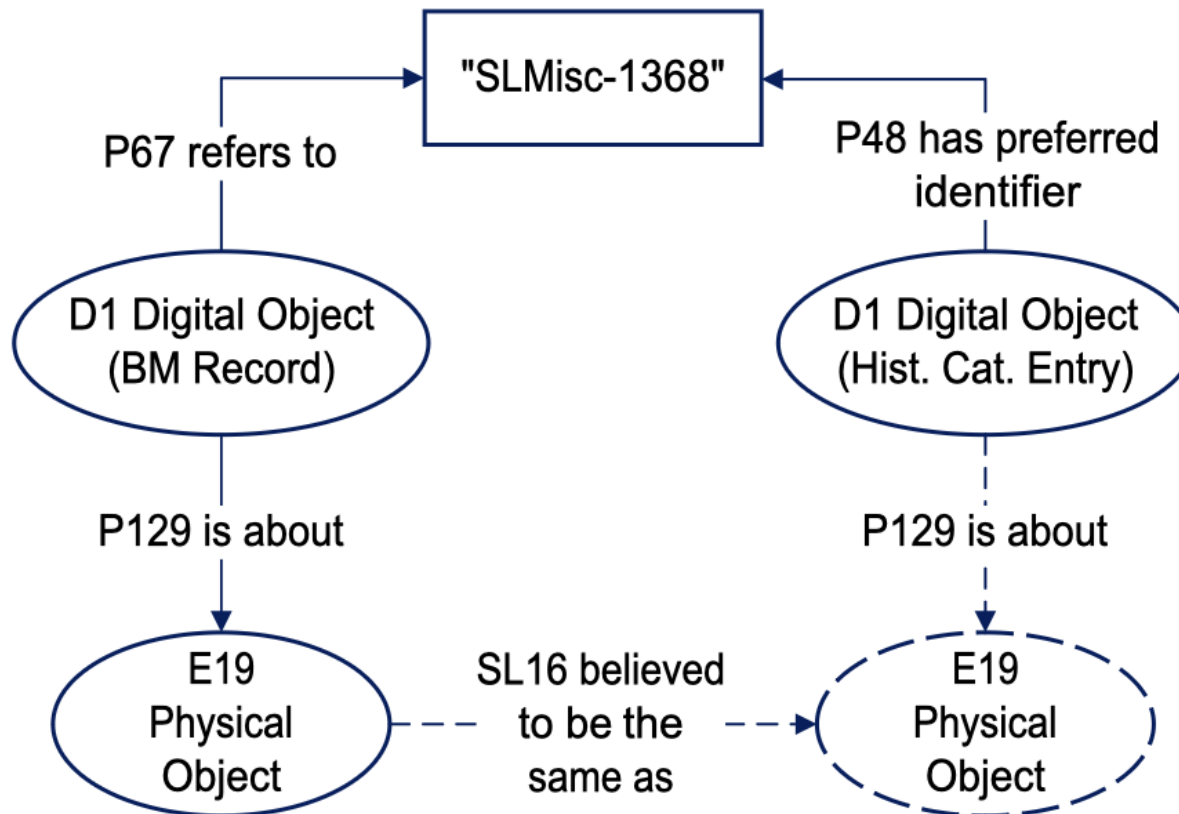
Sloane Lab – Catalogue Item



Linking Multiple Records

- Multiple **records** (or entries) can be linked to each other when they refer to the same **physical object** or when they mention the same entities (people, places, etc.).
 - However, linking through the physical object is problematic because many physical objects have not been preserved, and in some cases the connection is uncertain.
- For these reasons, instead of going through the physical object, we have decided to link the records (or entries) through the **identifier** (E42).

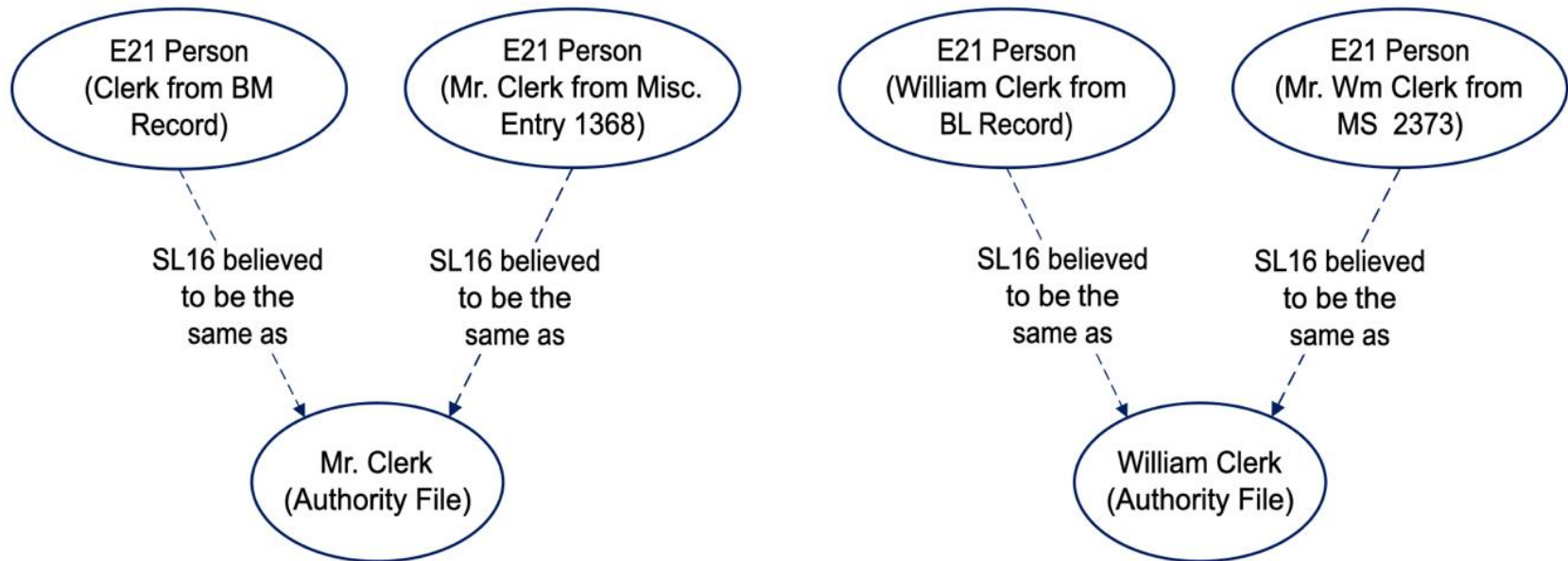
Linking of a museum record to a historical catalogue entry



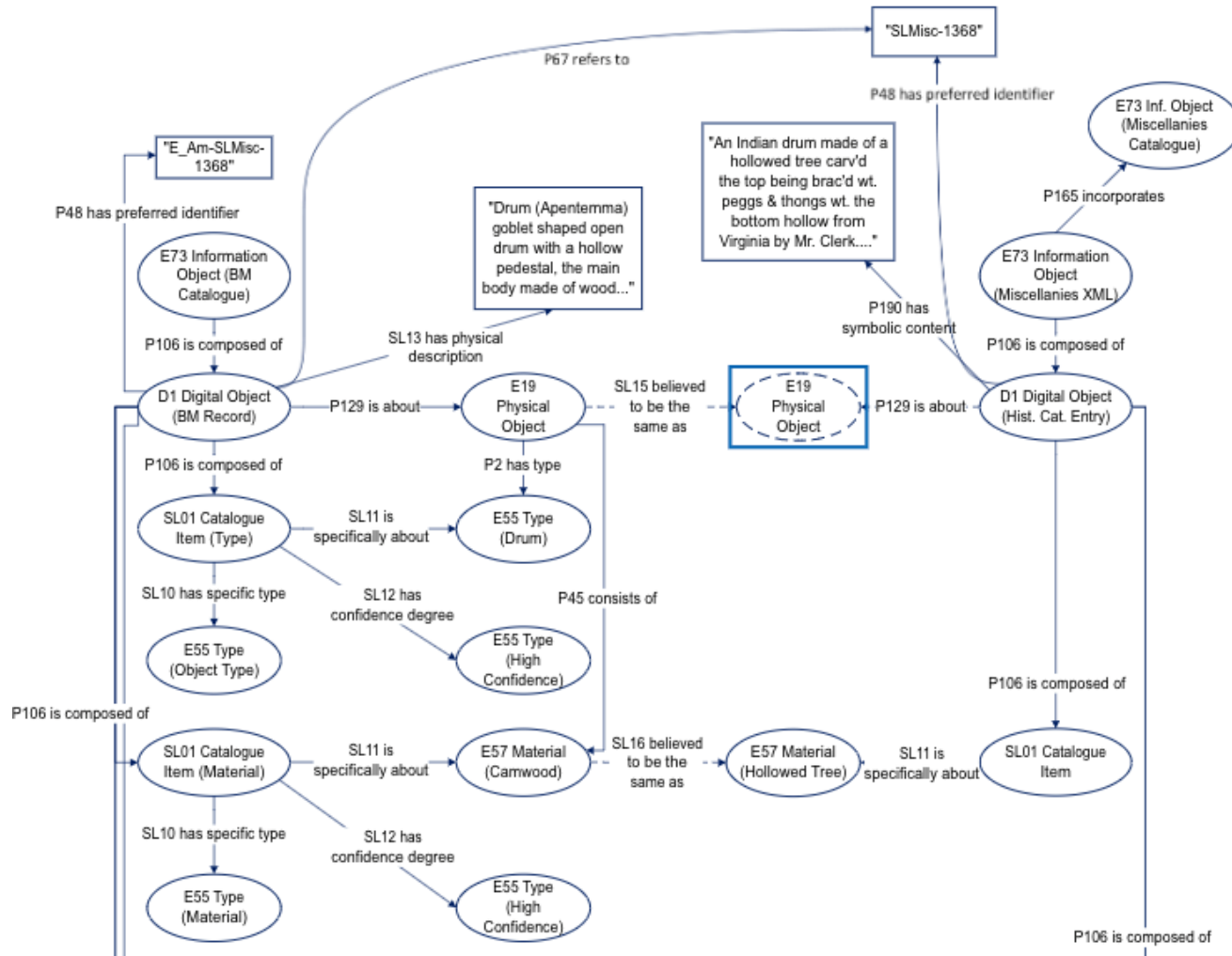
Modelling Identity

- In many cases, two **records** (or entries) will mention the same **entity**. For example, Miscellanies 1368 mentions a person called “Mr. Clerk”, while the BM record for the corresponding object mentions “Clerk” as the field collector.
- Two entities that are believed to be the same can be linked together through the property **SL16 believed to be the same as**. However, it would be impractical to directly connect each “Clerk” entity to all related “Clerk” entities

Linking of entities mentioned in different records



Example – SL Semantics



Conclusions

- CIDOC- CRM a de facto standard in CH
 - Extension can unlock domain specific challenges
- Persistent Identification
 - different identifiers used by institutions, make sense of them, and connect them to the Sloane Lab identifiers
- The role of museum-based vocabulary resources towards the realisation of unified collections
 - opportunities for semantic enrichment, linking and interoperability

Semantic technologies for historical collections:

A case study from the Sloane Lab Knowledge Base

Dr Andreas Vlachidis and Dr Daniele Metilli

UCL Department of Information Studies

a.vlachidis@ucl.ac.uk

d.metilli@ucl.ac.uk