

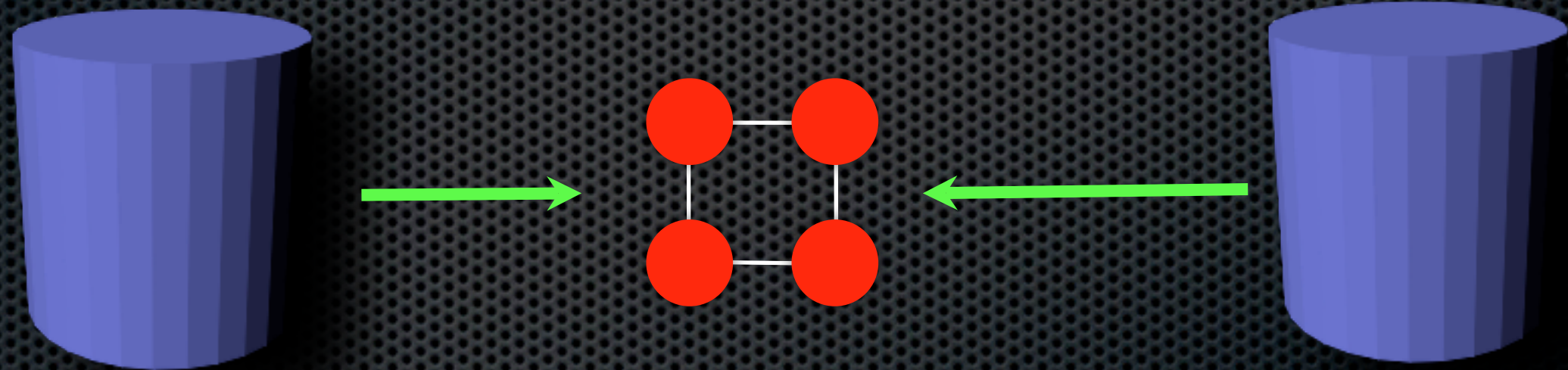
Archaeology, Formality & the CIDOC CRM

Leif Isaksen, Kirk Martinez & Graeme Earl
ECS/Archaeology
University of Southampton

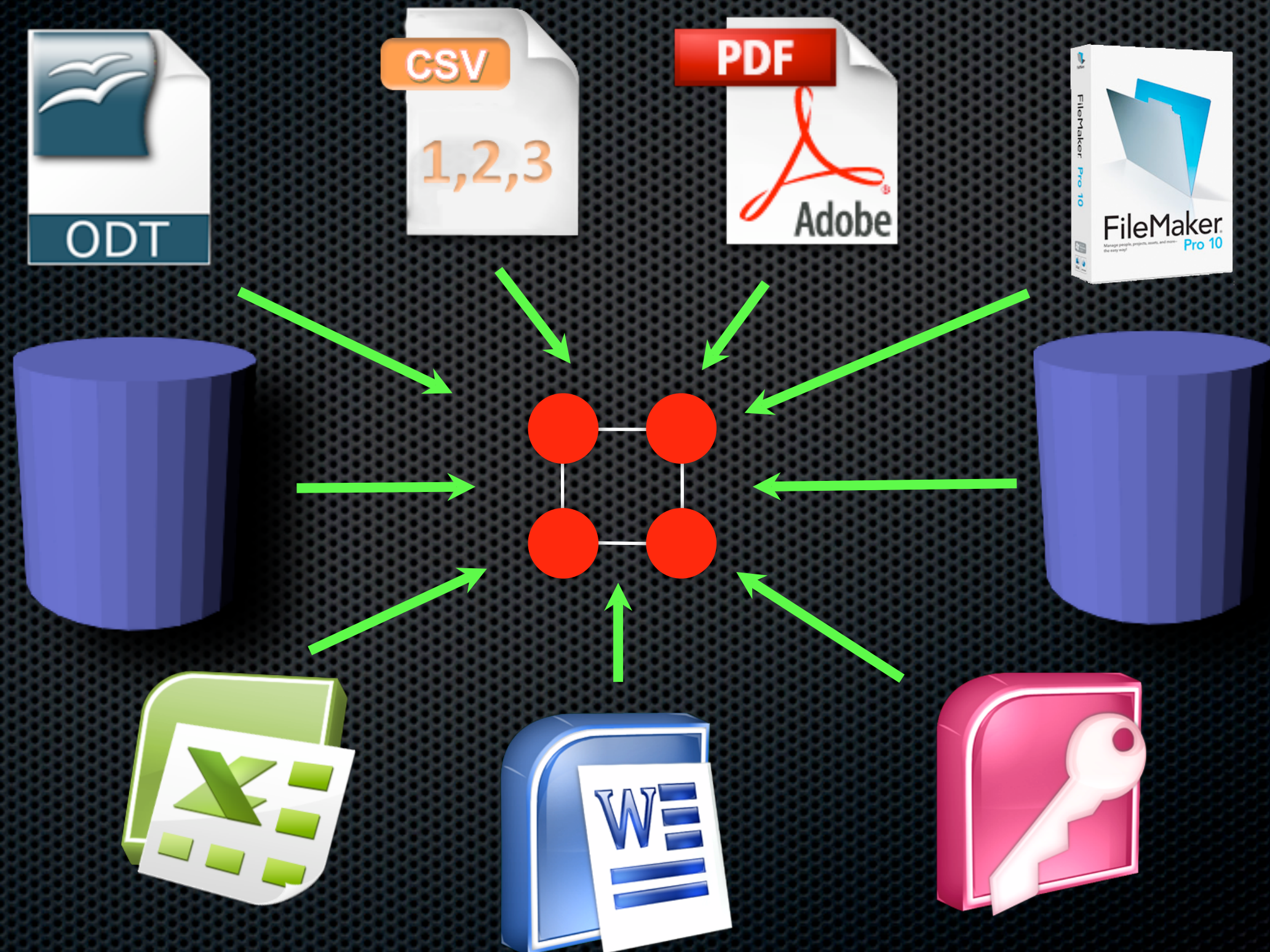
Is the CIDOC CRM too hard?

- “The initial idea that the Domain Experts would be able to discuss their Domain in CRM terms proved difficult” (Cripps et al. 2004)
- “We found that ontology mapping requires close collaboration between computer scientists[...]museum professionals[...]and external experts who understand the CRM. “Collaboration among these parties is time and effort consuming...” (Addis et al. 2005)
- “Es soll aber nicht verschwiegen werden, dass sich das CRM für Laien auf dem Gebiet der Informationsmodellierung nur nach ernsthafter Arbeit erschließen wird.” (Stein et al. 2005)
- “This strong interdependency between the mapping process and the implementation requires several feedback cycles between the mapping experts and the application developers” (Nussbaumer & Haslhofer 2007)

DB -> Ontology Map



~~DB~~ Stuff -> Ontology Map



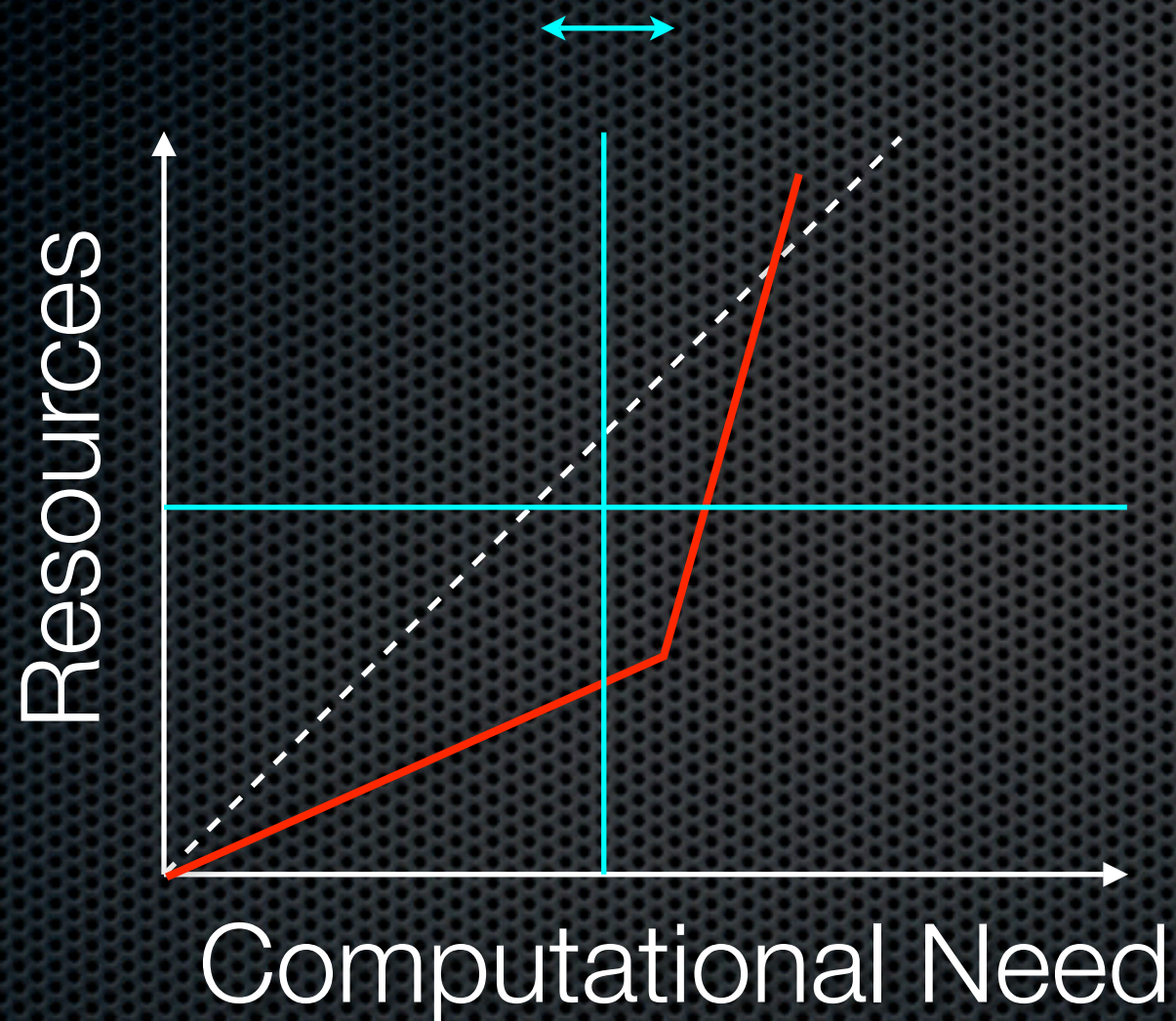
Interoperate with whom? - The role of Microproviders

- ✦ 'Long Tail' of the Academic Community
- ✦ Small but valuable datasets
- ✦ Extremely limited resources
- ✦ Fidelity to source material is paramount

Formality Considered Harmful?

- ✦ Marshall & Shipman 1999
- ✦ Formality as a trade-off:
 - ✦ Improves computational power?
 - ✦ Increases difficulty for users?
 - ✦ Every user finds an equilibrium
 - ✦ If effort increases with power, interoperability (i.e. ∞ computing power) becomes self-defeating?

The Goldilocks Effect

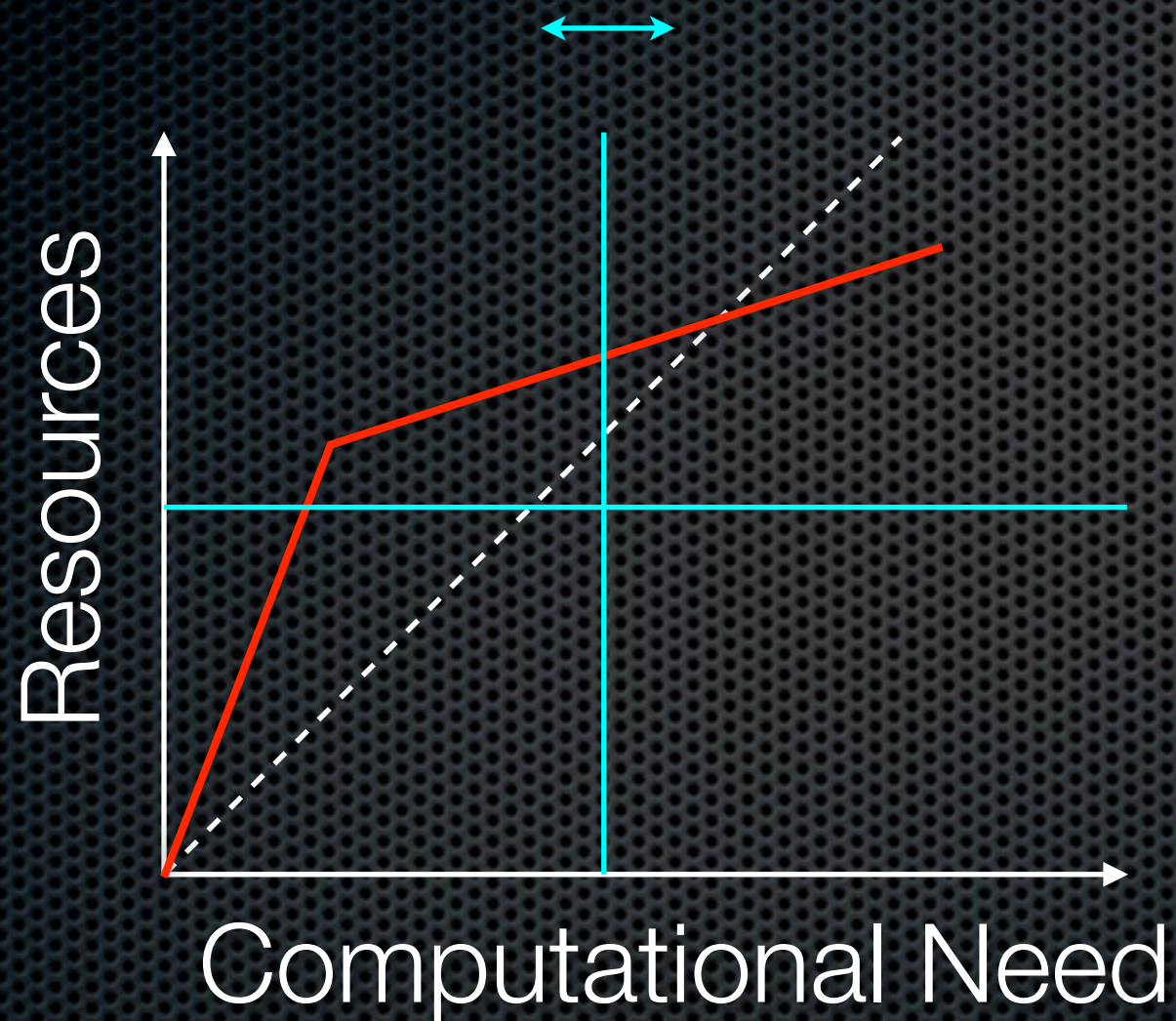


Cost/Benefit path of 'easy adoption' technology (e.g. MS Access, HTML)

'Complexity debt' causes long-term scalability problems

— User-dependent threshold

The Goldilocks Effect

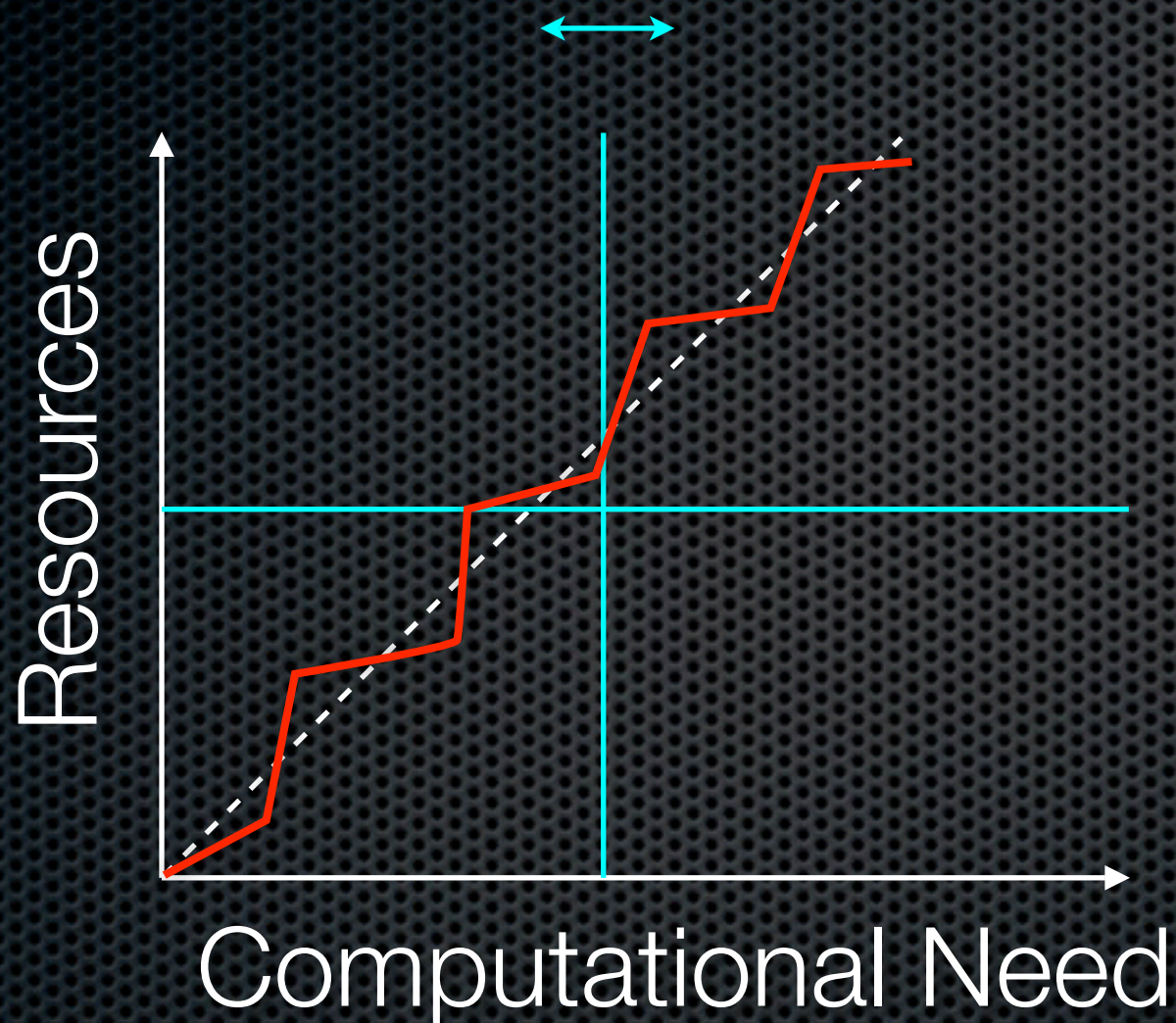


Cost/Benefit path of 'Front-loading' technologies (e.g. Semantic Web)

High barrier to entry reduces no. of participants

— User-dependent threshold

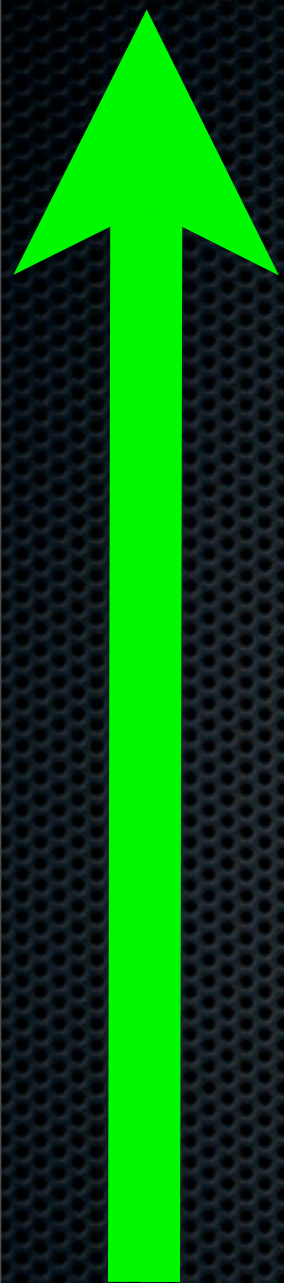
The Goldilocks Effect



Ideal Interoperability
Cost/Benefit path has
shorter 'wavelength',
increasing the no. of
pay-off points

—— User-dependent threshold

3 Steps to Heaven



3. Load

2. Transform

1. Extract

8 Steps to Heaven

Hosting

Validation

RDF generation

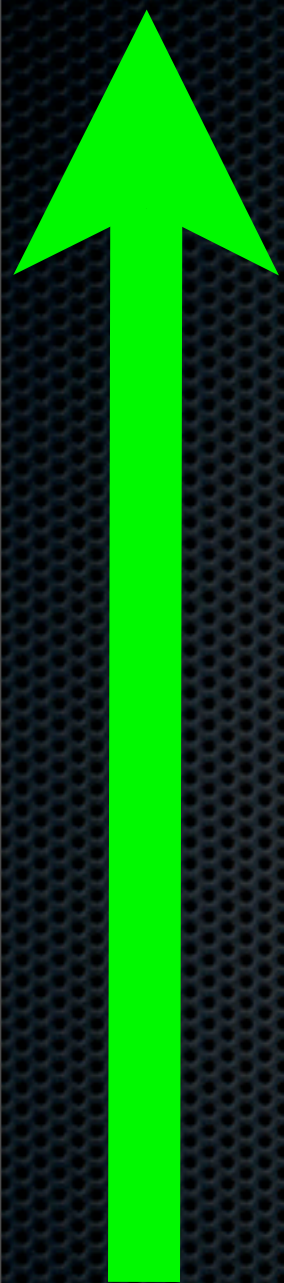
Literal Manipulation

Canonical URI Mapping

Instance URI generation

Schema -> Ontology mapping

Comprehension of task/workflow/ontology



FCH 1: Difficulties Arising

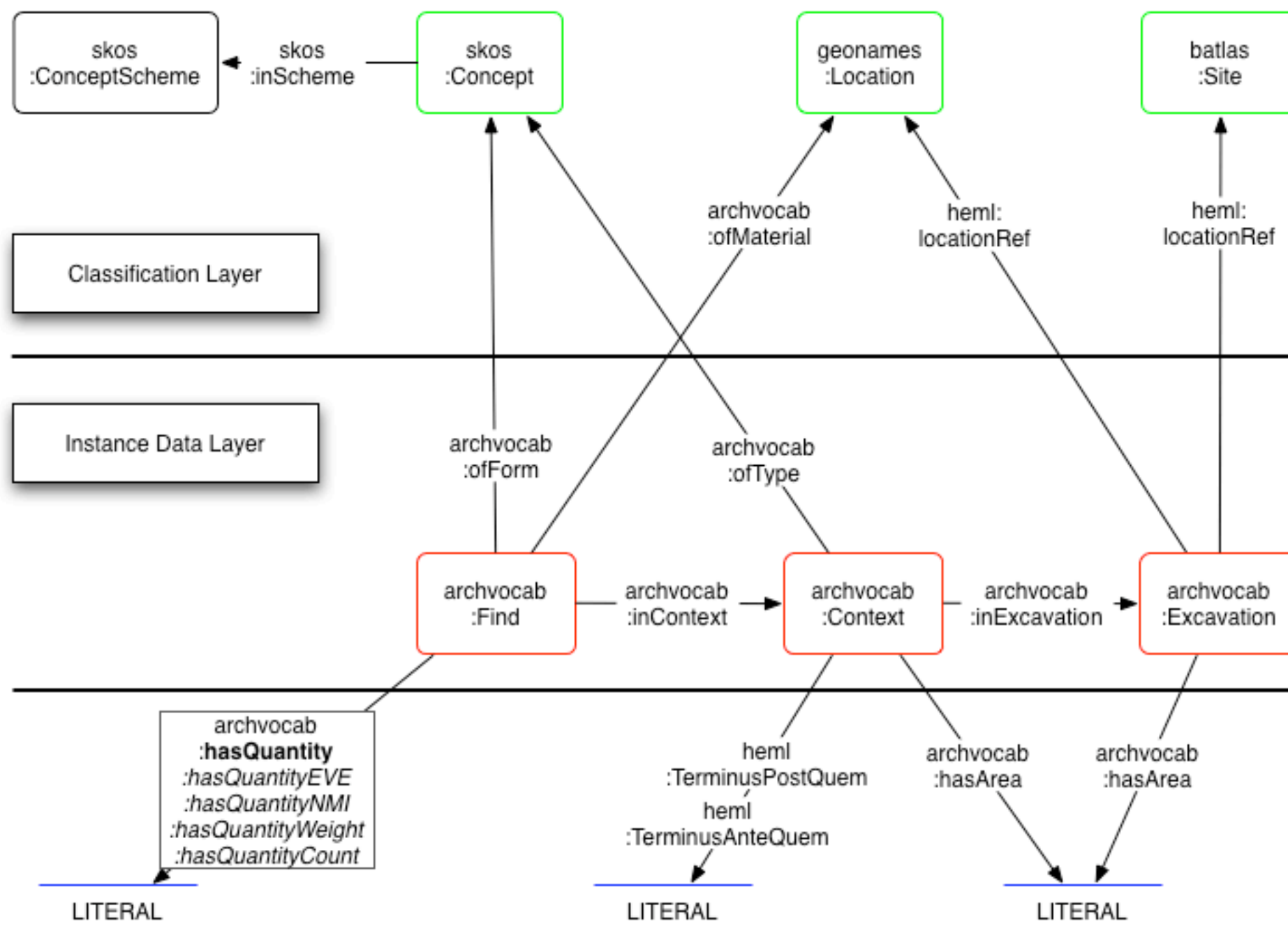
1. Cognitive Overhead
2. Tacit Knowledge
3. Enforcing Premature Structure
4. Different People, Different Tasks: Situational Structure

FCH 2:

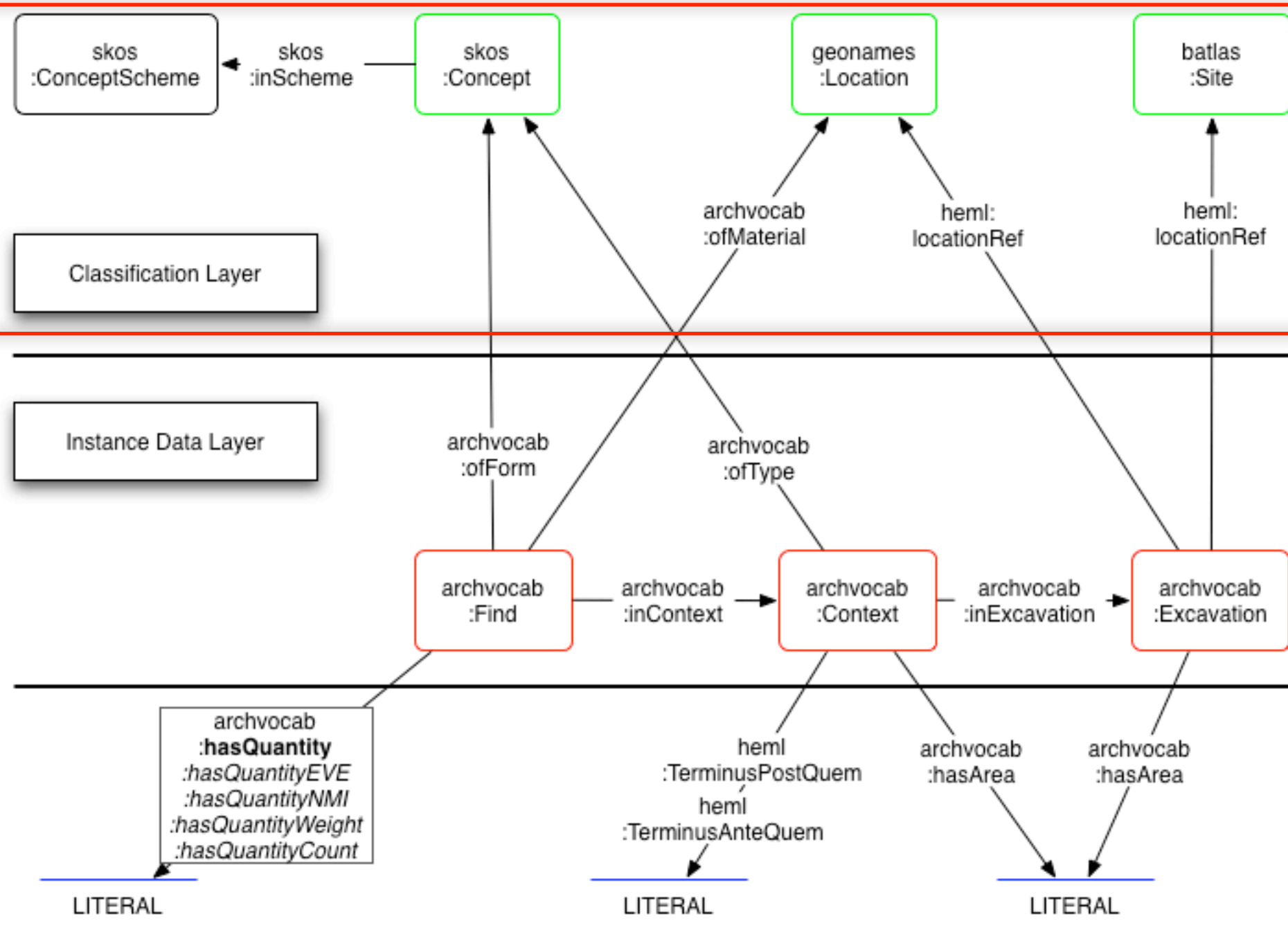
Mitigation Strategies

1. Identify Essentials for Task
2. Evaluate Cost/Benefit Trade Off
3. Gradual Formalization and Restructuring
4. Ephemeral Structure on Demand
5. Training, Facilitation and Intervention

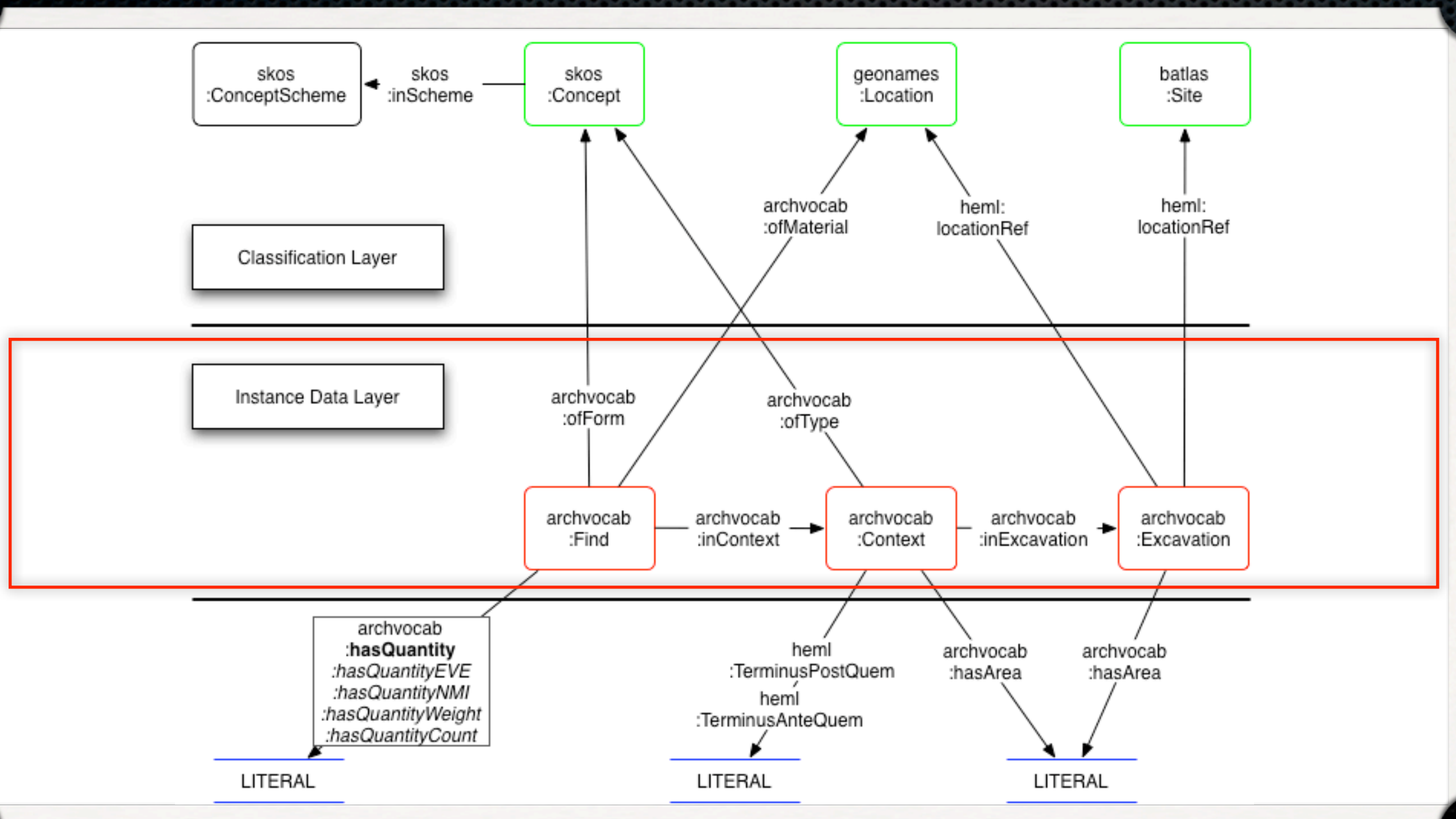
Identify Essentials for Task



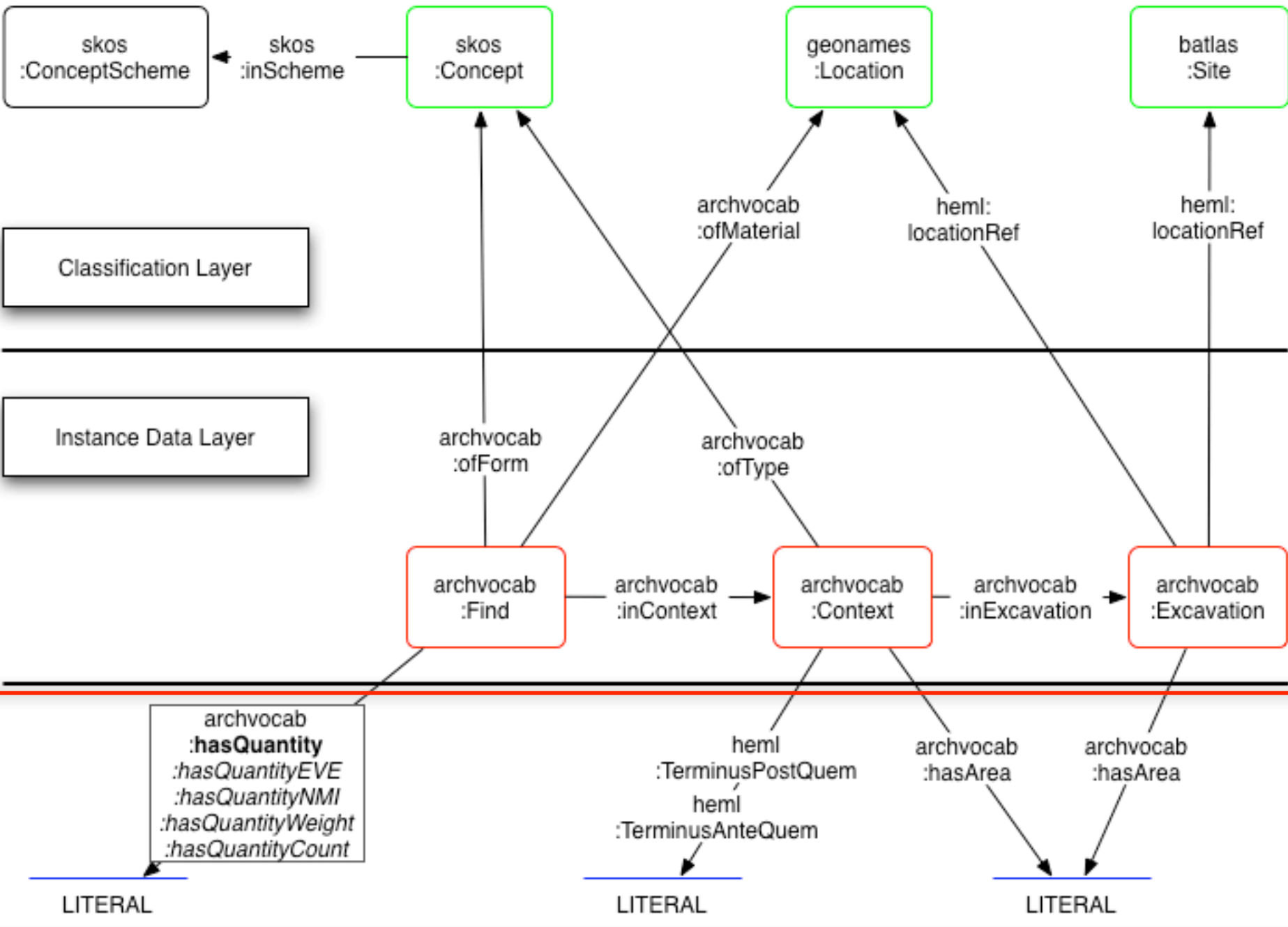
Identify Essentials for Task



Identify Essentials for Task



Identify Essentials for Task



Cost/Benefit Trade Off

Map Concepts to DB Fields

Please select the correct table or query from the database: finds

Subject	Predicate	DB Mapping
Find	ofMaterial	type
Find	hasQuantity	_MANUAL_
Find	ofForm	type
Find	inContext	context
Context	ofType	_MANUAL_
Context	hemi:TerminusPostQuem	_MANUAL_
Context	hemi:TerminusAnteQuem	_MANUAL_
Context	hasArea	_MANUAL_
Context	inExcavation	site
Excavation	hasArea	_MANUAL_

skos:Concept

skos:Concept

Find

inContext (context)

inExcavation (site)

Excavation

hasQuantity (-)

hemi:TerminusPostQuem

hemi:TerminusAnteQuem

hasArea (-)

hasArea (-)

rdfs:Literal

rdfs:Literal

rdfs:Literal

rdfs:Literal

rdfs:Literal

next

site

Table -> Ontology Mapping

Gradual Formalization/ Restructuring

Map Excavation Location(s)

Excavation	hasArea (m2)	locationRef	GeoNames ID
A95	500	Barcino	✓ 3128760: Barcelona (seat
A96			
AB88			
Adm.94			
Argenter			
BLL61			
C.Vell			
CAL'A93			
CAT89			
Font97			
JAU-I			
Llimona			
PA88			
PC90			
PIA			
RC89			
Regomir			
SA93			
SCR81-98			
SM89			
SP96-98			
Sant lu			
Tinell			
Xirau97			

next

Map Contexts

Context TPQ TAQ

Context	inExcavation	hasArea (m2)	ofType
11006	Plaza Encarnacion		Construction
11007	Plaza Encarnacion		Public
11010	Plaza Encarnacion		Public
11012	Plaza Encarnacion		Public
11014	Plaza Encarnacion		Public
11015	Plaza Encarnacion		Dump
11019	Plaza Encarnacion		Funerary
11024	Plaza Encarnacion		Funerary
11026	Plaza Encarnacion		Domestic
11035	Plaza Encarnacion		Transport
11050	Plaza Encarnacion		Transport
11078	Plaza Encarnacion		
11094	http://departamento.us.es/dpreyarq/web/anforas/excavation/plaza+encarnacion		
11114	Plaza Encarnacion		
11116	Plaza Encarnacion		
11124	Plaza Encarnacion		
11133	Plaza Encarnacion		
11134	Plaza Encarnacion		
11138	Plaza Encarnacion		
11201	Plaza Encarnacion		
11205	Plaza Encarnacion		
11211	Plaza Encarnacion		
11213	Plaza Encarnacion		
11215	Plaza Encarnacion		
11219	Plaza Encarnacion		

next

Excavation URI generation

Context URI generation

Ephemeral Structure on Demand

Map Types to Type Series

Records Mapped: 94%

Local term	Canonical Term	Include?	Records
bajo imperial	?	No	1
beltrán	Beltrán	Yes	251
bonifay	Bonifay	Yes	1
común	?	No	2
comúndrss	?	No	1
dolio	?	No	1
dr	Dressel	Yes	4
drees	?	No	1
dress	Dressel	Yes	624
egea	?	No	2
egeo	?	No	1
g	Gaulish	Yes	87
h	Haltern	Yes	167
huelva	?	No	3
indetermiando	?	No	1
indeterminado	_UNKOWN_	Yes	259
islámica	?	No	1
islámicas o cristianas	?	No	2
k	Kapitän	Yes	864
lra	Late Roman Amph...	Yes	80
m	Kuzmanov	Yes	2
ndra 1 ficado	Laietana	No	1
os lra 1 ó similar	Lamboglia	Yes	1
pe lra 12	✓ Late Roman Am	Yes	10
pr lra 2	Leptiminus	No	2
pu lra 3	Locale	Yes	1
ri lra 4	Lomba do Canh	Yes	4
rodas	Lusitanian	No	3
rodia	?	No	3
similar a g	?	No	1
similar a dress	?	No	1

next

Typology prediction

Map local terms to Types

Type	Canonical Term
▶ Beltrán	
▶ Bonifay	
▶ Dressel	
▶ Gaulish	
▶ Haltern	
▶ Kapitän	
▼ Late Roman Amphora	
Ira 5/6	_UNKNOWN_
Ira 1	Late Roman Amphora 1
Ira 1 ó similar	Late Roman Amphora 1
Ira 2	Late Roman Amphora 2
Ira 3	Late Roman Amphora 3
Ira 4	Late Roman Amphora 4
Ira 12	Late Roman Amphora 12
▶ Maña	
▶ Ostia	
▶ Peacock and Williams	
▶ Puerto Real	
▶ Richborough	
▶ Spatheion	
▶ _UNKOWN_	

next

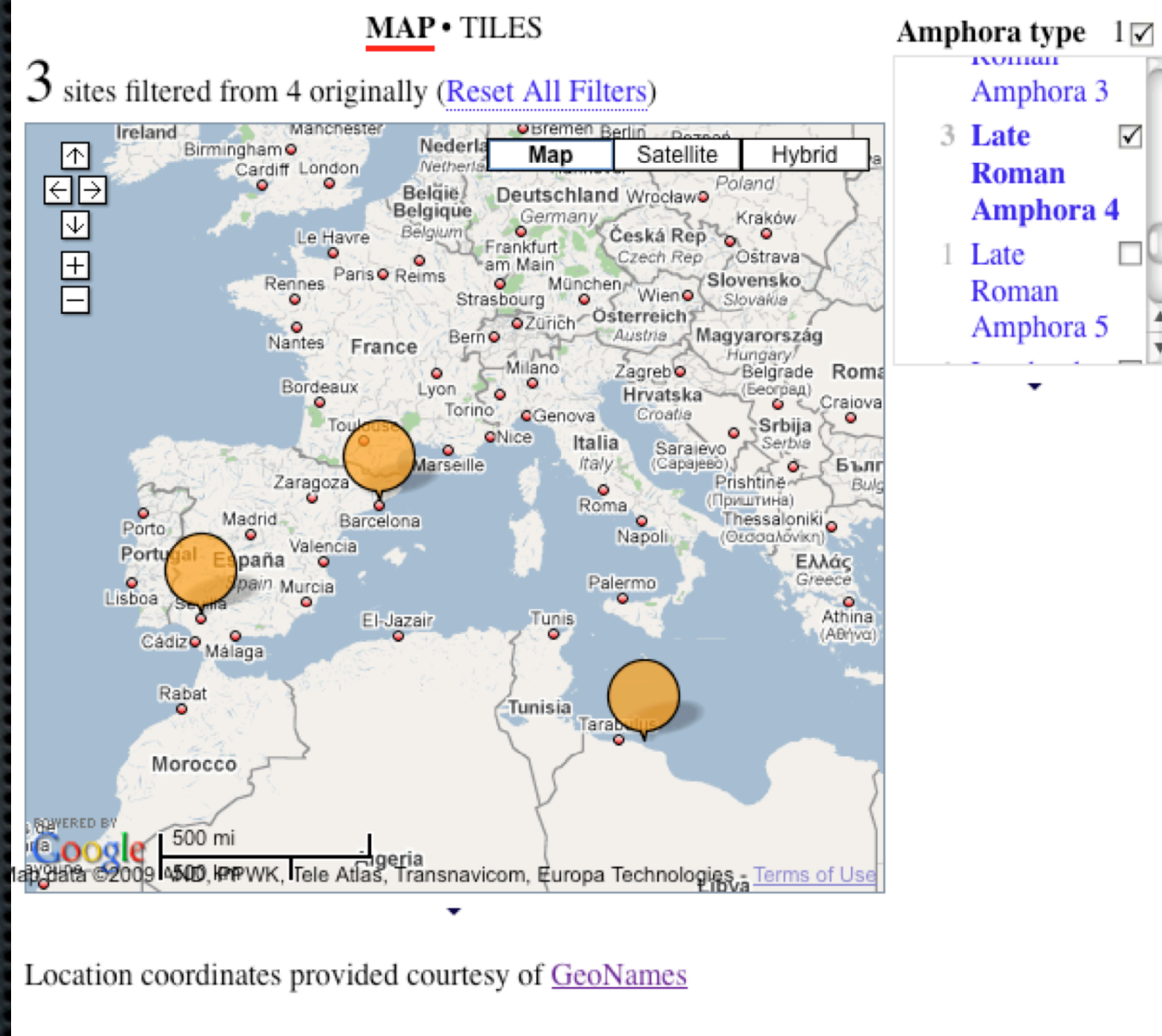
Type prediction

Training, Facilitation & Intervention

- ✦ online help
- ✦ Guides to best practice
- ✦ Recipe books
- ✦ <http://linkeddata.org/guides-and-tutorials>
- ✦ <http://pedantic-web.org/>

Comparison & Visualisation

Amphora Finds at Roman Harbours



‘Semantic Infrastructures in Archaeology’

Session at CAA 2010

Granada, Spain

6-9 April

Join us!

References

- Matthew Addis et al., 2005. “New Ways to Search, Navigate and Use Multimedia Museum Collections over the Web,” in *Museums and the Web 2005*. Vancouver, Canada
- Paul Cripps et al., 2004. “Ontological Modelling of the Work of the Centre for Archaeology” Centre for Archaeology, English Heritage.
- Philipp Nussbaumer and Bernhard Haslhofer, 2007. “CIDOC CRM in Action – Experiences and Challenges,” in *Research and Advanced Technology for Digital Libraries*.
- Frank M. Shipman and Catherine C. Marshall, 1999. “Formality Considered Harmful: Experiences, Emerging Themes, and Directions on the Use of Formal Representations in Interactive Systems,” *Computer-Supported Cooperative Work* 8, no. 4.
- Regine Stein et al., 2005. *Das CIDOC Conceptual Reference Model: Eine Hilfe für den Datenaustausch?*, Mitteilungen und Berichte aus dem Institut für Museumskunde Berlin: Staatliche Museen zu Berlin.