GLOBALISE

Data modelling at the start of the project - Pilot

GLOBALISE Symposium, Leon van Wissen (l.vanwissen@uva.nl), June 27, 2021

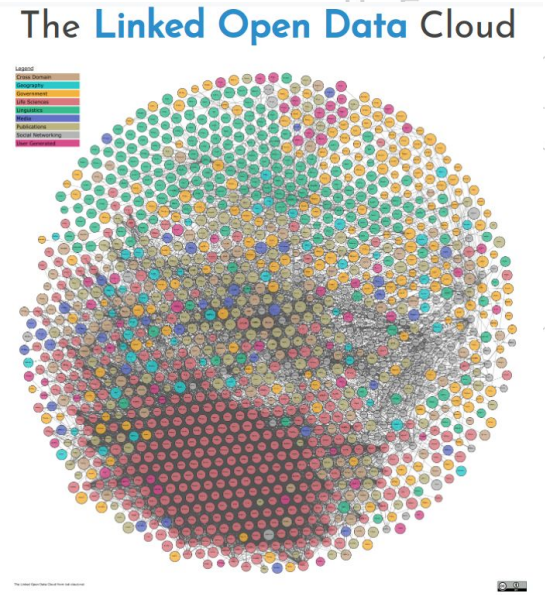
Work package contents

Semantic Contextualisation

- Provide *machine readable* entry to our (reference) data (WP3, Historical Contextualisation)
  → needed for question answering in our infrastructure
- Building a 'Knowledge Graph'
  → needed for interpret results (relation with *the world*)

And, cf. Linked Open Data

- Integration with other projects
  - E.g. NDE's Dataset Register
- Making our data linkable for similar/other domains
- Linking to other datasets
- Authority in the domain of the VOC



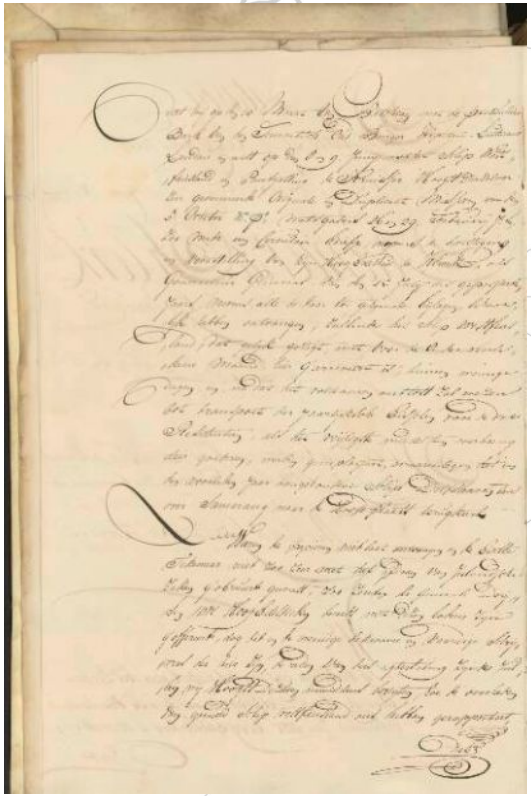
https://lod-cloud.net/

Focus: Data modelling

Example material

- https://www.nationaalarchief.nl/onderzoeken/archief/1.04.02/invnr/8141/file/NL-HaNA_1.04.02_8141_0014

Wij bedienen ons van het **vertrek** van het **Zeilree**
leggende schip **Delfshaven** , als de eerste voorkomende
gelegenheid om uwe Hoog Edelheden te **verwittigen** ,
dat wij op den **30 : Maart** Van **Batchian** met de partenlere
Bark van den **Ternaatsch Oud Burger Kapitein Luitenant**
vandan en eerst op den d en **9 . Junij** met het schip **West** „
„ **Friesland** en **Pantijalling** de **Kruisser** Hoogst Derzelver
heer gevenereerde **Originele en Duplicaat Missiven** van den
5 : October d „ o p „ r mitsgaders 2**6 en 29 . Februarij** 7 : 6 ,
zoo mede een Circulair briefje , nopens de bevestiging
en voorstelling van zijn **Hoog Edelheid de Klerk** , als
Gouverneur Generaal , van den **25 . Julij** des gepasseerden
Jaars , nevens alle de daar toe gehorende bijlagen , behoor „
„ lijk hebben **ontvangen** ; zullende het schip **westfries**
„ **land** , dat gelijk gezegt , niet voor de **8 . der verwe** „
„ **kene maand** , hier g ' **arriveert** is , binnen weinige
dagen , en , na dat het volkomen **ontlost** zal wezen
tot **transport** der Jaarlijksche Eisschen voor de **drie**
Residentien , als het veijligste middel ten **overbreng**
nter goederen , werden **gemploijeert** , waarentegen het in
den voorleden Jaar aangehoudene schip **Delfshaven** lans
over **Samarang** naar de hoofd plaats **terugkeert**



How?

Among others entities observed:

- Dates
- Places
- Ship names
- Situations / events
    - Pilot: focus on ship movements

Needed:

- Event-based model
    - *Common* in DH/GLAM
- Flexible / generic for now
- Integration with NLP part (VU) in the project

Ontologies and vocabularies

Off-the-shelf ontology

- Simple Event Model (sem)
  - Event-based
  - Generic
  - Tested

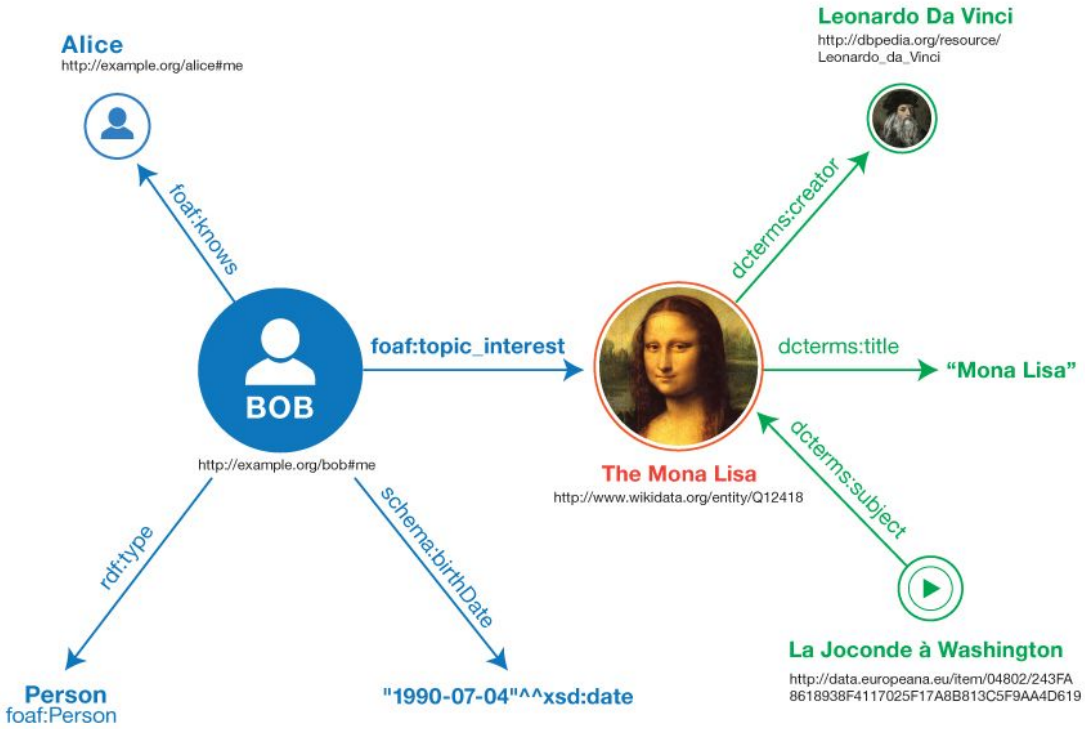
And our own thesaurus

- Thesaurus in SKOS
  - cf. CIDOC-CRM approach

Compare: https://dutchshipsandsailors.nl/

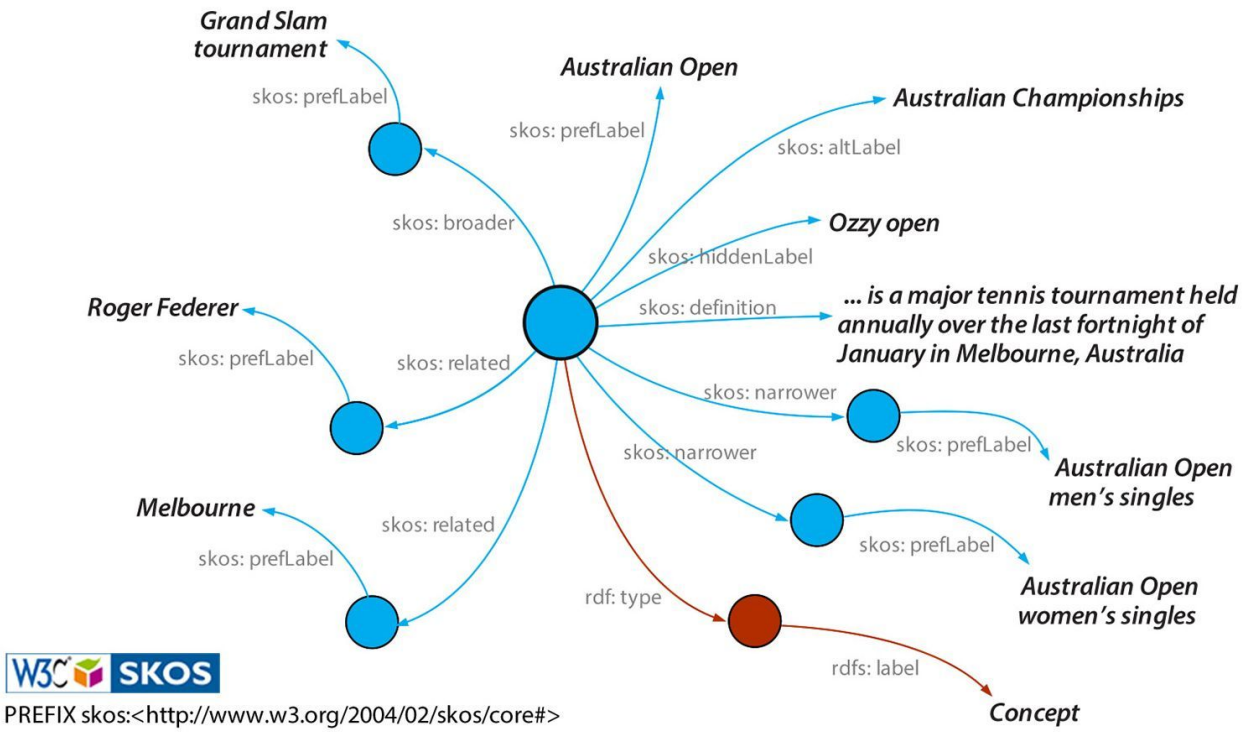
- 2014 and 2021 version

Resource Description Framework (RDF)



Taken from: https://www.w3.org/TR/rdf11-primer/

Simple Knowledge Organization Scheme



Taken from: https://www.poolparty.biz/skos-and-skos-xl

SEM Classes

https://semanticweb.cs.vu.nl/2009/11/sem/

Event

- Date properties
  - timestamp in xsd:date [=machine readable!]
  - (earliest/latest) begin/end

Actor / Object

- Link to reference data

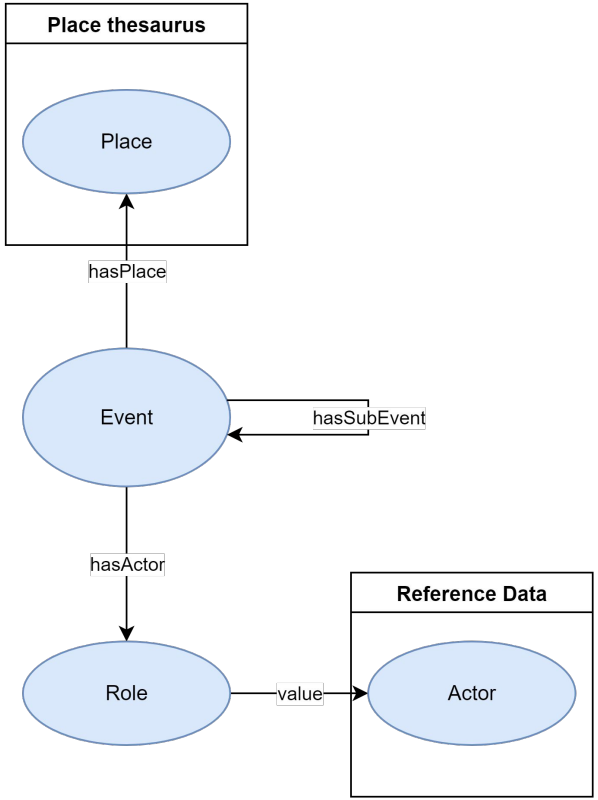
Role

- Can be used for contextual information on the actor

Place

- Link to place thesauri

Question: Should we use the View class of sem?



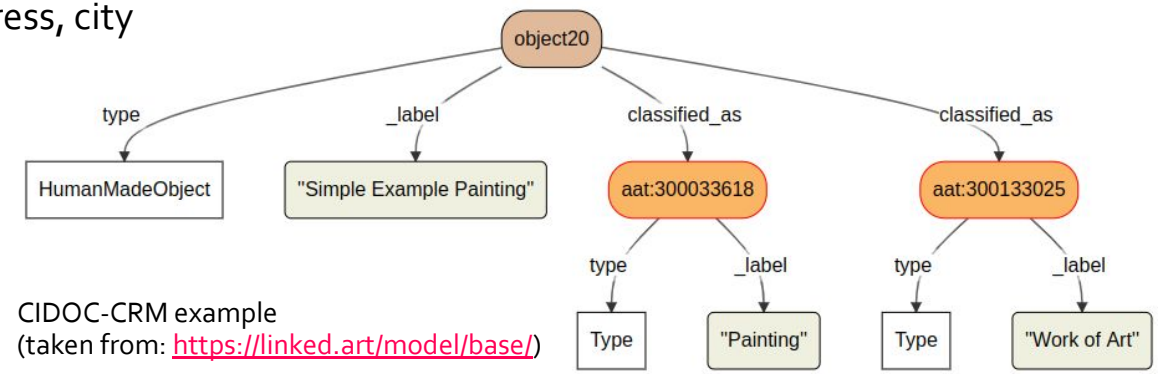
Adding Types

Focus on *vocabulary* not *ontology*

Stick to generic concepts 'Event', 'Actor', 'Role', 'Place'

Our types

- Event types: e.g. departure, arrival, voyage, transport
- Actor types: e.g. person, ship, polity
- Role types: e.g. captain, transporting vessel, cargo
- Place types: e.g. harbour, fortress, city



CIDOC-CRM example
(taken from: https://linked.art/model/base/)

Research

Research Home ▸ Tools ▸ Art & Architecture Thesaurus ▸ Full Record Display

Art & Architecture Thesaurus® Online
Full Record Display

New Search Previous Page Help

Click the icon to view the hierarchy.

Semantic View (JSON, RDF, N3/Turtle, N-Triples)

ID: 300162867
Page Link: http://vocab.getty.edu/page/aat/300162867

Record Type: concept

journeys (events (activities), Events (hierarchy name))

Note: Courses of travel, having a beginning and end in space and time, and and taken as a whole. Extended to excursions, expeditions or tours.

Terms:
journeys (**preferred**,C,U,LC,English-P,D,U,U)
journey (C,U,English,AD,U,U)
trips (C,U,English,UF,U,U)
voyages (C,U,English,UF,U,U)
reizen (gebeurtenissen) (C,U,Dutch-P,D,U,U)
reis (gebeurtenis) (C,U,Dutch,AD,U,U)
viajes (C,U,Spanish-P,D,U,PN)
viaje (C,U,Spanish,AD,U,U)

Facet/Hierarchy Code: K,KM

Hierarchical Position:
Activities Facet
.... Events (hierarchy name) (G)
..... events (activities) (G)
..... journeys (G)

Related concepts:
meaning/usage overlaps with .... **travel**
..... (culture-related concepts, <culture and related concepts>, Associated Concepts (hierarchy name)) [300080094]

http://vocab.getty.edu/aat/300162867

GLOBALISE Pilot datasets

Boekhouder Generaal Batavia (BGB, https://bgb.huygens.knaw.nl/)

- Transport of goods
- Name of the ship (not unique)
- Cargo

Dutch-Asiatic Shipping (DAS, http://resources.huygens.knaw.nl/das/)

- Voyage + Stopover
- Ship (unique)
- Master

Generale Zeemonsterrollen (GZM, Van Rossum 2014)

- Mustering
- Name of the ship
- Crew (counts)

Example: https://bgb.huygens.knaw.nl/bgb/voyage/2865

Boekhouder Generaal Batavia (BGB)

Voyage Details

Number	2865
Book year	1776/1777
Source	10793
Folio number	60
Ship name	Amsterdam
Departure date	1-11-1776
Departure place and region	Batavia, Batavia
Arrival date	-
Arrival place and region	Amsterdam, Republiek
Total value Dutch guilders	131.248,7,8
Total value Indian guilders	-
Remarks Voyage	-
Voyage in DAS	go to DAS voyage 7976.1

Cargo details

Quantity		Product	Specification	Value Dutch guilders	Value Indian guilders
6.800	lb	kaneel	-	1.963,10	
10.318	lb	foelie	-	5.107,8	
65.700	lb	garioffelnagel	-	21.681	
24.165	lb	nootmuskaat	in soort	3.987,4,8	
194.745	lb	peper	in soort	26.945,4	
1.000	lb	benzoë	-	749,19	
12.875	lb	garen	katoenen Javaans	5.677,15,8	

Example: http://resources.huygens.knaw.nl/das/detailVoyage/98796

Dutch-Asiatic Shipping (DAS)

Details of voyage 7976.1 from Batavia to Texel

Number	7976.1
Name of ship	AMSTERDAM
Master	Verburg, Jurriaan
Tonnage	1150
Type of ship	
Built	1772
Yard	Amsterdam
Chamber	Amsterdam
Date of departure	02-11-1776
Place of departure	Batavia
Arrival at Cape	14-01-1777
Departure from Cape	07-03-1777
Date of arrival at destination	19-06-1777
Place of arrival	Texel
Chamber for which cargo is destined	Amsterdam (131,223)
Particulars	The 6 passengers included 2 female slaves.
Previous outward voyage	4185.1

On Board	I	II	III			
Seafarers	108	8	6			
Soldiers	11					

Conversion to RDF

Data available as *database dumps* in csv

- Plus: every entity has a unique id already
- Though, not everything is what it says it is…

Conversion to RDF using RDFLib (Python library)

- Pipeline (metadata + data)

Data (will be) permissively available on our GitHub

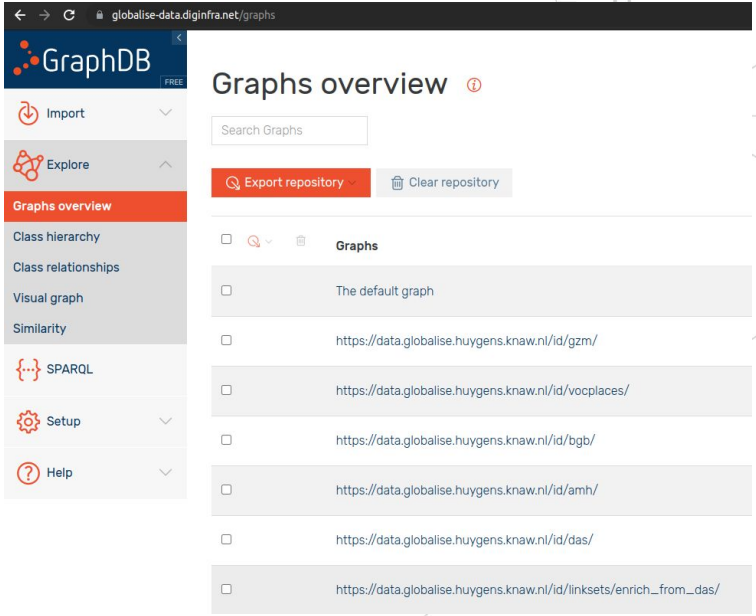
- https://github.com/globalise-huygens/

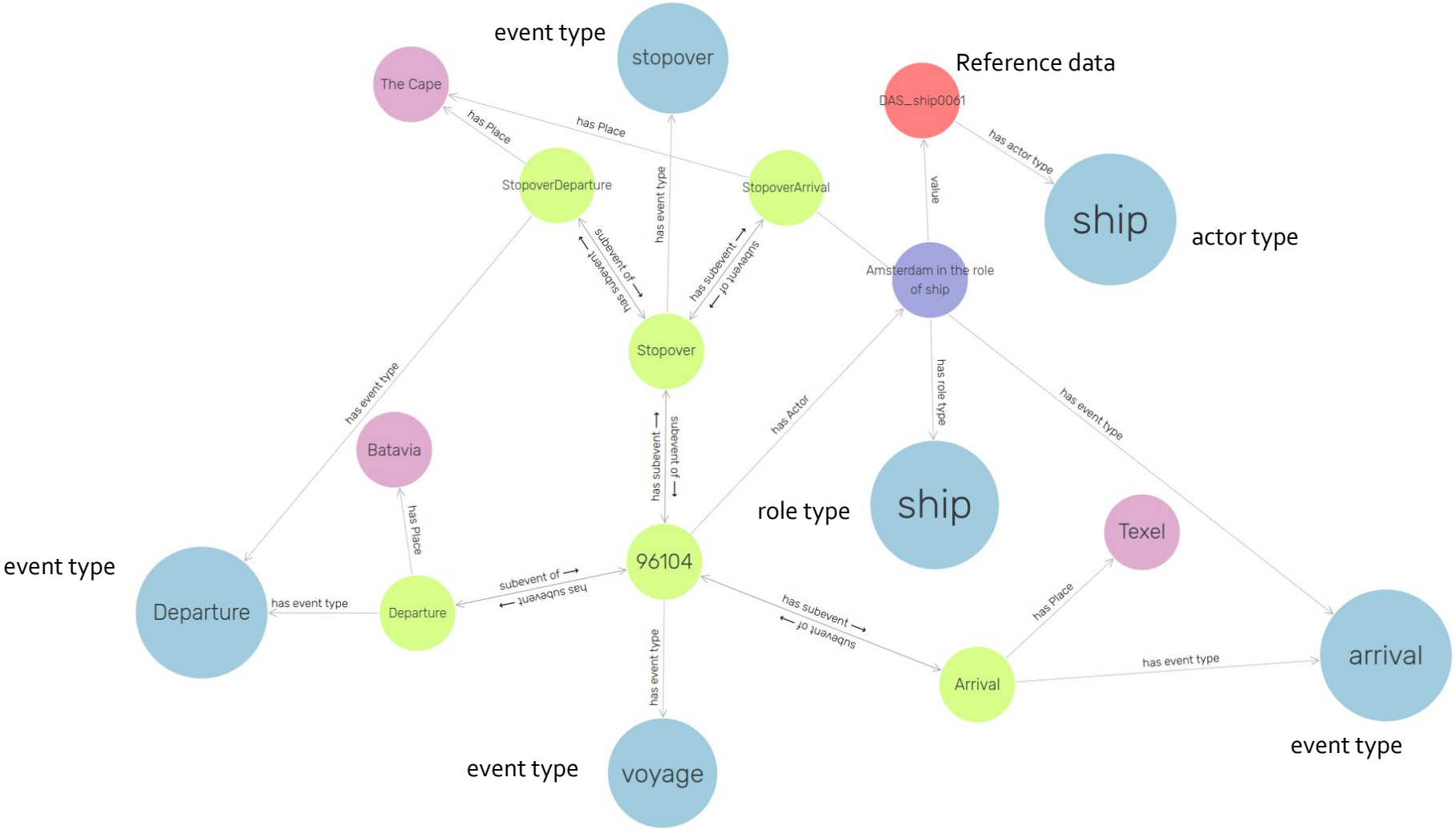
Overview in our triplestore (development)

https://globalise-data.diginfra.net/

- Separated in *named graphs*

Metadata and basic *provenance* (dataset level) according to NDE's Dataset requirements (https://netwerk-digitaal-erfgoed.github.io/requirements-datasets/




event type
stopover
Reference data
The Cape
DAS_ship0061
has Place
has Place
has actor type
StopoverDeparture
has event type
StopoverArrival
value
ship
actor type
subevent of
has subevent
has subevent
subevent of
Amsterdam in the role of ship
Stopover
has event type
has role type
has event type
Batavia
has subevent
subevent of
has Actor
role type
ship
Texel
has Place
event type
Departure
has event type
Departure
subevent of
has subevent
96104
has subevent
subevent of
has Place
Arrival
has event type
arrival
has event type
event type
voyage
event type

```
<bgb/cargo/679315> a sem:Object ;
        sem:actorType glob:vocabulary/actortype/goods ;
        ontology:cargoQuantity 1868.0 ;
        ontology:cargoUnit "lb" ;
        ontology:lichtGuldens 233 ;
        ontology:lichtStuivers 10 ;
        ontology:order 2 ;
        ontology:productType "lood" .
```

Question: how to best model a group?

To store number of:
- Passengers
- Soldiers
- Seafarers etc.

Example queries

1. All ships in service

   Query (in dev triplestore)

   SPARQL-query (.rq)

2. All places in East-Asia

   Query (in dev triplestore)

   SPARQL-query (.rq)

Next steps: Other domains

Current focus in the pilot: Ships

New domains for our thesauri and reference data

Places

- Integration with WHG (https://whgazetteer.org/)

Polities

Persons (and their roles)

Units (e.g. lb, pees, currencies)

Time

- Other than Gregorian
- Extension to *Heideltime*? (https://github.com/HeidelTime/heideltime)

Commodities

Feedback into NLP pipeline

Used for training of entity resolution methods

- Extension of CLARIAH pilot (https://github.com/CLARIAH/wp6-missieven)
- Useful for (pre-)annotation

Input for semantic role labeling / frame annotation

- List of (historical) words (verbs/nouns) that signal an 'event'
- Detecting 'situations' (e.g. transport of goods, change of power, conflicts)
- *Who* did *what* to *whom*?
    - Reference data as contextualisation layer

Future developments / next steps

Publishing our thesaurus and datasets

- Integration in PoolParty / NDE's termennetwerk https://termennetwerk.netwerkdigitaalerfgoed.nl/)

Modelling our data in a CIDOC-CRM vocabulary (generic)

- Build on top of the SeaLiT ontology (https://doi.org/10.5281/zenodo.5964239)

Developing our own ontology (specific)

- Focus on events/situations (cf. Frame Semantics)
- Depends on research questions and infrastructure

in collaboration with
dutch digital heritage network

Rijksdienst voor het Cultureel Erfgoed
Ministerie van Onderwijs, Cultuur en Wetenschap

RCE Erfgoedthesaurus

CIDOC CRM
CONCEPTUAL REFERENCE MODEL