

597 Norwegian excavation databases and CIDOC-CRM

—

a practical exercise

Ariandeplus workshop: Semantic mapping of excavation data 15.06.2022

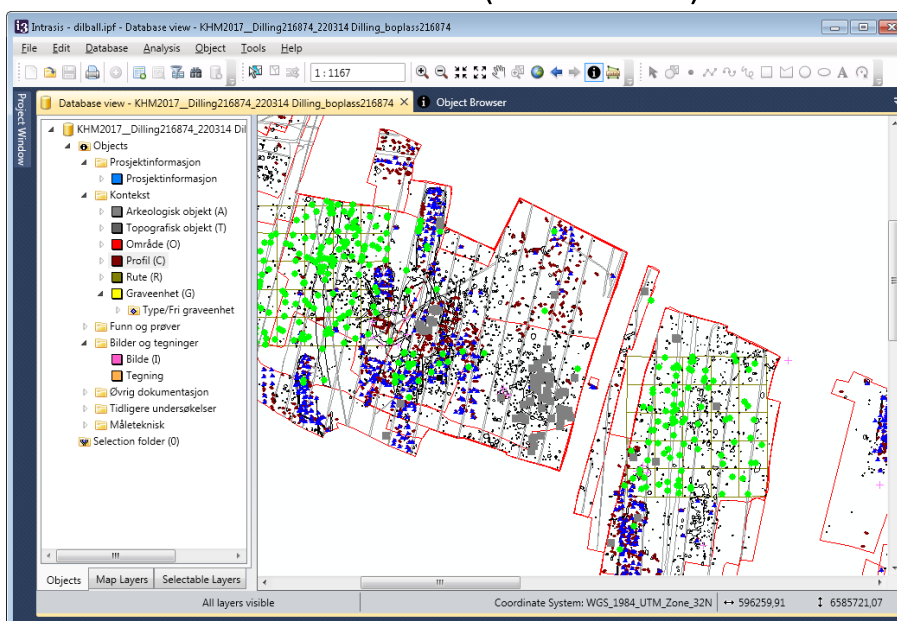
Christian-Emil Ore
University of Oslo



INTRASIS - excavation documentation

- Swedish system (INTRASIS.com)
 - Introduced around year 2000, extremely flexible
- For each instance one may define a template with
 - Class hierarchy
 - Class and subclass
 - Separate class for geobjects
 - A class can have named attributes
 - A subclass inherits attributes and may have its own
 - Attributes may take their value from
 - Closed vocabularies
 - Free values
 - Relations between classes (and subclass)

Excavation data sets (INTRASIS)



Situation in Norway

- Excavations and fieldwork
 - 5 museums, 14 counties, 1 research institute
 - No private commercial contract archaeology
- Since 2000 increasing number of excavations documented in the Swedish system INTRASIS (GIS), now the de facto standard
- In 2010 the museums decided to use a common INTRASIS template
 - has local variations
 - Evolves over time
- No common repository for INTRASIS or other excavation datasets

ADED

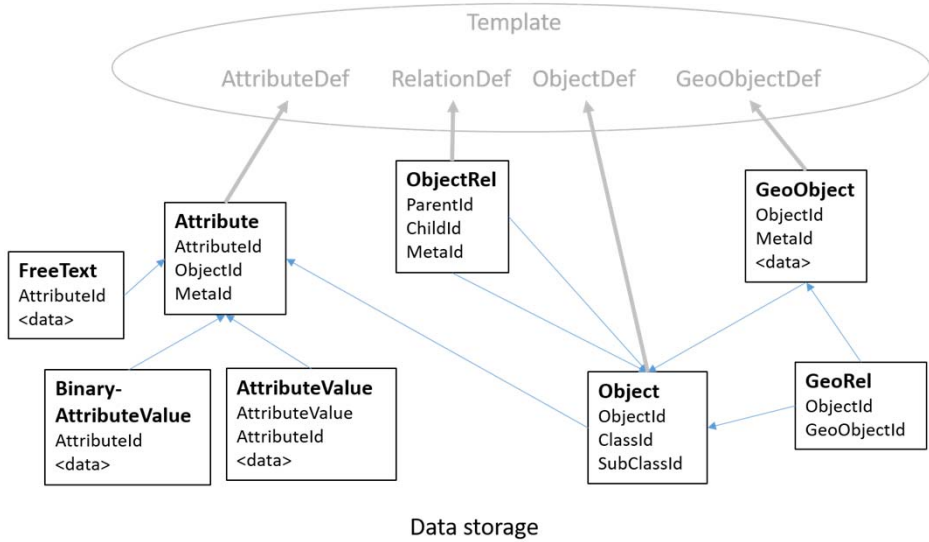
– Archaeological Digital Excavation Documentation project

- Create an infrastructure for digitally born excavation data sets in Norway
- The data sets are produced and stored in INTRASIS
 - Geo-references objects
 - Structures, profiles, layers, points
 - Structural relationships between object
 - Text and visual documentation of the objects
- The number of available datasets used in this project is around 600.
- The objectives is that the project should be a permanent activity

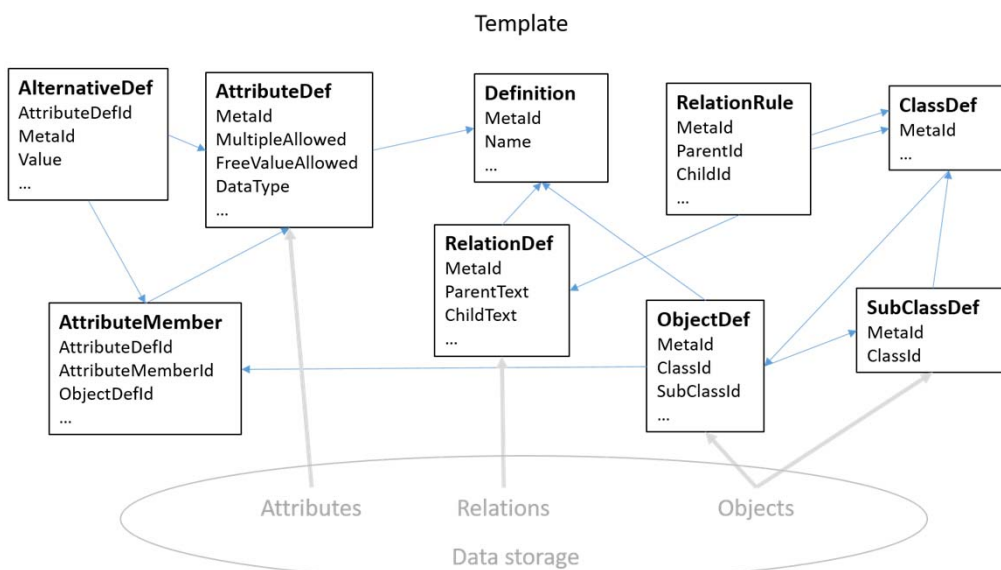
ADED Archaeological Digital Excavation Documentation project

- Tasks
 - The data sets are imported into into a single searchable information system PostgreSQL/PostGIS
 - Map the datasets to CIDOC CRM compliant data structures
 - xml intermediate format
 - rdf-triple store

INTRASIS – data storage



INTRASIS – semantic template



Mapping

- A mapping is defined on the basis of structural information in the template
 - Depending on the desired level of detail one may need to take into consideration the names and datatypes (and values) of the attributes
- The actual instances of classes, subclasses, attributes and relations are linked to the template via a few foreign keys
 - Once the mapping is defined, the production of rdf-triples (or other output) is an “industrial” process
- The snag:
 - The 597 templates have a lot of variations and a common template is needed
 - This required data cleaning and a conversion system. Not impossible, a conversion table consisting of 4 500 lines, required a month of work

A first analysis of the 597 INTRASIS templates

- Only two levels of classes
 - superclass and subclass
 - Relations are only between superclasses
 - From a model theoretical point of view subclasses can be interpreted as types
- A named attribute with values from closed vocabularies
 - represents a set of types
 - Each of the values is a type
- A named attribute with unrestricted values
 - Represents a set of strings
 - Each string will have the name and datatype of the attribute as types

Mapping: INTRASIS class to CRM class

Klassenavn	Class Name	CRM class	Type
Arkeologisk objekt	Archaeological Object	A8 Stratigraphic Unit	
Bergkunst_Motiv	Rock Art Motif	E73 Information Object	
Bergkunst_Område	Rock Art Area	S20 Rigid Physical Feature	
Bilde	Image	E36 Visual Item	
Båt	Boat/Ship	E22 Human-Made Object	
Dagbok	Diary	E73 Information Object	
Dokument	Document	E73 Information Object	
Funn	Find	E22 Human-Made Object	
Funnenhet	Find Unit	E22 Human-Made Object/ A8 Stratigraphic Unit	
Georeferanse	Geo reference	E73 Information Object	
Geo-objekt	Geo Object	E53 Place + E62 String (geojson)	
Graveenhet	Excavation Unit	S20 Rigid Physical Feature	
Hendelse	Event	E7 Activity	

Mapping: INTRASIS relations to CRM properties

Meta-Id	Relation name	CDOC-CRM domain	INTRASIS parent	CIDOC CRM properties	CIDOC CRM range	INTRASIS child
495	Has/is taken from	A8 Stratigraphic Unit	000015 Find unit	O4i sampled at (was sampling location of) : S2 Sample Taking. O5 removed (was removed by)	S13 Sample	000016 Sample
500	Shows/is shown at	E36 Visual Item	000021 Drawing	P138 represents (has representation)	A10 Excavation Interface	000022 Profile
503	cuts/is cut by	A8 Stratigraphic Unit	000011 Archaeological Object	AP7i produced (was produced by): A4 Stratigraphic Genesis. AP28i occurs before (occurs after) : A4 Stratigraphic Genesis. AP7 produced (was produced by) and / or simply P121 overlaps with	A8 Stratigraphic Unit	000011 Archaeological Object
...

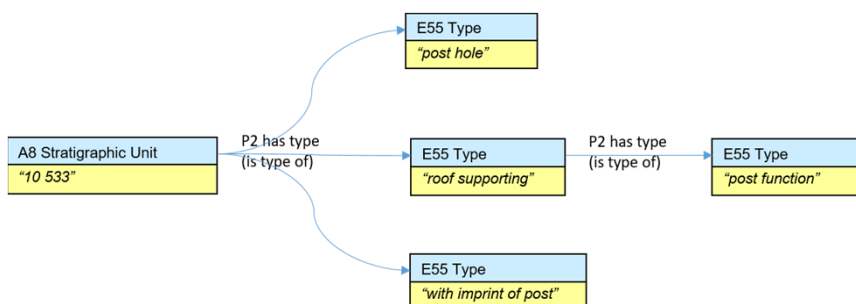
Mapping: attributes with values from a closed list

A structure interpreted as a post hole for a roof supporting post

In INTRASIS:

An object (id:10533) of the subclass "post hole" of the class "Archaeological Object" with the attribute "post function" with the value "roof supporting".

In CIDOC CRM:

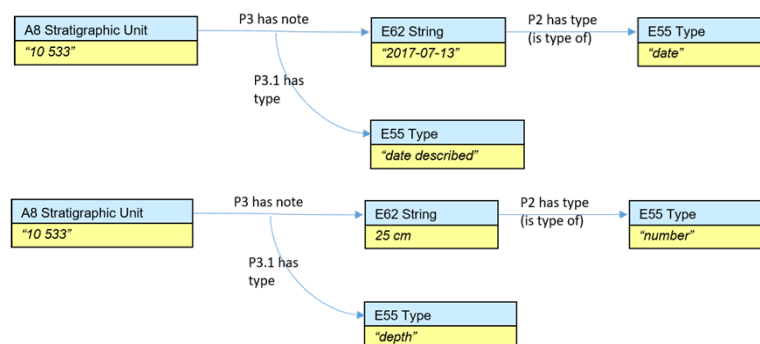


Mapping: attributes with free values

In INTRASIS:

An object (id:10533) of the subclass "post hole" of the class "Archaeological Object" with the attribute "described at" with the value "2017-07-13" and attribute "depth" with the value "25 cm".

In CIDOC CRM reflecting the INTRASIS structure:

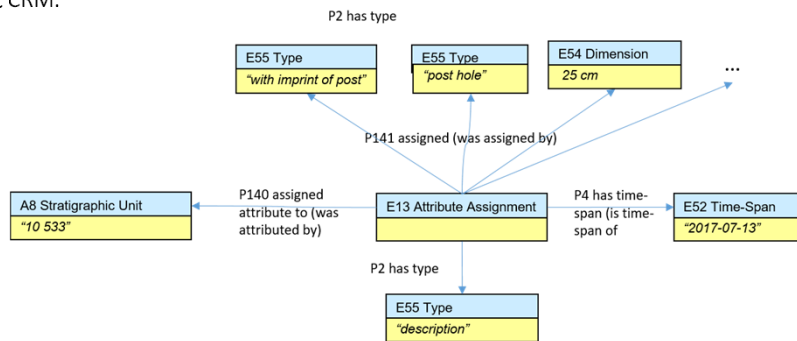


Mapping: attributes with events

In INTRASIS:

An object (id:10533) of the subclass “post hole” of the class “Archaeological Object” with the attribute “post function” with the value “roof supporting”, with the attribute “described at” with the value “2017-07-13” and attribute “depth” with the value “25 cm”

In CIDOC CRM:

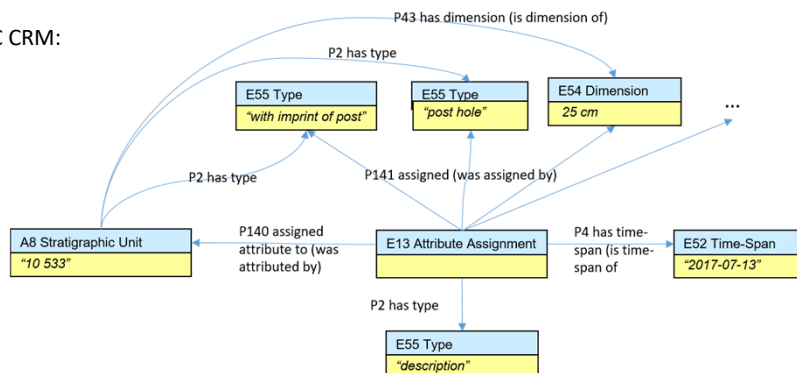


Mapping: attributes – ideal

In INTRASIS:

An object (id:10533) of the subclass “post hole” of the class “Archaeological Object” with the attribute “post function” with the value “roof supporting”, with the attribute “described at” with the value “2017-07-13” and attribute “depth” with the value “25 cm”

In CIDOC CRM:



XML (intermediate) target format

`<node universalid=... cid=... lid=... oid=... datatype=... (filename=...)>...</node>`

- universalid is the universally unique identifier created as described above
- cid is the class number for example 'E7'
- iid is the unique name of the INTRASIS database
- oid is ObjectId in the given INTRASIS database
- filename is a pointer to a file and is used for CLOBs and BLOBs

`<edge universalid=... pid=... iid=... relid=... did=... rid=.../>`

- universalid is the universally unique identifier created as described above
- pid is the property number for example 'P7' or 'P7i'
- iid is the unique name of the INTRASIS database
- relid is the internal INTRASIS identifier, that is, the triple `<MetalId, ParentId,ChildId>` which also is the complex primary key of the table `ObjectRel`
- did is the universally unique identifier for the parent object
- rid is the universally unique identifier for the child object

Considerations

- INTRASIS is geo oriented.
 - All class objects are georeferenced (x,y,z) as points, lines or (closed) polygons
 - Spatial relationships between objects can (in principle) be calculated in a 3D GIS system.
 - Thus most of the spatial relations are (in principle) redundant
- The current template is object centric and contains via attributes, information about
 - Structures, profiles, layers, finds, sampletaking
- The detailed catalogue information of the finds is stored in the museums' CMS and not in INTRASIS

Considerations – mapping to CIDOC CRM

- Information about the excavation process (who, when, what) is
 - hidden in the objects' text attribute "description"
 - Implicit in via "date described" and the classification attributes.
 - A detailed mapping to CIDOC CRM requires a detailed analysis of free text
- Map the geo-data to CRMgeo or easier and more relevant to use GIS-systems?
- As shown, it is not complicated to map the information about structures and objects in Intrasis to CIDOC CRM

Considerations – most important caveat

- To link
 - Excavation datasets
 - artefact information (museums),
 - Site and monument information
 - report archives
- One needs
 - Unique identifiers denoting the same item in all databases
 - Catalogue information connected to common authorities like AAT, TGN, GeoNames...
- Slogan: data cleaning and digital discipline