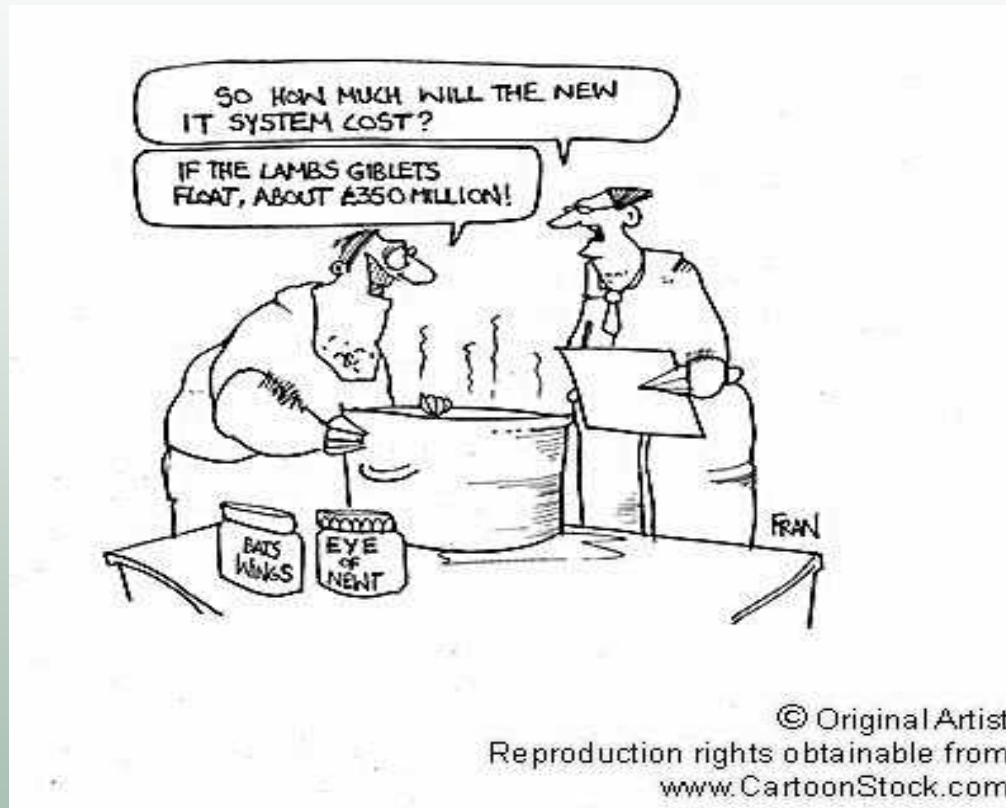


**Experiences of the CIDOC-CRM
at English Heritage**
**Methodologies for Ontological Modelling of
archaeological data**
Presented by Keith May

**Based on research work of English Heritage staff
especially Paul Cripps
and STAR project work of
Doug Tudhope, Ceri Binding
at Glamorgan University**



1. Background – why do it?



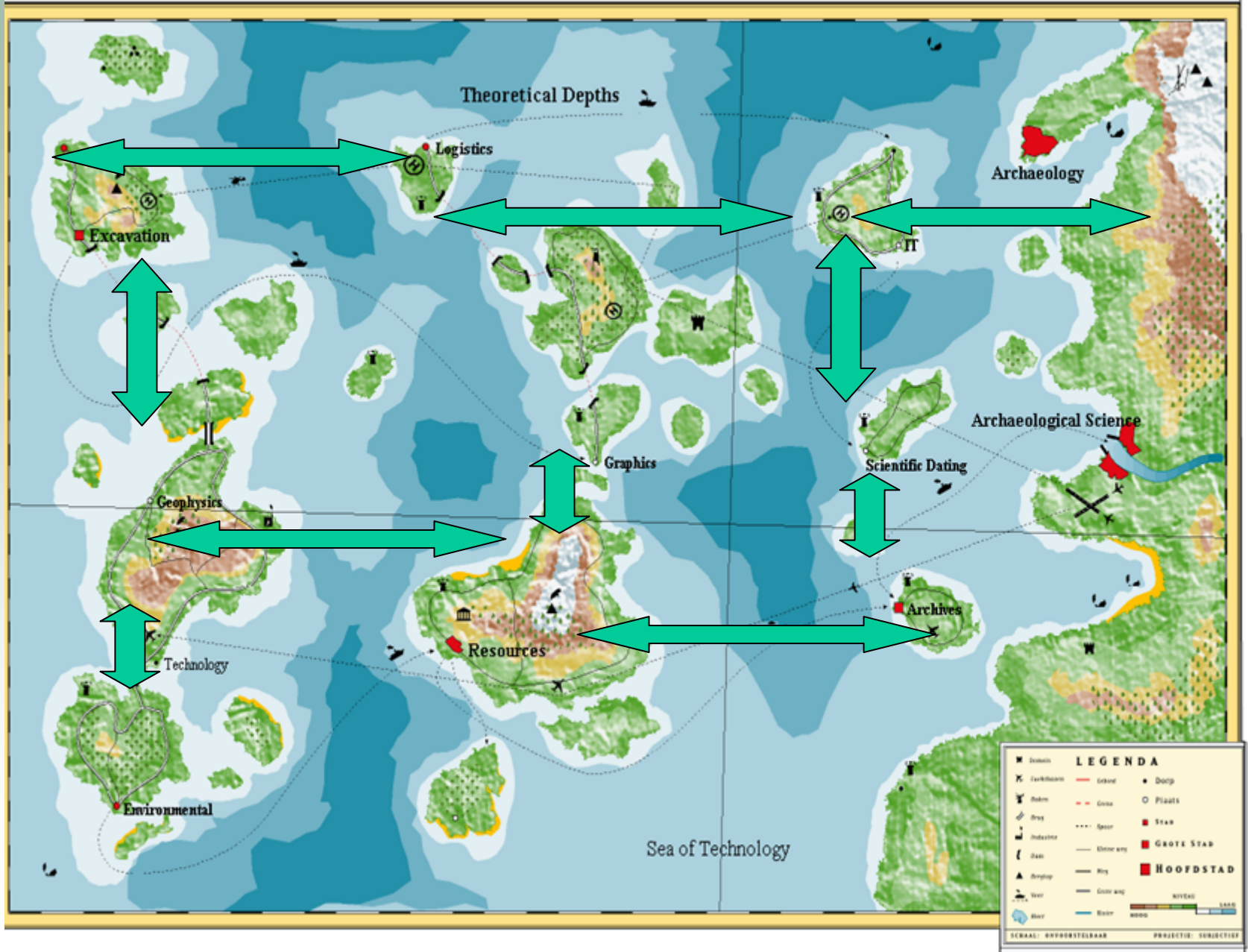
In the beginning - Revelation?

“Revelation” is an English Heritage project to provide a coherent digital information system that will make the capture, analysis and dissemination of our research faster and more effective.

Integrating archaeological data records



The Archaeological Archipelagos



Why the CIDOC CRM- Ontology?

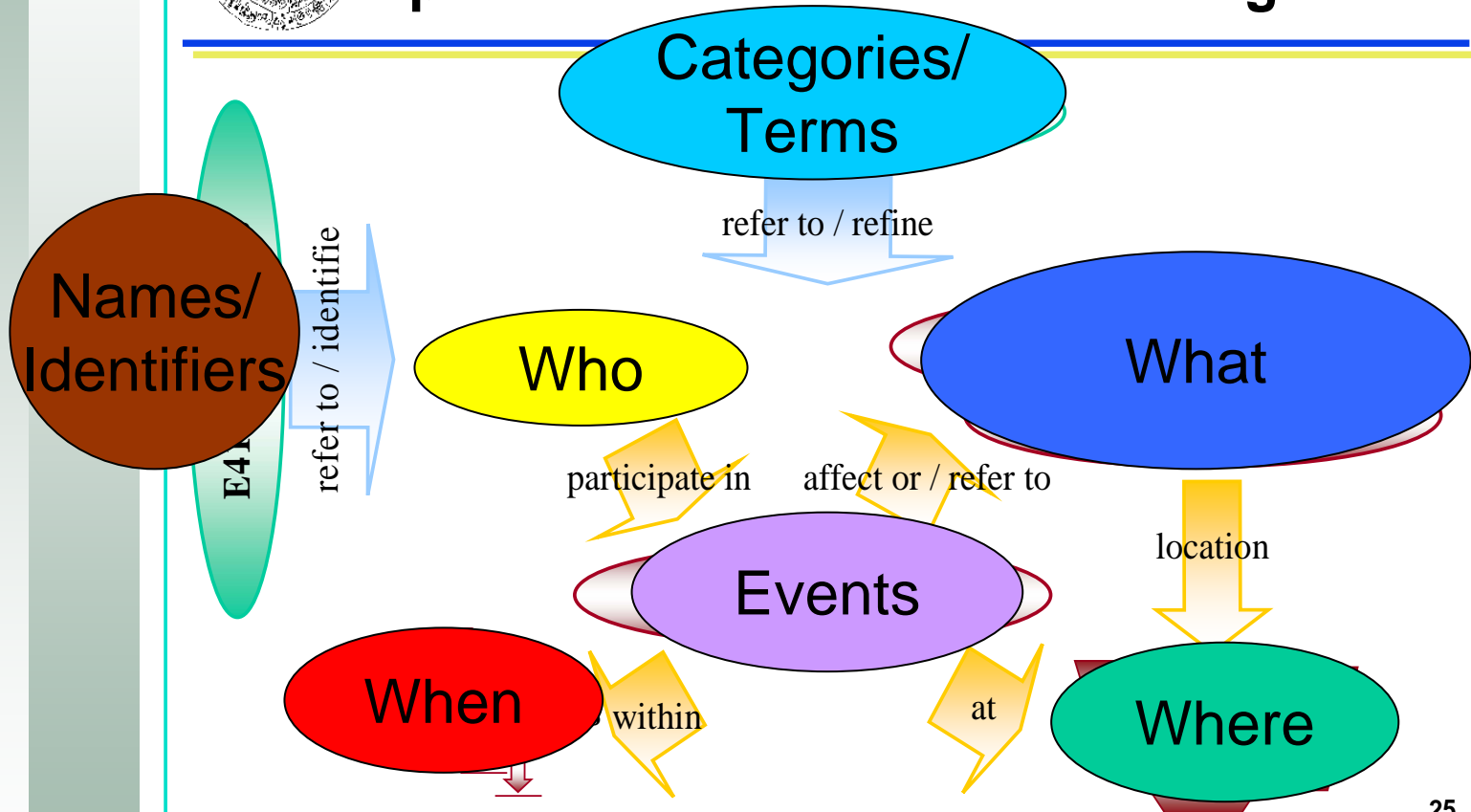
- Encapsulating & re-using domain expertise
- Shared understanding of *existing* information
- Build and integrate available entity modelling
- Modelling conceptual archaeological processes & events
- Enabling Semantic Web searching by non-domain experts
- Relating archaeology to other domains
- existing ontology provides greater standardisation and interoperability





The CIDOC CRM

Top-level Entities relevant for Integration



Practical Integration issues

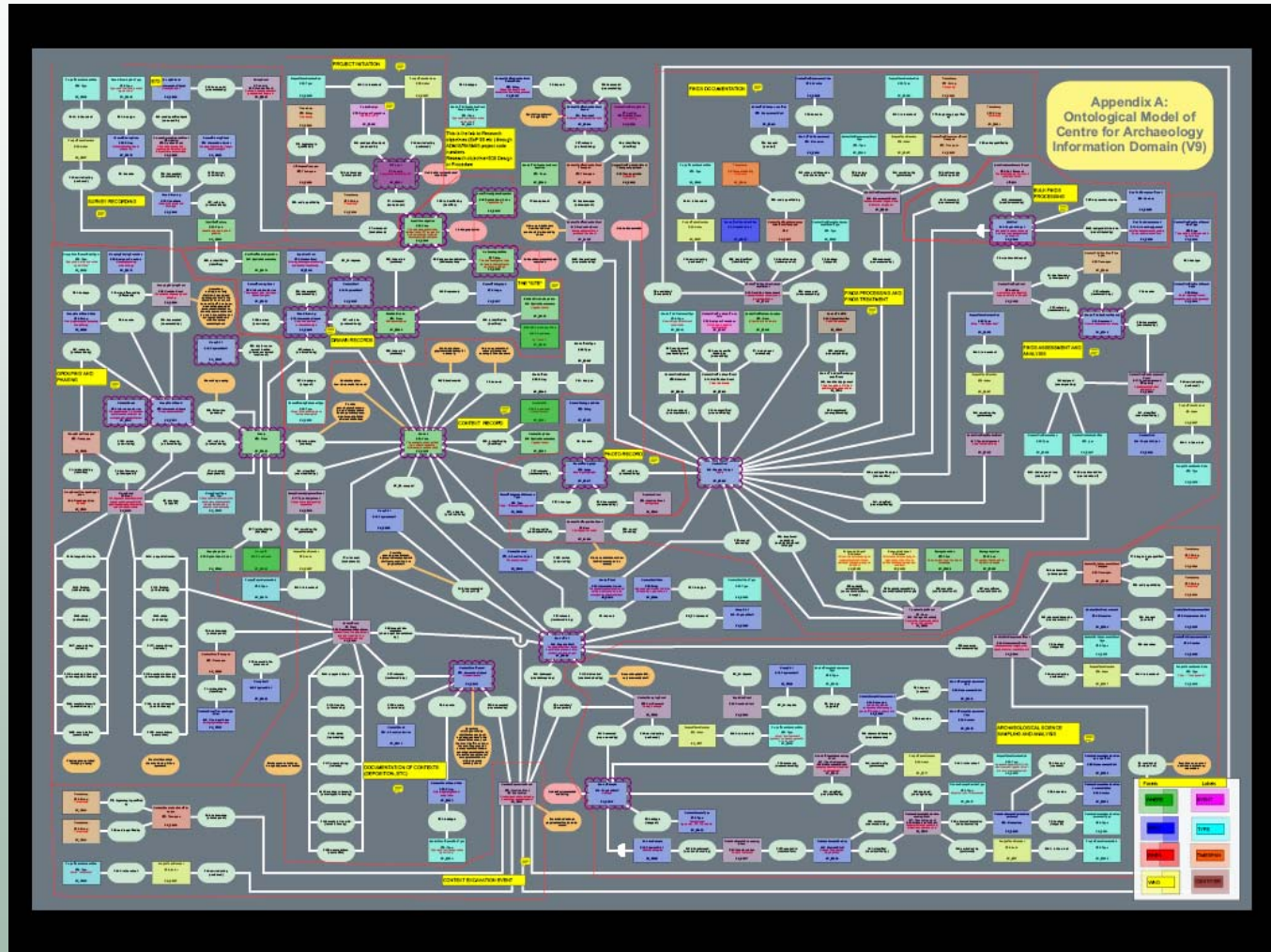
– old data & new requirements

- Some data in EH systems over 20 yrs old
- Need to produce a model that reflected continuing best archaeological practice
- Not fossilize existing systems
- *Model* new information requirements
- *Map* to old data as required
- Integration between old & new projects



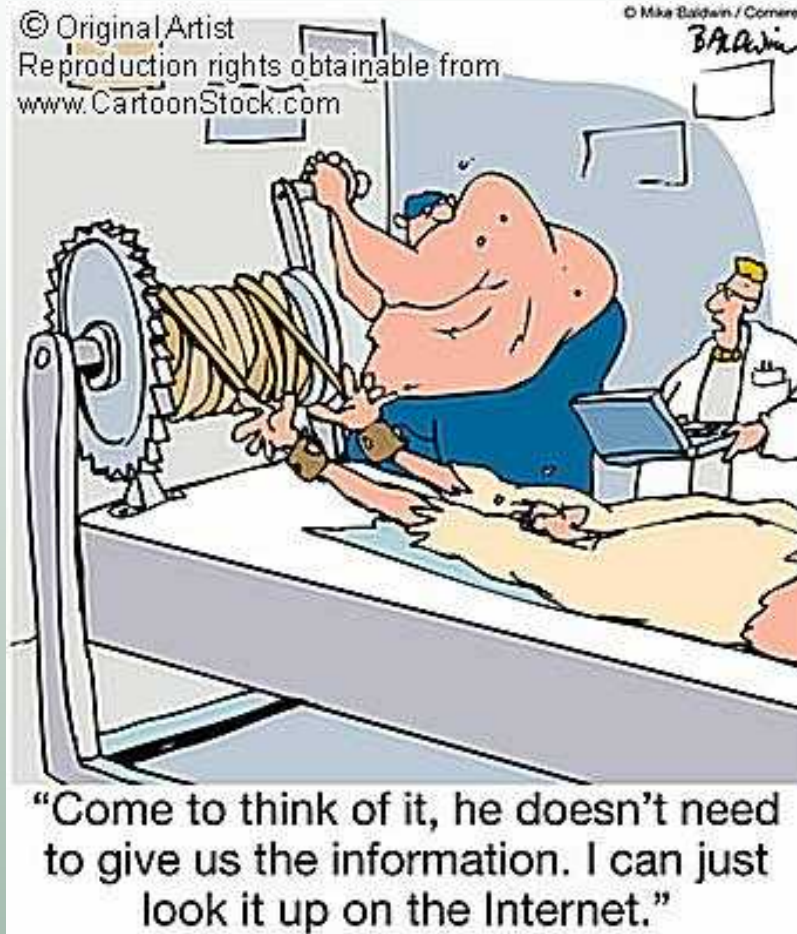
CRM diagram of Archaeological Information Domain

(http://cidoc.ics.forth.gr/technical_papers.html)



126 CRM-EH archaeological extension entities + 4 extension properties

2. Defining the Information Domain



The English Heritage Archaeological Information domain

- “Archaeologist” covers a range of ‘specialisms’
- Reflected in the Archipelagos problem
 - Field archaeologists
 - Geophysicists
 - Architectural Investigators
 - Field surveyors & Informatics
 - Finds specialists – many specialisms
 - Archaeological scientists – many specialisms
 - Aerial Survey
 - Cultural Resource Managers
 - Etc, Etc



Domain experts – Who decides?

- Interviews with key staff
- Review of existing systems documents
- Iterations of the model back and forth
- Modelling team staff from different specialisms
- But...
- In the end someone needs to make decisions
- Hopefully ‘moderator’ not ‘lone wolf’...
- ...Others agree

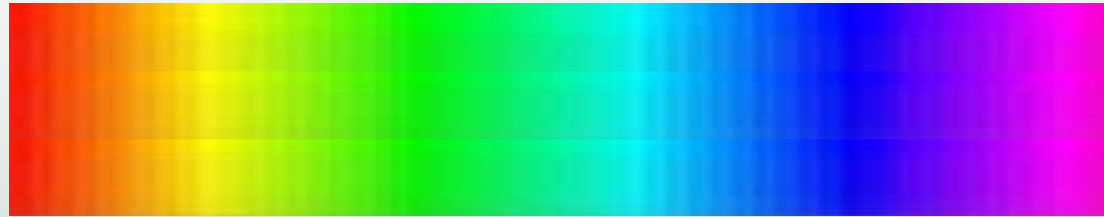


A picture of Order or Chaos?

- Need integrated systems for the archaeological process
- Contract archaeology's pressing requirements tend to be project specific
- EH - a more *research* driven approach (within limits)
- But not wishing to impose one simple way to record - others collect data in their own ways and systems and will continue to do so – *but hopefully demonstrate value*



Two ends of the archaeological spectrum in Work Practice & Knowledge Generation



- Preservation by Record
Most common approach to archaeological data collection in England

- Investigation by Record
Excavation strategies more research question driven, according to data assessed on site



General trends across Archaeological sector in UK

- Project focused approach to data
 - creates silos
- Divisions of the information within the process: Fieldwork-Analysis-Dissemination
- Planning geographic/spatial scope of organisational data helps in further integration



Field record data modelling

- Common ‘core’ of Arch process’
- Context record sheet modelled as CRM Information Object (E73)
- Limited degree of minute detail
- *Matrix stratigraphic relationships*
- Model already complex enough - most archaeologists find it a little daunting



176th 200x

DEPOSIT AND CUT FORM

Site Name	02. Project Code	A1. Year	01. Context No						
A2. Context type DEPOSIT CUT	05. Simple name		03. SSD						
04. Co-ordinates	E	N	E	N					
06. L	m	07. W	m	08. Diam	m	09. H/d	m		
DEPOSIT	12. Compaction								
10. Colour								; Munsell	
11. Texture									
13. Inclusions									
30. Contamination: Probable Possible Unlikely									
CUT	A3. Shape in plan		29. Orientation						
A4. Profile									
16. Comments									

Initial Interpretation ▼ STRATIGRAPHIC RELATIONSHIPS ▼ Revised Interpretation

<p>This context</p>	<p>This context</p>
40. Same as	
PHYSICAL RELATIONSHIPS	
34. Filled by	
35. Cut by	
42. Fill of	
43. Cuts	
EXCAVATION DETAILS	
31. Division of	32. Divided into
86. Method of excavation	A5. Weather
A6. Excavated by/date	68. Recorded by/date
61. Drawing Nos	A7. Checked by/date
63. Photo Nos	

176th 200x



Archaeological Context as 2 entities – Deposits (physical) & Cuts(spatial)

Context as a spatial entity - E53 Place

(e.g. pit cut)

(Cls(E1.CRM_Entity))

- (Cls(E53.Place))
 - (Cls(**Context_Class_EHE0007**))

Context as a physical entity

- E18 Physical Stuff

(e.g. pit fill)

(Cls(E1.CRM_Entity))

- (Cls(E77.Persistent_Item))
 - (Cls(E70.Stuff))
 - (Cls(E72.Legal_Object))
 - (Cls(E18.Physical_Stuff))
 - » (Cls(**ContextStuff_Class_EHE0008**))



Scope, Scale and Granularity issues

- Granularity – i.e. What level of modelling or mapping to go to?
- Do we need to maintain balance in the levels of granularity of the model?
- E.g. Do “finds” equate to “ecofacts”?



Relating different Vocabularies

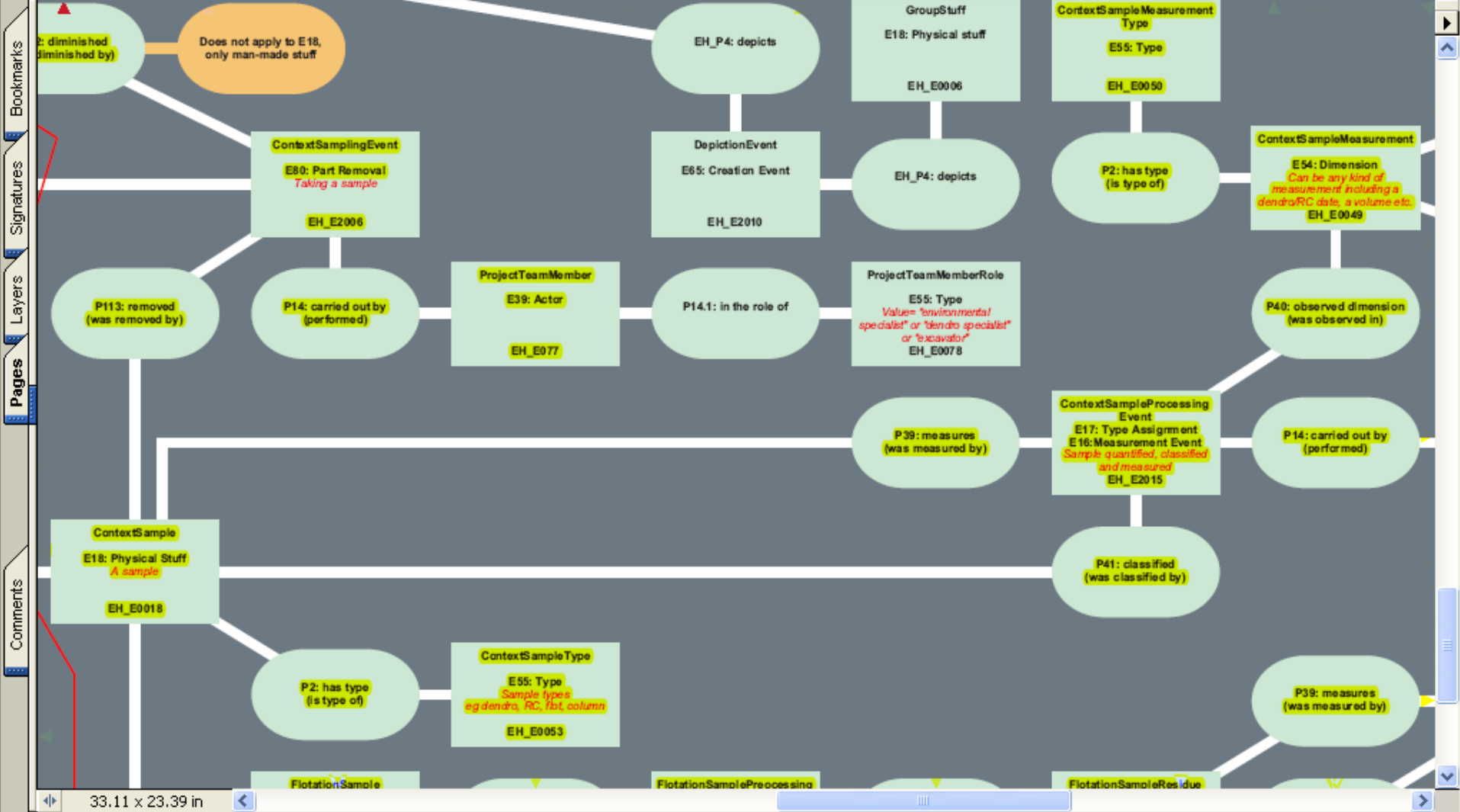
- Archaeological Science thesaurus
- Thesaurus terms Map to E:55 Type
- E.g. ContextSample *has Type* ContextSampleType
 - Dendro
 - Radio Carbon
 - etc



Open Save Print Create PDF Review & Comment Secure Sign

Select Text 119.38%

Note Tool Text Edits Show How To..?



3. Methodology for Ontological Modelling



CRM modelling project methodology – after M. Denny

- 1. Acquire domain knowledge - interviews**
- 2. Organize the ontological model
prototypes – software tools?**
- 3. Flesh out the ontological model -
iterate & add elements**
- 4. Check the work – versions & revisions**
- 5. Commit the ontological model -
disseminate**



Concept descriptions/scope

<i>EHProject - EHE0001</i>			
Concept	Project	CRM entity	E7: Activity
Notes	An archaeological project can be modelled as a Period, but as it results in a "change of state in the cultural, social or physical systems documented," the specialisation Event can be used. As it is intentionally carried out by people, the specialization Activity should be used.		
Relationship	To	Notes	
P7: took place at	EHE0003: AreaOfInvestigation (IsA E53: Place)	A project takes place at a particular location, described as a site, the extent of which represents the area of investigation.	
P16: used specific object	E29: Design or procedure	A project is defined by its project design (PD). Currently this is the only link to PDs although it is possible to use additional events to capture the process of writing and submitting PDs/UPDs according to procedural guidelines.	
P1: is identified by	E75:Conceptual Object Appellation	The project is identified by a project number UID	
P15: influenced by	E28:Conceptual Object	Research Aim	



Iterations and Extensions

- Iterations as part of the methodology
- Extensions to the CIDOC-CRM
- E.g.
 - EHE0003: AreaOfInvestigation (IsA E53: Place)



Some drawbacks encountered?

- Archaeology domain experts not very ontology aware – Don't use the 'O' word
- Methodology still being developed & tested
- Practical issues of producing, presenting and disseminating the model
- Getting wider participation & acceptance
- Still a degree of R&D

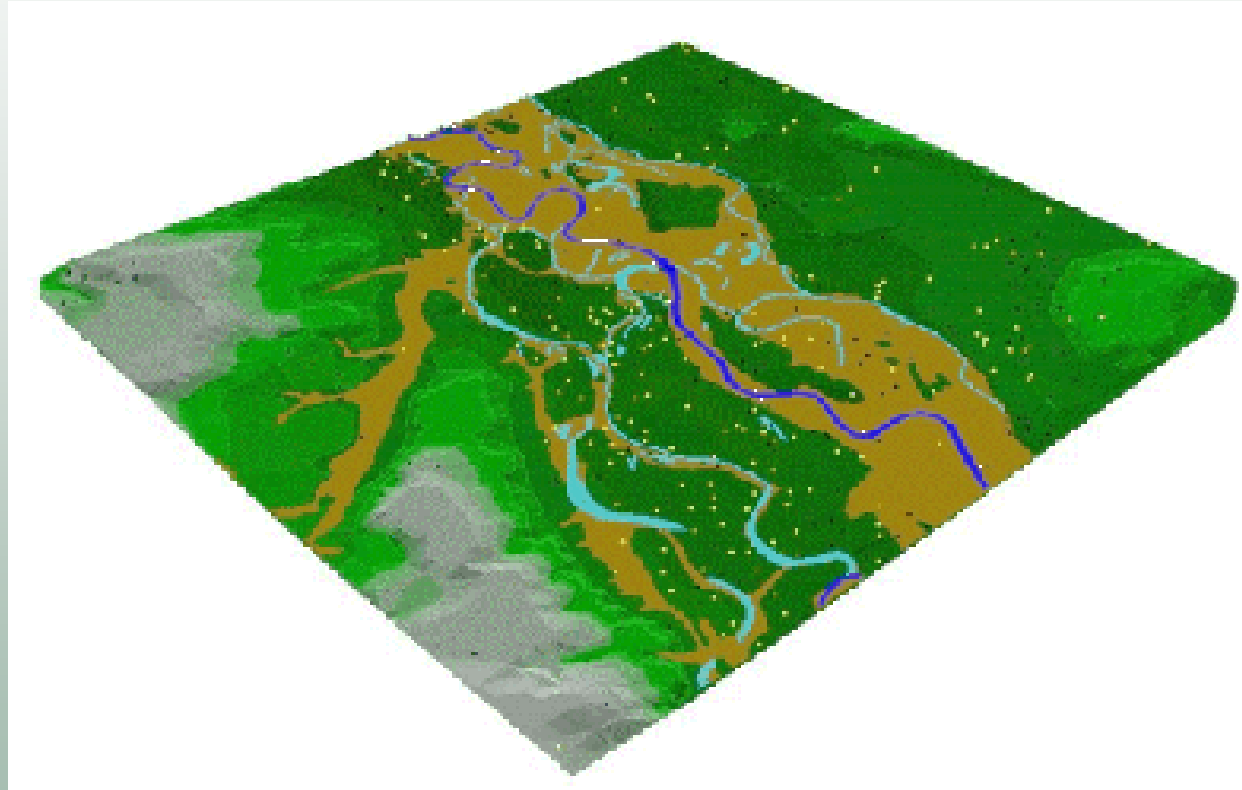


Positive outcomes of modelling

- Model is agreed across EH Archaeologists
- Applicable to wider archaeological sector
- Useful for other domains using CIDOC-CRM
- It serves as a information “map” for other EH KOS developments
- Makes explicit the links between the
Archaeology archipeligos
linking Field - Finds - Scientific analytical data



4. Modelling versus Mapping

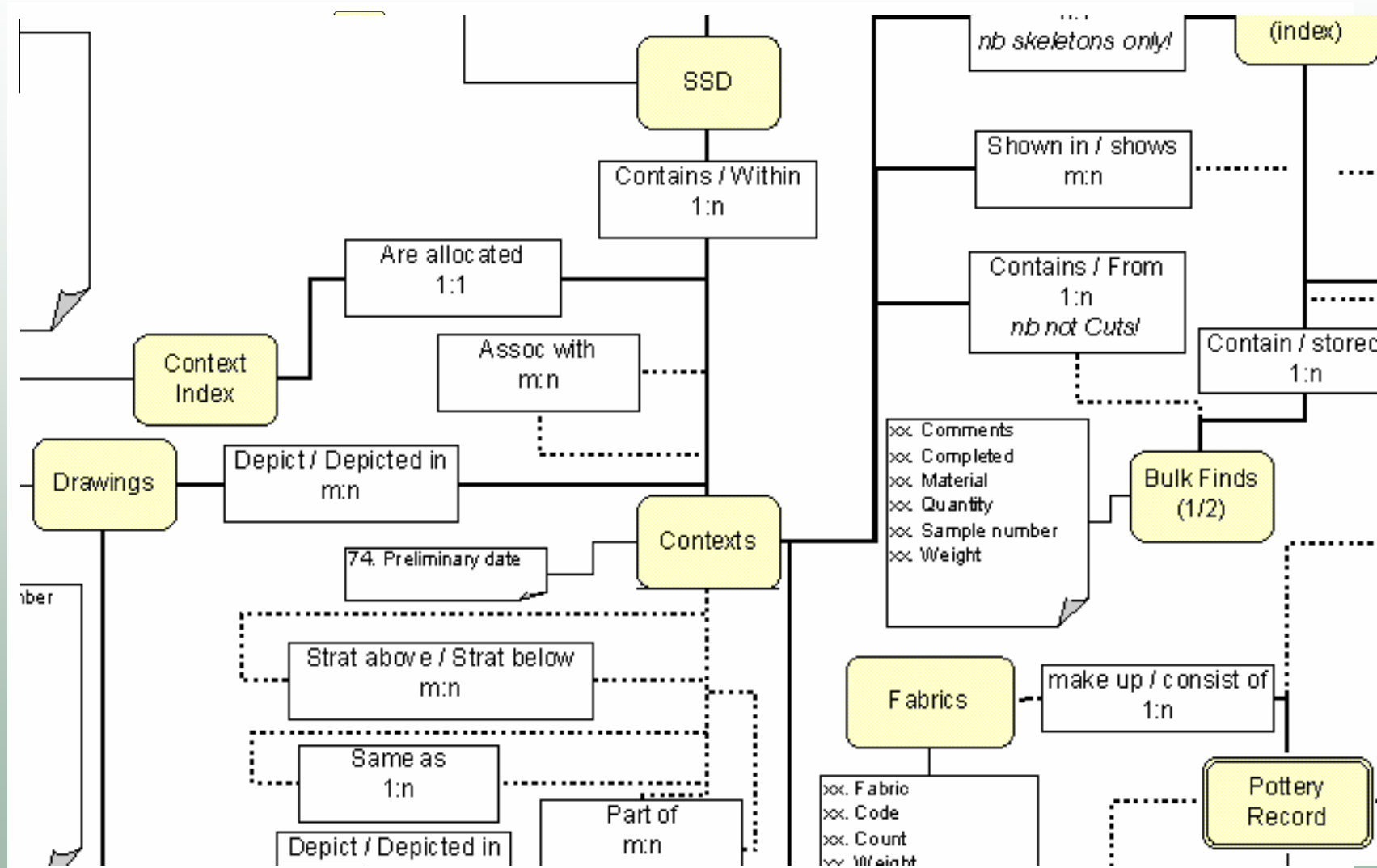


Model systems – Map to data

- *Model* new systems requirements
- *Map* to legacy or existing data records
- Mapping to Context recording system
- Representing different degrees of Granularity (detail)



Complex Entity Relationship Modelling

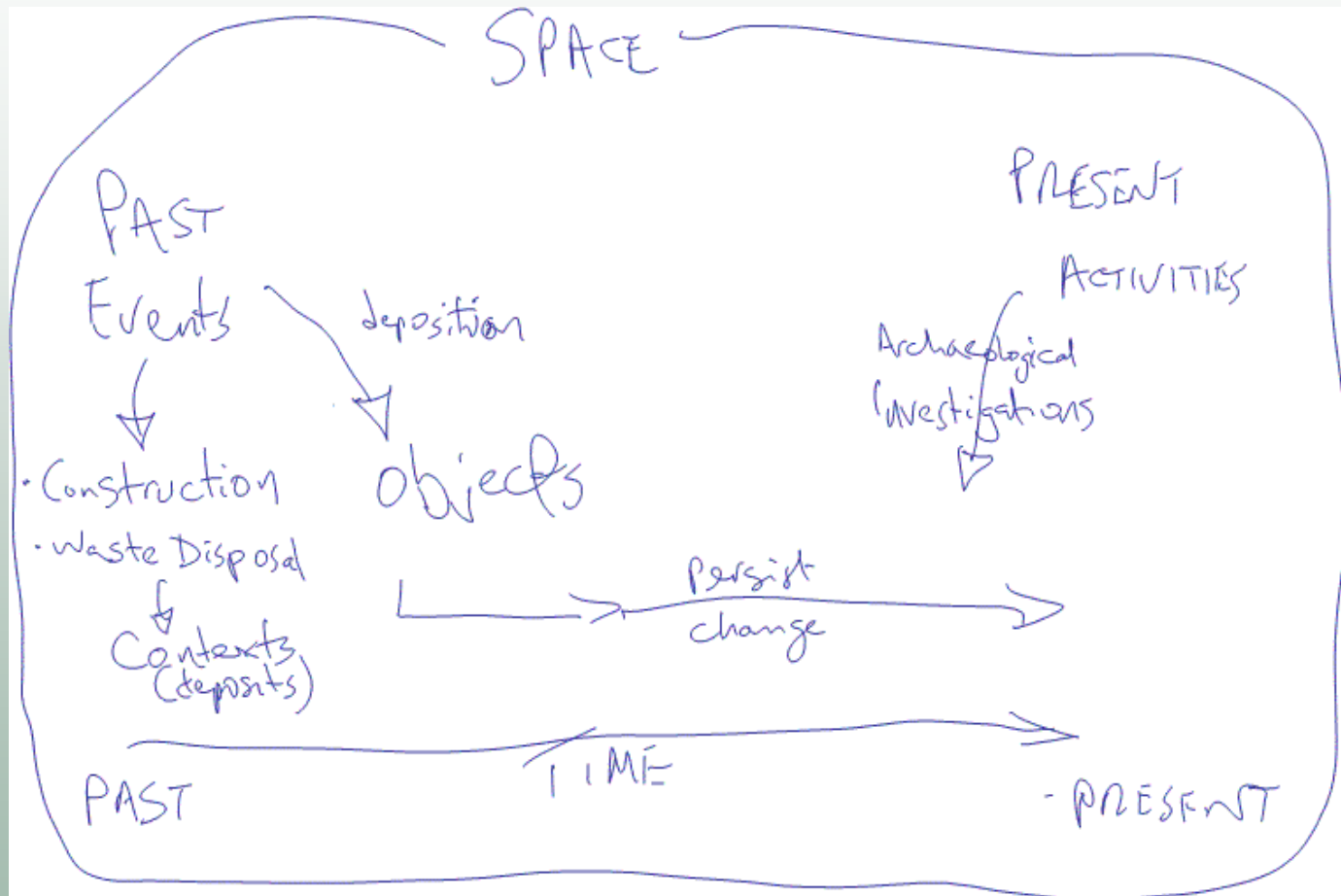


How to Model the archaeological process

- Events in the past result in remains in the present
- Activities in the present engage with and investigate the remains of the past
- Effectively two groups of events, one in the present, one in the past, related by the place in which they occur and the physical remains in that place.



Archaeological Processes





Events in the present

Excavation // Drawing and Photography
Survey // Sampling
Treatments & Processing
Classification // Grouping, & Phasing
Measuring, including scientific dating
Recording of observations
Dissemination // Interpretation/Analysis



5. Semantic Tools, Technologies and STAR (Semantic Technologies for Archaeological Resources)



New Mapping work - STAR

- Semantic Technologies for Archaeological Resources
- Building upon Ontological Modelling
- Some initial semantic modelling software tools available such as Protégé



Protégé – Dynamic annotations

The screenshot displays the Protégé 3.2 beta interface. The main window is titled "CombineOntologies Protégé 3.2 beta (file: D:\Documents\EH_laptop\PROJECTS\STAR%20-%20Doug%20Tudhope\RDFs_Ceri\CombineOntologies\C...". The interface is divided into several panes:

- SUBCLASS EXPLORER:** Shows the class hierarchy for "CombineOntologies". The class "crm:E7.Activity" is selected, and its subclasses are listed, including "ecrm:EHE2007.SurveyEvent", "ecrm:EHE2005.ContextFindDatin...", "ecrm:EHE2014.ContextFindUseA...", "ecrm:EHE1005.ContextFindUseE...", "ecrm:EHE0001.EHProject", "ecrm:EHE2008.ProcessSurveyD...", "crm:E10.Transfer_of_Custody", "crm:E11.Modification", "crm:E13.Attribute_Assignment", "crm:E65.Creation", "crm:E66.Formation", "crm:E8.Acquisition", "crm:E9.Move", and "crm:E52.Time-Span".
- CLASS EDITOR:** Shows the class "crm:E7.Activity" (instance of rdfs:Class). The "Annotations" pane displays a table of annotations:

Property	Value	Lang
rdfs:comment	hiérarchie des classes au-dessus de Événement. This class comprises actions intentionally carried out by instances of E39 Actor that result in changes of state in the cultural, social, or physical systems documented. This notion includes complex, composite and long-lasting actions such as the building of a settlement or a war, as well as simple, short-lived actions such as the opening of a door.	en
rdfs:label	Activité	fr
rdfs:label	Δράση	el
rdfs:label	Activity	en

The "rdfs:subClassOf" pane shows "crm:E5.Event" as a subclass of "crm:E7.Activity". The "Properties" pane displays a table of properties:

Property	Cardinality	Type
crm:P138B.has_represen...	Multiple	crm:E36.Visual_Item
crm:P140B.was_attribu...	Multiple	crm:E13.Attribute_Assign...
crm:P141B.was_assigne...	Multiple	crm:E13.Attribute_Assign...
crm:P15B.influenced	Multiple	crm:E7.Activity
crm:P17B.motivated	Multiple	crm:E7.Activity
crm:P1F.is_identified_by	Multiple	crm:E41.Appellation
crm:P2F.has_type	Multiple	crm:E55.Type
crm:P3F.has_note	Multiple	rdfs:Literal
crm:P41B.was_classified...	Multiple	crm:E17.Type_Assignment
crm:P4F.has_time-span	Multiple	crm:E52.Time-Span
crm:P62B.is_depicted_by	Multiple	crm:E24.Physical_Man-Me...
crm:P67B.is_referred_to...	Multiple	crm:E73.Information_Object
crm:P70B.is_documented...	Multiple	crm:E31.Document
crm:P7F.took_place_at	Multiple	crm:E53.Place
crm:P8F.took_place_on_o...	Multiple	crm:E19.Physical_Object
crm:P9B.forms_part_of	Multiple	crm:E4.Period
crm:P9F.consists_of	Multiple	crm:E4.Period

The Windows taskbar at the bottom shows the "start" button and various application icons, with the system clock displaying "13:37".



Protégé modelling – pros and cons

- + Pros: modelling much more updateable – dynamic
- + easy to disseminate in RDF formats for developers
- + Protégé is open source
- Cons: Existing PDF model complex enough but viewable
- Protégé is open source – difficult to maintain as a standard?
- Protégé networking – a whole further project at EH
- How much use to the wider Archaeological or Heritage community?



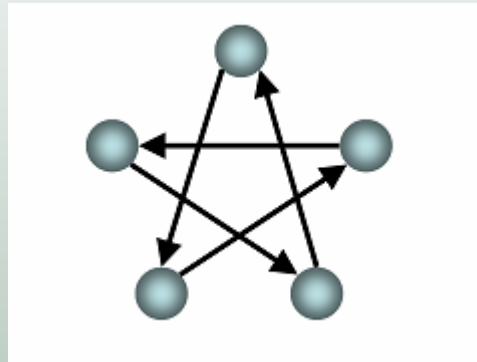
How do we Research Archaeological Grey Literature and excavation Reports?

If you *do* search Grey Literature, do you search for

- a. Exact Parallels?
- b. Finding ‘things’ in similar circumstances or similar types?
- c. Synthesis – finding broader associations for something?
- d. Background information on a subject/type/place, etc?
- e. Links to more detailed data sets that need more specific research?
- f. Other? – please specify

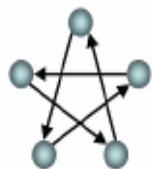


STAR - Semantic Technologies for Archaeological Resources

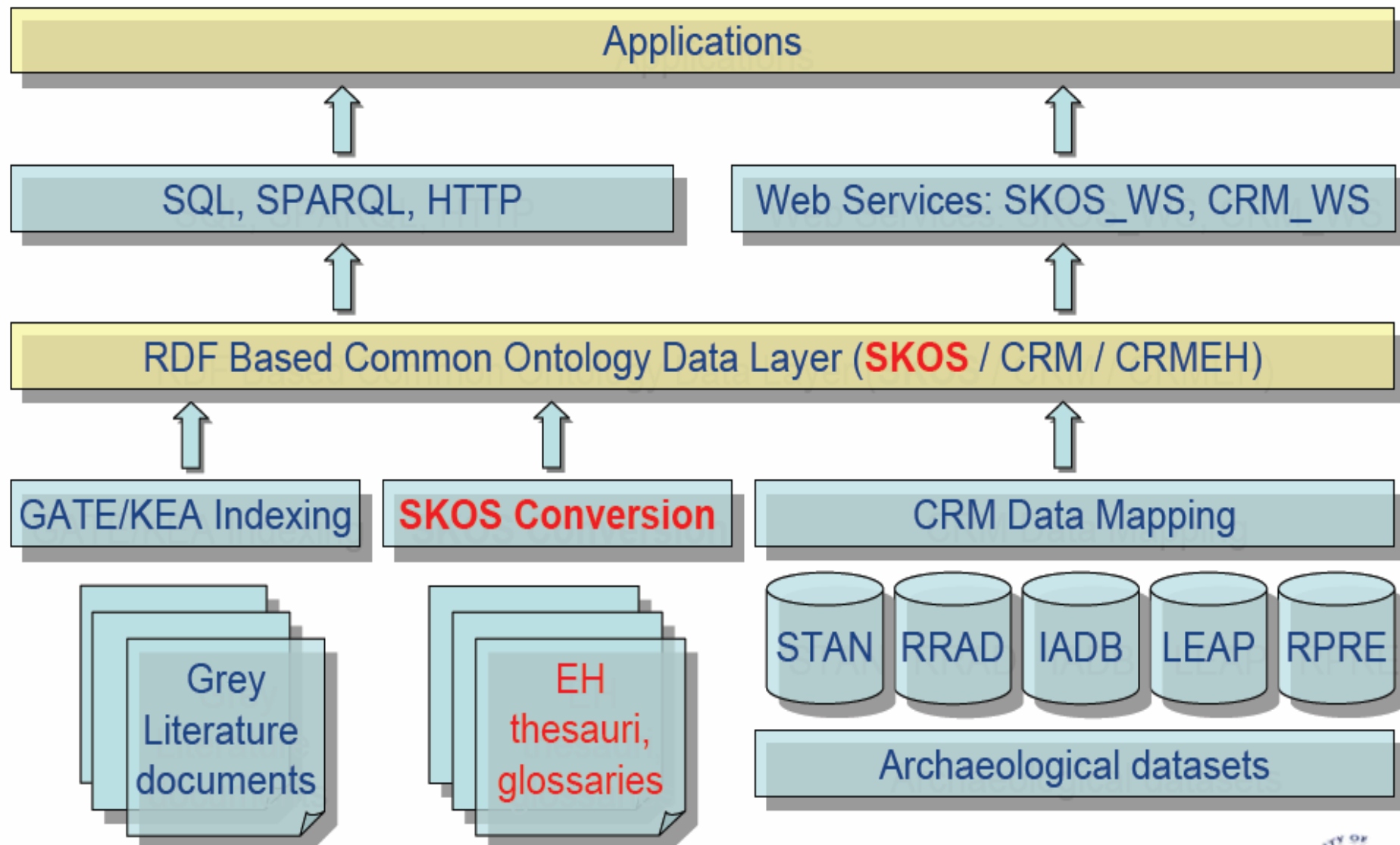


STAR - Semantic Technologies for Archaeological Resources

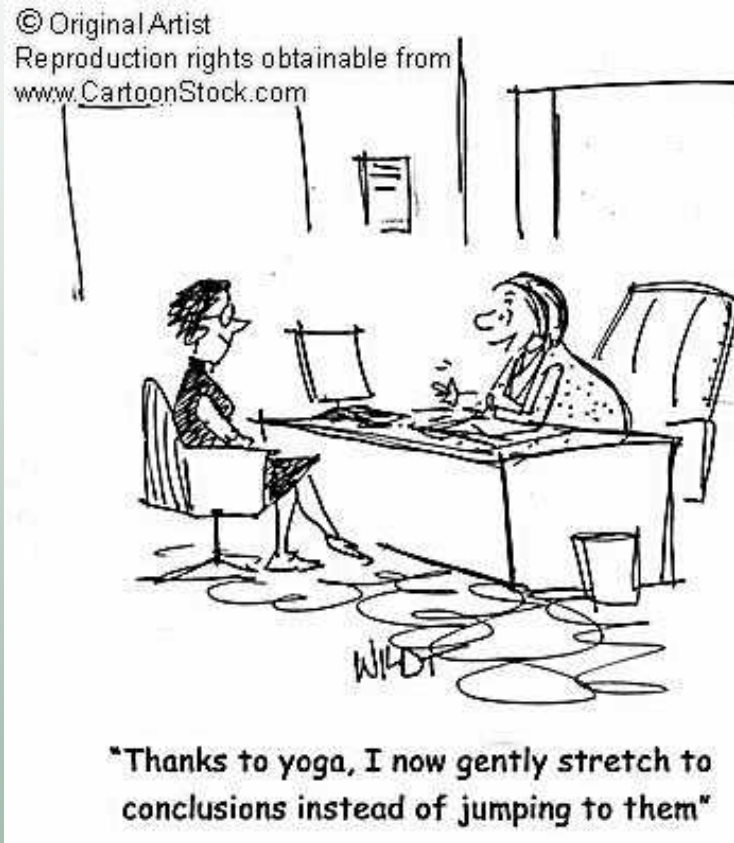




STAR - General Architecture



6. Conclusions and considerations for further work



Dissemination tools to better enable user endorsement of CIDOC-CRM

- Need for wider engagement with CRM
- EH hope other archaeologists adopt
- Need to identify “cost-benefits”
- Dissemination issues with size of model
- Need more graphical modelling tools
- Protégé helps – but not for archaeologists



How to enable semantic searching across/between Reports and Data?

- User Needs assessments
- How do people want to search?
- Example of a hyperthetical semantic search
- A seed from a plant of type W, found within the fill of a cut (type X), associated with a ritual burial of type Y, under or associated with a floor of a building of type Z
- Question – Could this room or building have been used for activities of a religious nature – e.g. a shrine?



Granularity issues

- Being explicit about the levels of entities within a model or mapping
- Do we need to explicitly show the current granularity of the model - (How) ?
- Explicitly defining the ‘granularity’ of your dataset as part of the metadata
- Can existing Artefact modelling (eg. pots & coins, etc) suffice for Ecofacts?



Verification & Dissemination

- Verified by CIDOC CRM SIG
- Published updated model with scope notes as RDFs online
- Other possible publication & dissemination routes ?
 - Interpreting Stratigraphy?
 - Internet Archaeology?
 - Others depending upon dissemination tools?



Bibliography / References

- Revelation assessment report – Cross et al, EH 2003.
- Denny, M. 2002. *Ontology Building: A Survey of Editing Tools*.
<http://www.xml.com/pub/a/2002/11/06/ontologies.html>
- CIDOC CRM v3.4.9 - <http://cidoc.ics.forth.gr/>
- CRM-EH Ontological model - Cripps, et al 2005
http://cidoc.ics.forth.gr/technical_papers.html
- STAR - <http://hypermedia.research.glam.ac.uk/kos/star/>

