

Semantics and the Humanities

some lessons from my journey 2000-2012

Guus Schreiber
VU University Amsterdam

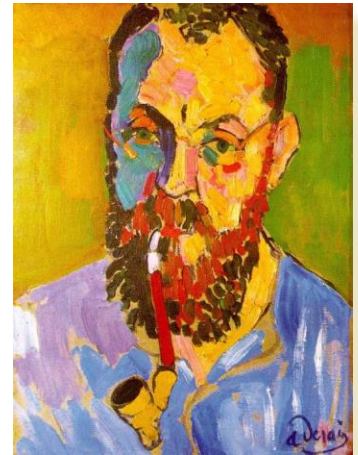
The Web: resources and links



URL



Web link



URL

The Semantic Web: typed resources and links

Painting
“Woman with hat
SFMOMA

Dublin Core

creator

ULAN

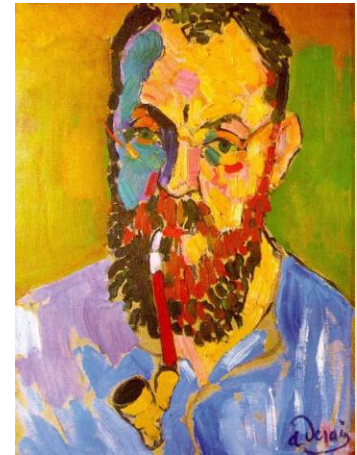
Henri Matisse



URL



Web link



URL

2000

photo annotation study

Ape use case

- A person searches for photos of an “orange ape”
- An image collection of animal photographs contains snapshots of orang-utans.
- The search engine finds the photos, despite the fact that the words “orange” and “ape” do not appear in annotations



Annotation with vocabulary (Iconclass)

MIA tool -- file:photos/ar5611-002.jpg

File ▾ Annotate ▾ Ontology ▾ View ▾ Help ▾ Photo_annotation ▾

Subject matter | Photo features | Medium features

Create: Passive agent Active agent

young scratching_the_head chimpanzee

Agent
chimpanzee

With

Life stage
young

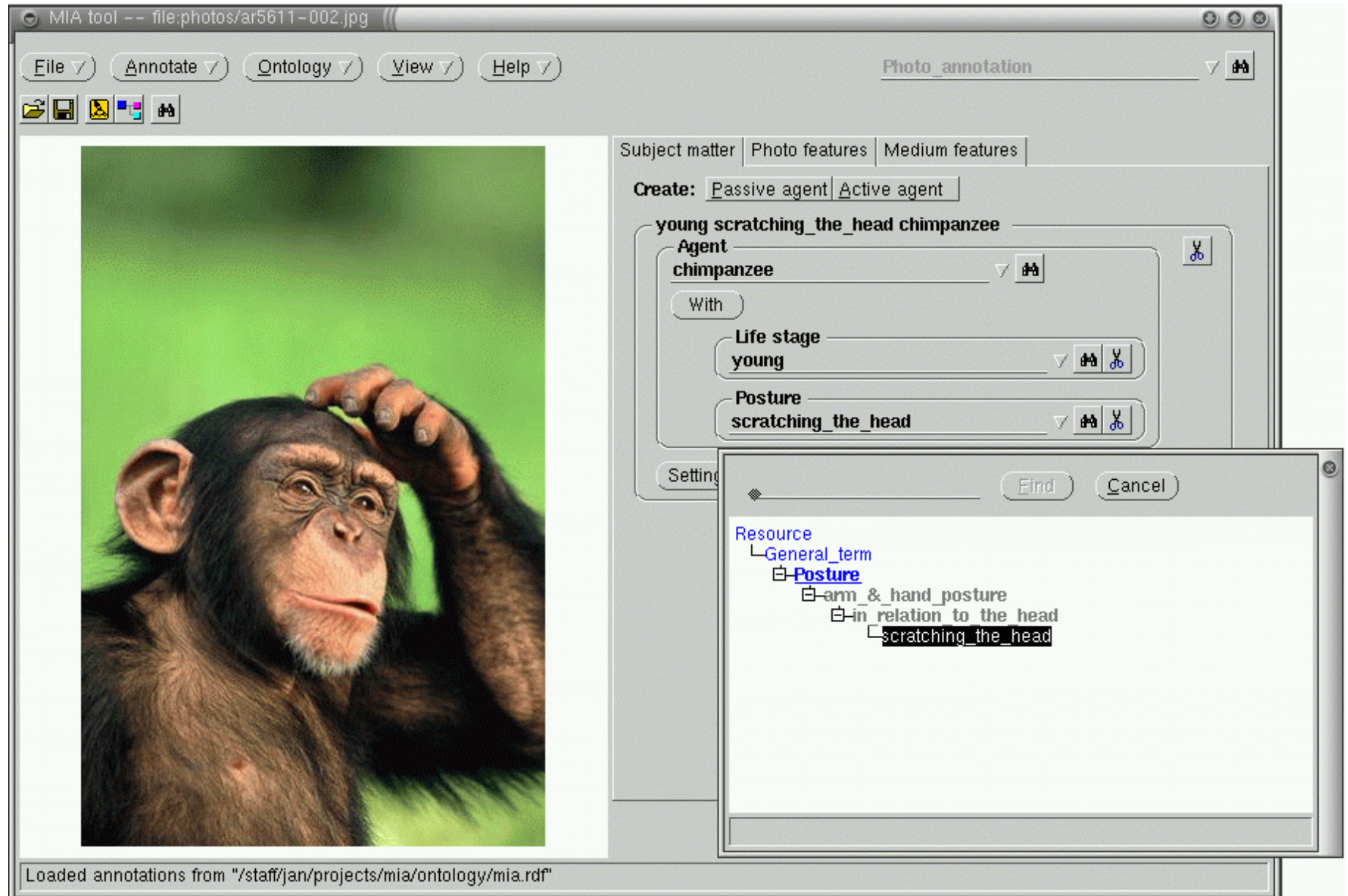
Posture
scratching_the_head

Setting

Find Cancel

Resource
└ General_term
 └ Posture
 └ arm & hand posture
 └ in relation to the head
 └ scratching_the_head

Loaded annotations from "/staff/jan/projects/mia/ontology/mia.rdf"



Qualitative comparison with image search engines

- AltaVista image finder:
 - “great + ape”: 13 hits, 6 of which are images of apes
 - “great ape”: 45,000 hits. Precision 32% (estimate)
- Gettyone.Com
 - Uses some partial hierarchies
 - Good results on “ape”
 - No results “great ape”
 - Poor results on ape features
 - Result with “ape scratching”
 - No result with “ape AND hand AND head”
 - Knowledge of indexing scheme is necessary for successful complex queries

2003

Annotating paintings with help of multiple vocabularies

MIA tool -- file:guus/venus-cupid-bacchus.jpg

File Annotate Search Ontology Help

MIA

Information_Object

Production Physical Content Administration

Create: Active Agent Objects Setting

Active Agent

Agent
Venus

With

Action
lying

Object
bed

With

Recipient

With

Ok Reset

Found "bed" 8 times in ontology



Image content issues

Scope	Characteristic	Conceptual Level			Total
		general	specific	abstract	
Object	event	7.4%	0.1%	0.7%	8.3%
	place	1.9%	0.7%		2.6%
	time	0.2%		0.2%	0.4%
	relations	1.7%		0.1%	1.8%
	uncharacterized	40.1%	14.2%	2.8%	57.1%
Subtotal		51.5%	15.0%	3.8%	70.3%
Scene	event	4.7%		0.3%	5.0%
	place	7.7%	0.9%	1.2%	9.9%
	time	4.3%	0.3%	1.1%	5.8%
	relations	0.3%			0.3%
	uncharacterized	5.8%	0.1%	2.8%	8.8%
Subtotal		22.9%	1.4%	5.4%	29.7%
Total		74.4%	16.4%	9.2%	100.0%

Annotation and search architecture

**Knowledge
corpora**

AAT

ICONCLASS

WordNet

**Annotation
Template**

VRA 3.0

Scene descriptors



*RDF
Schema*

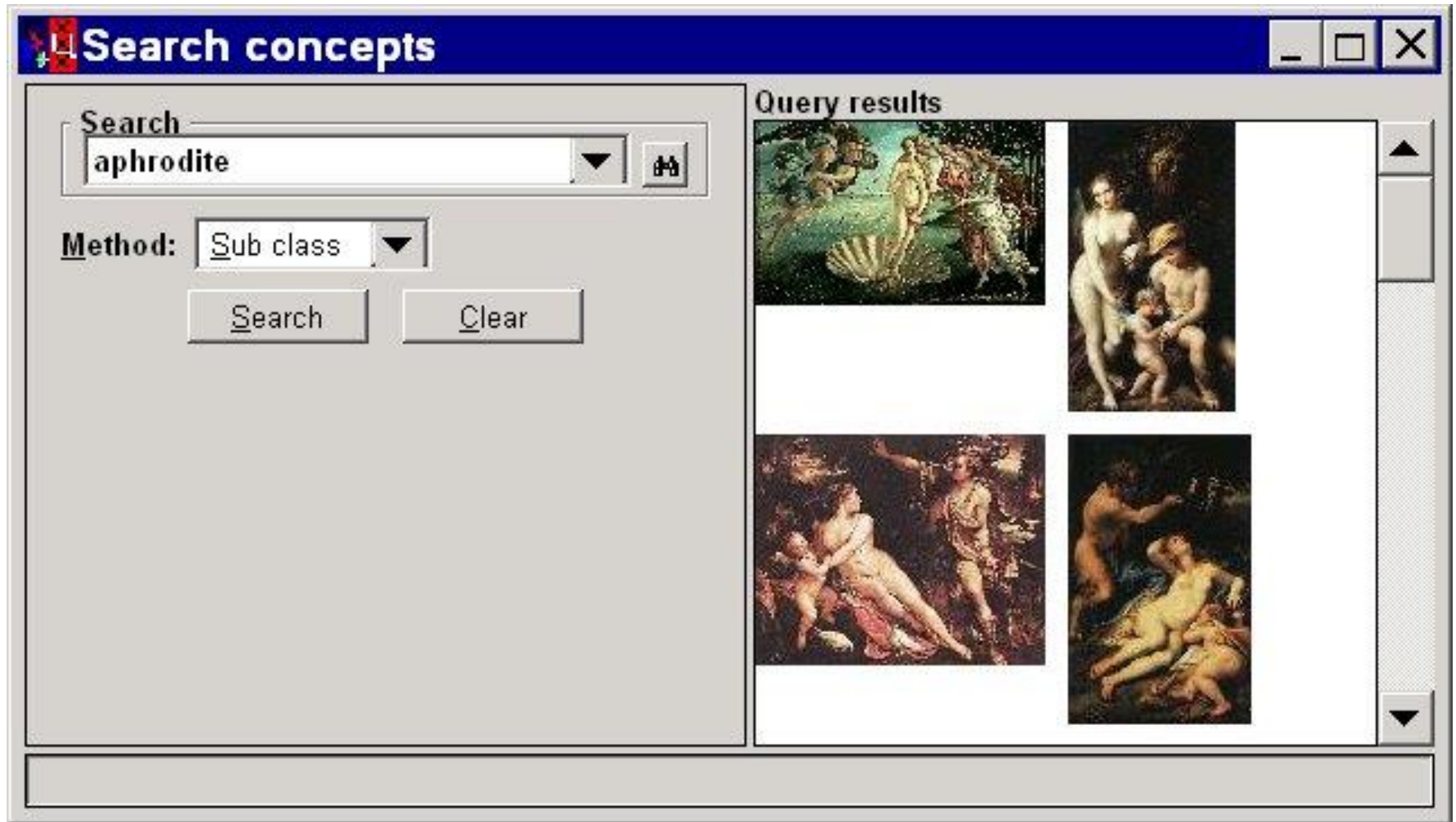


**Annotation
& search tool**



RDF image
annotations

Semantic search attempts



Knowledge representation builds on a long tradition of vocabulary construction



Searching the new Iconclass Browser using these keywords - *lifting* and *hat* - will find us the concept shown here in its hierarchical context.

- 0 Abstract, Non-representational Art
- 1 Religion and Magic
- 2 Nature
- 3 Human Being, Man in General
- 33 relations between individual persons
- 33A non-aggressive relationships
- 33A1 saluting

The myth of a unified vocabulary

- In large virtual collections there are always multiple vocabularies
 - In multiple languages
- Every vocabulary has its own perspective
 - You can't just merge them
- But you can use vocabularies jointly by defining a limited set of links
 - “Vocabulary alignment”
- It is surprising what you can do with just a few links

Ontological commitment

- Ontologists tend to be "trigger-happy"
 - i.e. define as many axioms as possible
- Over-commitment makes ontologies less usable
 - The art of being minimal
 - “In der Beschränkung zeigt sich der Meister”
- Design criterion SKOS: minimal commitment



Navigation

- [Main Page](#)
- [Recent changes](#)
- [Tools](#)
- [Books](#)
- [Validators](#)

Other w3c resources

- [Activity news](#)
- [Publications](#)
- [Logos, buttons](#)
- [Activity home page](#)

W3C RSS feeds

- [Activity newsfeed](#)
- [W3C blogs](#)
- [Use cases, case studies](#)

SKOS/Datasets

< [SKOS](#)

Contents [\[hide\]](#)

[1 Find & Share SKOS Data](#)

- [1.1 STW Thesaurus for Economics](#)
- [1.2 Thesaurus for the Social Sciences](#)
- [1.3 Country Codes](#)
- [1.4 Drug Administration Forms](#)
- [1.5 Vocabularies for GeoSciML Geoscience information interchange](#)
- [1.6 DDC Dewey Decimal Classification](#)
- [1.7 Libris' vocabularies](#)
- [1.8 National Széchenyi Library's vocabularies](#)
- [1.9 Library of Congress' vocabularies](#)
- [1.10 German national Library' subject headings](#)
- [1.11 French National Library's subject headings](#)
- [1.12 VIAF person authorities](#)
- [1.13 Wikipedia categories](#)
- [1.14 New York Times subjects](#)
- [1.15 IVOA astronomy vocabularies](#)
- [1.16 NASA taxonomy](#)
- [1.17 GEMET General Multilingual Environmental Thesaurus](#)
- [1.18 AGROVOC Agricultural Thesaurus](#)

Wordnet

RDF/OWL Representation of WordNet

W3C Working Draft 19 June 2006

This version:

<http://www.w3.org/TR/2006/WD-wordnet-rdf-20060619/>

Latest version:

<http://www.w3.org/TR/wordnet-rdf/>

Previous version:

This is the first published version

Editors:

[Mark van Assem](#), Vrije Universiteit Amsterdam

[Aldo Gangemi](#), ISTC-CNR, Rome

[Guus Schreiber](#), Vrije Universiteit Amsterdam

Copyright © 2006 W3C[®] ([MIT](#), [ERCIM](#), [Keio](#)), All Rights Reserved. W3C [document use](#) rules apply.

2006

[illegible]

ENHANCED FEATURES

Interoperability
XML/RDF/OWL



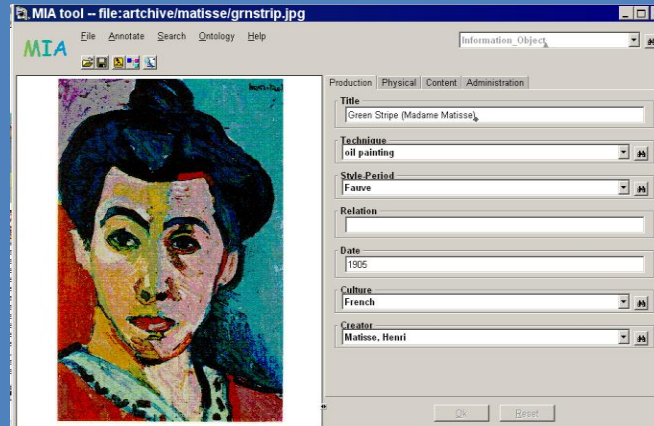
Scalability
> 10,000,000 triples

Ontologies
WordNet, AAT, TGN
ULAN, Dutch labels

Search strategies
sibling search
semantic distance

Dublin Core
specializations
dumb-down

BASELINE



semantic annotation

**DIGITAL HERITAGE
COLLECTIONS**

semantic search



NEW FEATURES

Natural-lang proc.
automatic annotation
text strings → concepts

Distributed
cultuurwijzer.nl collections
OAI-based access

Reasoning support
time/space reasoning

Web interface
support for web collections

Presentation facilities
semantic presentation
device-specific



Semantic annotation

Woman with a Fan

<http://t-d-b.org?http://www.artchive.com/artchive/p/picasso/womanfan.jpg>



Property	Value
<u>Creator</u>	<u>Picasso, Pablo</u>
<u>Location</u>	<u>Paris; The Hermitage Museum, St. Petersburg</u>
<u>Material</u>	<u>canvas; oil paint</u>
<u>Measurements</u>	<u>152 x 101 cm</u>
<u>Title</u>	Woman with a Fan

Relatively easy for
people and places

Type Title Creator Style/Period▼ Material▼ Date Description▼ Culture Measurements▼

Type
12
Work
SVCN Concept
AAT Concept
Place
Person

depicted place
World
Europe
France
Île-de-France
Ville de Paris, Dépar ...
Paris 12
Montparnasse 1
Montmartre 3
facet options

▼ Results grouped by depicted place ▼

Paris (8)



Self-Portrait with Se ...
Chagall, Marc



Paris Through My Window
Chagall, Marc



Homage to Bleriot
Delaunay, Robert



A Glimpse of Notre Da ...
Matisse, Henri



Avenue de l'Opera: Mo ...
Pissarro, Camille

Montmartre (3)



Boulevard Montmartre: ...
Pissarro, Camille



Boulevard Montmartre: ...
Pissarro, Camille



Boulevard Montmartre: ...
Pissarro, Camille

Term disambiguation is key issue in semantic search

- Pre-query
 - Ask user to disambiguate by displaying list of possible meanings
 - Interface is more complex, but more search functionality can be offered
- Post-query
 - Sort search results based on different meanings of the search term
 - Mimics Google-type search

Semantic search: result clustering based on retrieval path

cultural heritage on the web :: [humans](#)



[Pictures](#) and metadata of artworks

Works created by (95)



Three Women
Picasso, Pablo



Glass of Absinthe
Picasso, Pablo



Accordionist
Picasso, Pablo



The Aficionado
Picasso, Pablo

Works by professionally related artist (31)



Fruit Dish, Ace of Clubs
Braque, Georges



Man with a Violin
Braque, Georges



Bottle, Newspaper, Pipe,
Braque, Georges



Still Life BACH
Braque, Georges

Works created by artists with style/period [Surrealist](#) also used by artist (6)

Keyword search with semantic clustering

1. Btree of literals plus Porter stem and metaphone index
2. Find resources with matching labels
 - Default resources are “Work”s
3. Find related resources by one-way graph traversal
 - **OWL use: inverse, symmetric, transitive, same as**
Threshold used for constraining search
4. Cluster results (group instances)

Distributed vs. centralized collection data

- In the beginning we wanted to do it all distributed
 - accessed through protocol such as OAI
- In practice, external metadata access is cumbersome
 - Centralized (cached) metadata
 - Distributed data (images, ..)

Mobile museum tour



Supporting annotation: automatically deriving spatial relations

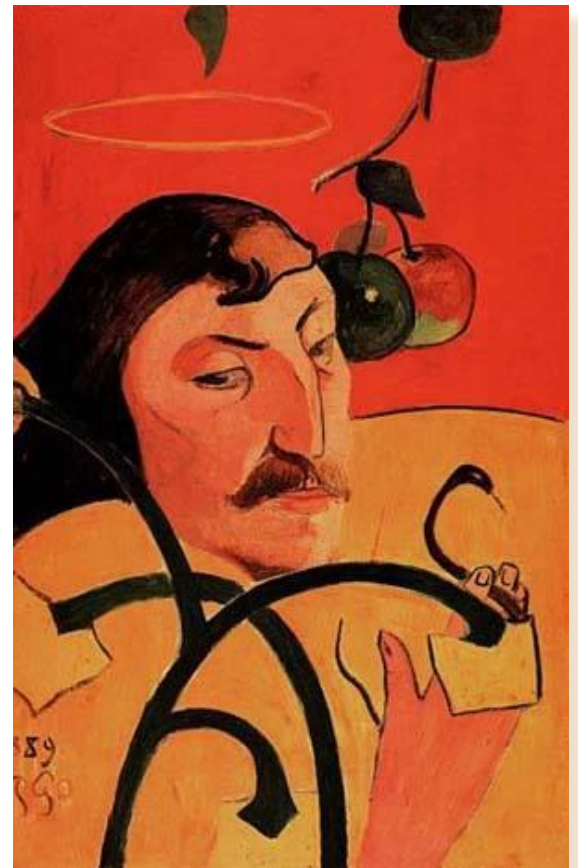
Object1
left
Object2



Other search paradigms: relation search



?

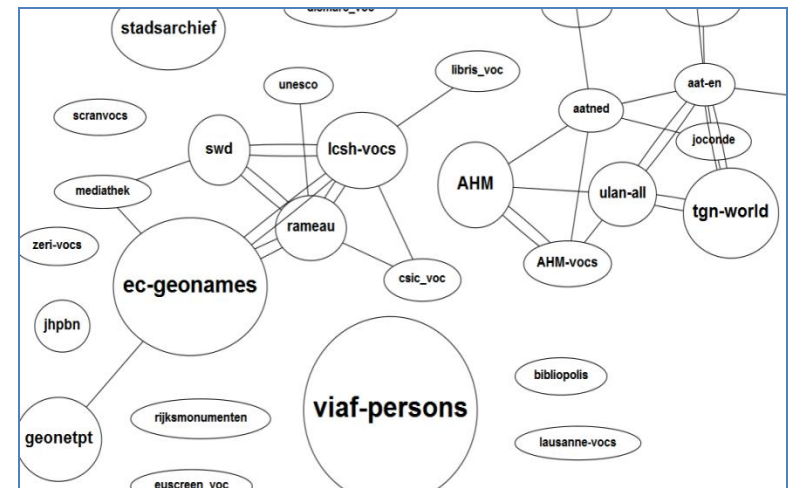
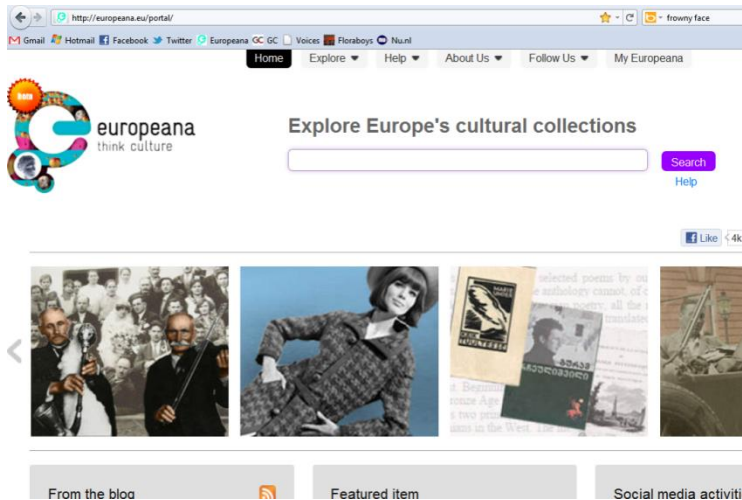


2009

Europeana

- “Europeana enables people to explore the digital resources of Europe's museums, libraries, archives and audio-visual collections.”

www.europeana.eu



From portal...



...to data aggregator.



This is a **research prototype** of **Europeana's** semantic search engine.
Enter a search term, for example: **Egypt**, **Rembrandt**, **window**.

Collections

Thesauri



Rijksmuseum
46,038 artworks



RKD
82,781 artworks



Louvre
11,327 artworks

Europeana Data Model (EDM) requirements

1. Distinction between “provided object” (painting, book, program) and digital representation
2. Distinction between object and metadata record describing an object .
3. Allow for multiple records for same object, containing potentially contradictory statements about an object
4. No information loss
5. Standard metadata format that can be specialized
6. Standard vocabulary format that can be specialized
7. EDM should be based on existing standards

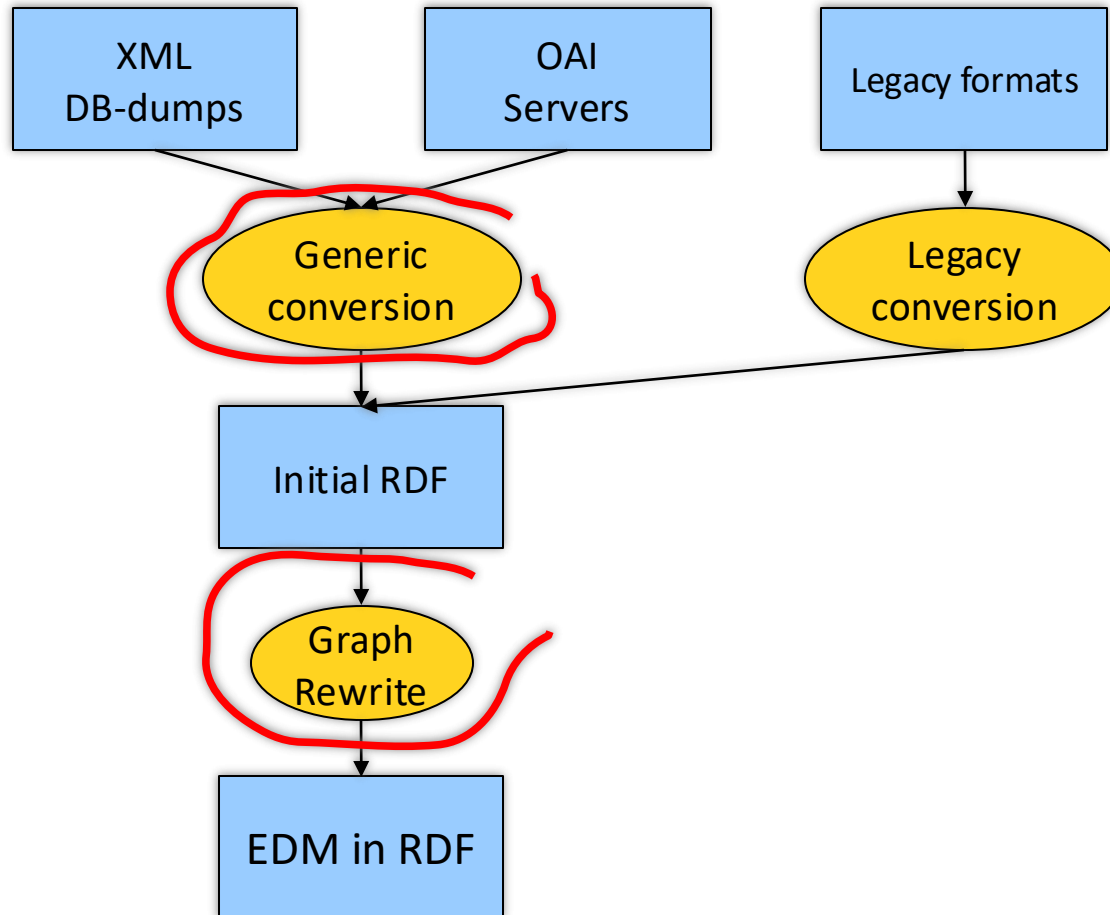
“not yet another standard” !

EDM basics

- OAI ORE for organization of metadata about an object
 - Requirements 1-4
- Dublin Core for metadata representation
 - Requirement 5
- SKOS for vocabulary representation
 - Requirement 6

OAI ORE, Dublin Core and SKOS together fulfil Requirement-7!

Data conversion process



Amsterdam Museum

In Europeana Data Model RDF

[Home](#)[Download \(git web\)](#)[Live search](#)[AM Collectie online](#)[Contact](#)

Europeana ThoughtLab

Victor, 20 Mar 2011

 Tweet 13

 Like 29

The Amsterdam Museum Linked Open Data set is now a part of the [Europeana ThoughtLab](#), a showcase of the technologies developed by the Europeana partners. [The Amsterdam Museum Linked Open Data](#) is a prime example of Linked Open Data of the type that Europeana will be producing.

More info at: www.europeana.eu/portal/thoughtlab_linkedopendata.html

Apps for Amsterdam

Victor, 18 Mar 2011

[AM as LOD](#) is one of the datasets available for the [Apps for Amsterdam](#) challenge. Apps for Amsterdam is an initiative to open up and use data about the city of Amsterdam. The apps-design competition runs up until 4 may 2011.

More info at: www.appsforamsterdam.nl/

Dataset added to CKAN

Victor, 17 Feb 2011

About AM data in EDM/RDF

The Amsterdam Museum (AM) data was retrieved from an XML-dump of the [museum's Adlib collection database](#). The data was converted to RDF compliant with the [Europeana Data Model \[pdf\]](#) (EDM). The AHM EDM data is published as [Linked Open Data](#). The conversion software is available as open source. [Europeana Data Model \[pdf\]](#)

Provenance Graph

- All steps (nodes) live visualized in a strategy/provenance graph

- Nodes include

- input vocabularies

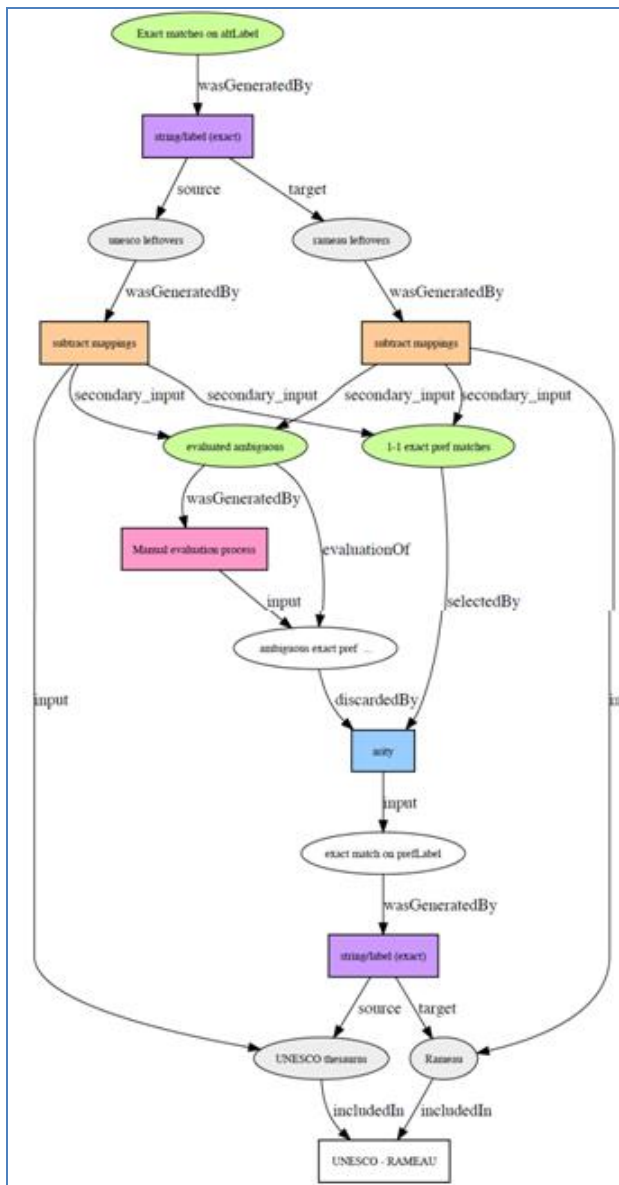
- lexical matchers

- (arity) selectors

- set operations (*subtraction*, union,...)

- mapping sets (intermediary and final)

- evaluation processes



Problems in alignment

- We have not agreed on an adequate alignment vocabulary
 - misuse of owl:sameAs
- We have no adequate methodology for evaluating alignments (Tordai et al., 2011)
 - How to handle lack of gold standard
 - Inter-observer agreement metrics
- In particular, people do not agree on how different classes align
 - and this is not because they don't do it “right”

Limitations of categorical thinking

- The set theory on which ontology languages are built is inadequate for modelling how people think about categories (Lakoff)
 - Category boundaries are not hard: cf. art styles
 - People think of prototypes; some examples are very prototypical, others less
- We also need to make meta-distinctions explicit
 - organizing class: “furniture”
 - base-level class: “chair”
 - domain-specific: “Windsor chair”

Other technical issues

- Information retrieval as graph search
 - more semantics => more paths
 - finding optimal graph patterns
- Information extraction
 - recognizing people, locations, ...
 - identity resolution
- Multi-lingual resources

Organizational issues

- Cultural heritage organizations find it difficult to “give away” their data
 - concerns for quality
- Re-orientation: Web is not derivative of physical presence; they should stand side-by-side
- Universal access: everyone should be able to enjoy the Rijksmuseum Amsterdam

Economic issues

- Primary access free
- Secondary services cost money
 - virtual museum shop can offer much larger collection
 - access to high-resolution images
 - tourist services on mobile devices

2012



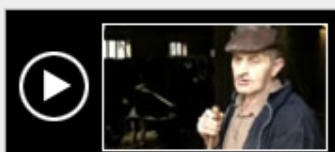
Man bijt hond

Woordentikkertje

Al **158.000** tags ingevoerd en **77.000** matches gemaakt

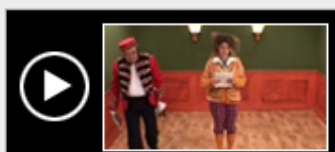
[inloggen »](#)
[registreren »](#)

Selecteer een fragment om het spel mee te starten



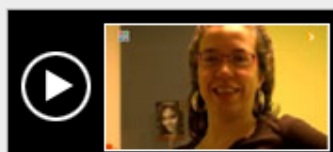
De tiet zal ut ûtwiezen

29 keer gespeeld
hoogste score: 2.915



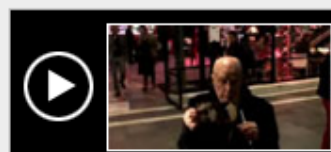
Burgemeester vlekken w

2 keer gespeeld
hoogste score: 35



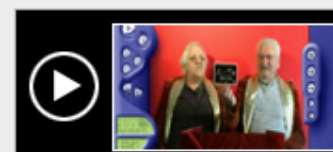
Een week later

4 keer gespeeld
hoogste score: 475



Wat sleept u met u mee?

0 keer gespeeld
hoogste score: 0



Spul wat ze inspuiten

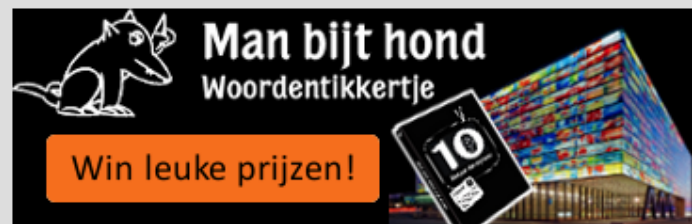
0 keer gespeeld
hoogste score: 0

Hoe werkt dit?

Selecteer een video en probeer zoveel mogelijk woorden in te voeren die beschrijven wat u ziet of hoort. Als anderen binnen tien seconden hetzelfde woord invoeren heeft u een 'match'. Matches leveren punten op. U kunt hierboven een fragment selecteren, of één uit de wachtrij, en direct beginnen met spelen. Om uw punten te kunnen bewaren dient u geregistreerd en aangemeld te zijn. Om mee te kunnen doen moet u de [Microsoft Silverlight plugin](#) geïnstalleerd hebben.

[registreren](#)

[inloggen](#)



Topscoreurs deze week

1



siempie

51.275



Populairste tags deze week

limburg

man

mensen

Doe direct mee (0 spellen)

Klik op een spel om mee te doen

Start in



Man bijt hond

Woordentikkertje



Hallo, Michiel!

uitloggen »

0

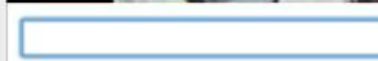
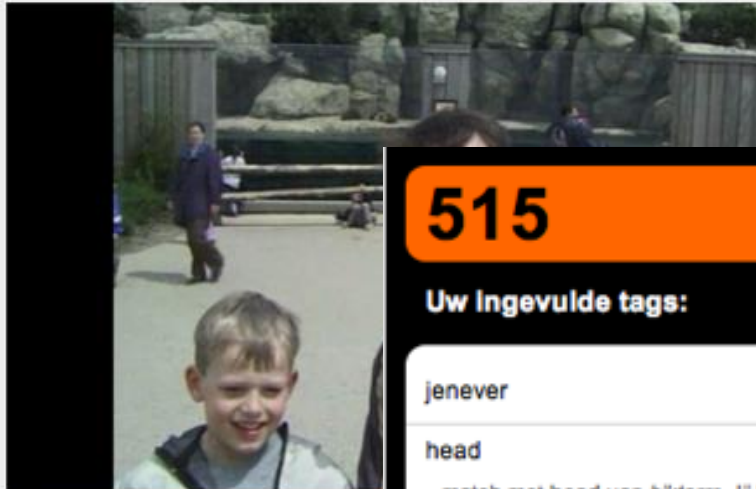
Uw puntentotaal



Winner EuroTV Competition
Best Archives on the Web Award

Hoe houdt u de stemming erin?

nog 00:15



Deelnemers

1  Michiel 2  biktorrr

510

Uw Ingevoelde woorden:

glasses

+5

babbelbox

+5

515

Uw Ingevoelde tags:

jenever

+5

head

+150

match met head van biktorrr. Jij hebt dit woord
geïntroduceerd

man

+5

woman

+5

beard

+150

match met beard van biktorrr. Jij hebt dit woord
geïntroduceerd

meeting

+5

bar

+25

geografische naam

wine

+150

Uw bijdrage in dit spel

punten

8 tags zonder match ?

40



1 geografische naam ?

25



0 persoonsnamen ?

0



3 matches met medespelers ?

150



3 pioniersmatches ?

300

Puntentotaal

515

Kijk hiernaast in het overzicht van 'Uw ingevoerde tags' voor een specifiekere toelichting per match.

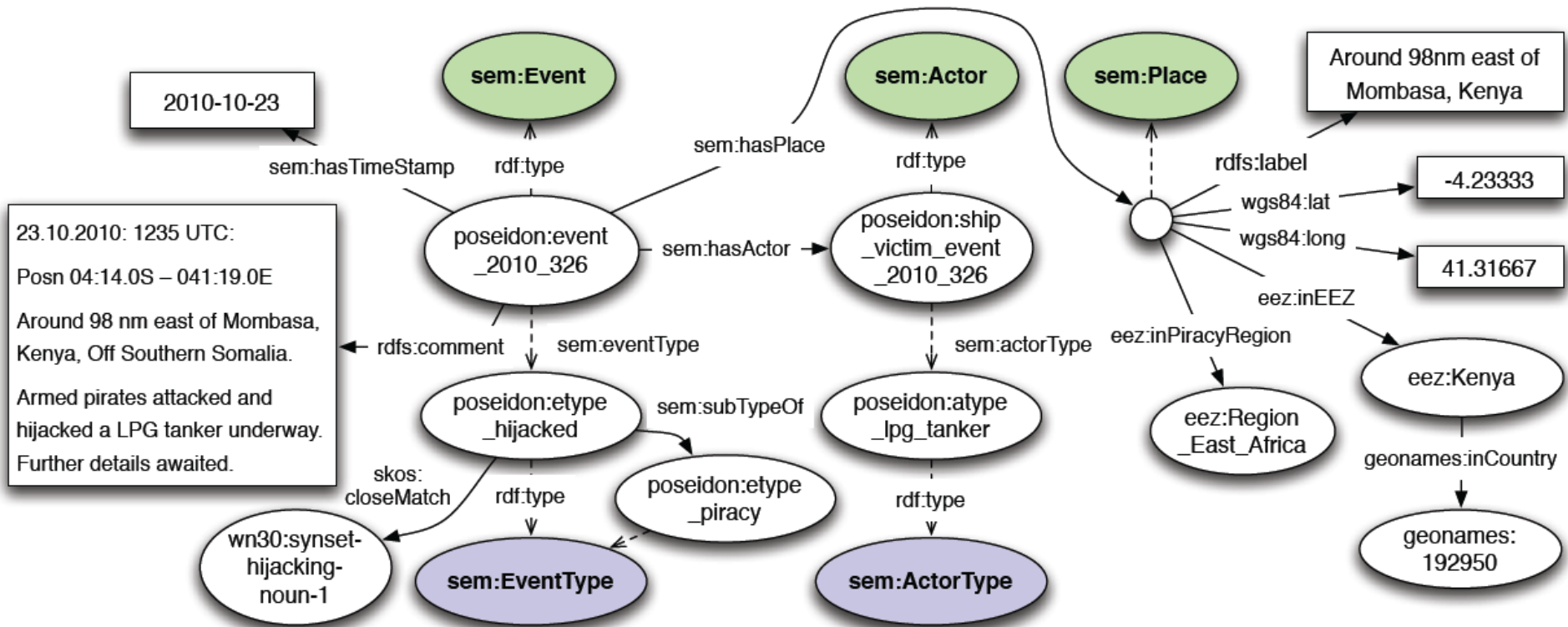
Meer punten verdienen?

Daag vrienden uit om dit spel te spelen en vergroot uw eigen kans op een hogere score ?

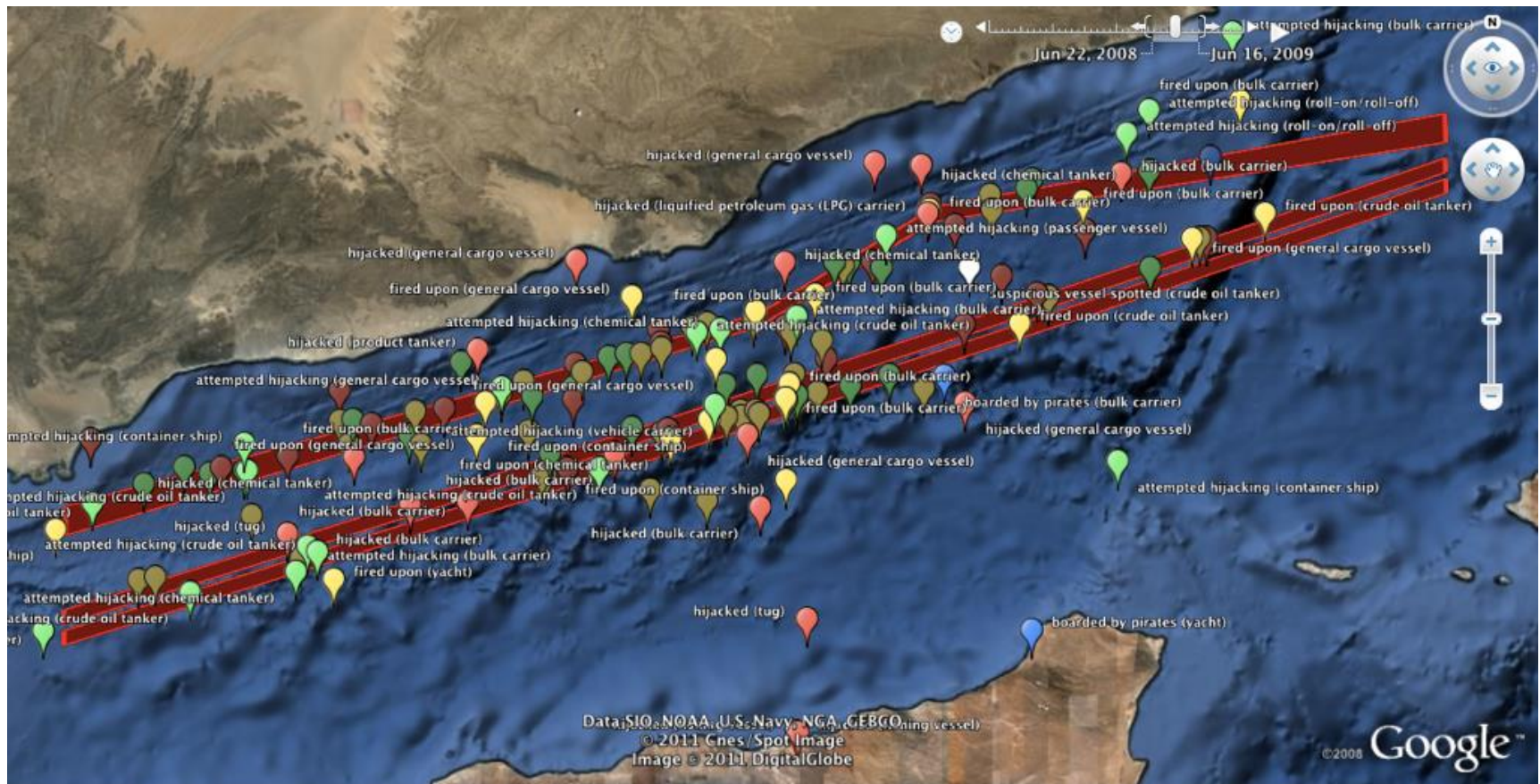


Events

from piracy reports & Web sources



Demonstrator from Poseidon: *Visualising piracy events*





Noordelijk Scheepvaartmuseum

U bevindt zich hier : [Home](#) » [Collectie en Onderzoek](#) » [Onderzoek](#) » [Monsterrollen](#)

- » [Home](#)
- » [Pers](#)
- » [Evenementen](#)
- » [Tentoonstellingen](#)
- » [Groningen in 1470](#)
- » [Kinderen en Jongeren](#)
- » [Onderwijs](#)
- » [Collectie en Onderzoek](#)
 - » [Collectie](#)
 - » [Onderzoek](#)
 - » [Bibliotheek](#)
 - » [Archief](#)
 - » [Scheepsbouwtekeningen](#)
 - » [Fotocollectie](#)
 - » [Monsterrollen](#)

Bekijk monsterrol

Informatie over schip en bemanning

Monsterrol : 1847-1
Datum : 26-01-1847
Scheepsnaam : Hoop op Fortuin
Scheepstype : smak
Grootte : niet vermeld
Bewaarplaats : Groningen, Groninger Archieven (Groningen)

Achternaam	Voornaam	Rang	Gage	Woonplaats	Leeftijd
Cuperus (Cuperes)	Albert	matroos	niet vermeld	Sneek (NL)	26
Stomp	Gerrit R.	kok	niet vermeld	Groningen (NL)	17
Stomp	Jolke R.	stuurman	niet vermeld	Groningen (NL)	23
Stomp	Roelf Jolkes	schipper	niet vermeld	Groningen (NL)	51

[Terug naar het zoekresultaat](#)



Gevonden voor **hoop op fortuin**

Gevonden krantenartikelen

[Deel](#)

1-10 van **22.821** resultaten

★ 0 favoriete kranten

👁 0 bekeken kranten

Sorteer op: [artikel](#) | [krantentitel](#) | [datum](#) | **[relevantie](#)** ⓘ

Weergave: ☐

1 2 3 4 5 6 7 ... 2283 >

Periode

- ☐ 18e eeuw (346)
- ☐ 19e eeuw (9330)
- ☐ 20e eeuw (13145)

Soort bericht

- ☐ Advertentie (859)
- ☐ Artikel (21829)
- ☐ Familiebericht (130)
- ☐ Illustratie met onderschrift (3)



UITGEZEILD.



UITGEZEILD. DUINKERKEN, 14 December; Viser; **Op Hoop van Fortuin** , Rotterdam. BORDEAUX, 15 December; Doewes, Johanna, Amsterdam. ROYAN , 15 December; Smaal, Lamoraal Ulbo , Amsterd. HULL , 16 December; ...



Soort bericht

artikel

Krantentitel

Algemeen Handelsblad

Datum

22-12-1837



This is the site of the Dutch Biography Portal. You can find scientific information here about various leaders and other prominent figures from Dutch history, from the earliest times up to the present day. The website will be constantly replenished with information and images.

News

- Due to maintenance works the Biography Portal is currently unavailable
- Recently added: the Compendium of office holders and civil servants (1428-1861)
- Recently added: the National Biographical Dictionary (NBW)
- Recently added: the Magistracy of Amsterdam (1578-1795) by Elias
- New search option: religion
- The data of the portal are downloadable using the BioPort API
- More than 110.000 biographies!

Search



[Advanced search](#)

Browse

[a](#) [b](#) [c](#) [d](#) [e](#) [f](#) [g](#) [h](#) [i](#) [j](#) [k](#) [l](#) [m](#) [n](#) [o](#) [p](#) [q](#) [r](#) [s](#) [t](#) [u](#) [v](#) [w](#) [x](#) [y](#) [z](#)

Participants in the Biography Portal are: Instituut voor Nederlandse Geschiedenis, Huygens Instituut, Internationaal Instituut voor Sociale Geschiedenis, Biografie Instituut, Centraal Bureau voor Genealogie, Digitale Bibliotheek voor de Nederlandse Letteren, Onderzoekscentrum voor Geschiedenis en Cultuur (UU), Parlementair Documentatie Centrum, Rijksbureau voor Kunsthistorische Documentatie, Data Archiving and Networked Services. This project was made possible by funding from the NWO, Prins Bernhard Cultuurfonds and SenterNovem ('Digitaliseren met beleid' / OCW).

Born on March 20



Adolfs, Jan (1634-1698)



Diepenhorst, Isaäc Nicolaas
Theodoor (1907-1976)



Holst, Gilles (1886-1968)

[more](#)

Mockup Prototype (II)

Navigation Path	Information	Sources	Sourcetext
<input type="checkbox"/> Added: Desiderius Erasmus	Born at 28-10-1466 <div> <div>1</div> <div>2</div> <div>2</div> </div>	<div> <div>Born in Rotterdam</div> <div>Author: Unknown Title: DBNL</div> </div>	Desiderius Erasmus geboren: 28 oktober 1466 te Rotterdam overleden: 12 juli 1536 te Bazel
	Born at 28-10-1467 <div> <div>1</div> <div>2</div> <div>2</div> </div>	<div> <div>Born in Rotterdam</div> <div>Author: B. Glasius Title: BW van Nederlandsche Godsgelerden</div> </div>	
	Born at 28-10-1469 <div> <div>0</div> <div>1</div> <div>4</div> </div>	<div> <div>Born in Rotterdam</div> <div>Author: K. ter Laan Title: Letterkundig Woordenboek voor Noord en Zuid</div> </div>	
	Born in Rotterdam <div> <div>3</div> <div>1</div> <div>0</div> </div>	<div> <div>Born in Rotterdam or Gouda</div> <div>Author: Unknown Title: Wikipedia</div> </div>	ERASMUS werd geboren in den nacht van 27 op 28 Oct., waarschijnlijk in 1466 , uit een ongeregelde verbintenis, waarvan de aard in 't duister ligt. In 1506 worden zijn ouders in een aan hem gerichte pauselijke bul beschreven als een ongehuwde leek (solutus) en een weduwe; een dergelijk document van 1517 doet de mogelijkheid veronderstellen, dat zijn vader priester geweest is.
	Born in Gouda <div> <div>0</div> <div>1</div> <div>3</div> </div>		

Links with other datasets

- Database with data about all Dutch ministers in the period 1575-1815 (Fred van Lieburg)
- Data converted for interoperability with other sources
- Geographical mobility computed with help of original place names and Geonames data (<http://www.geonames.org>)

Mobility of Dutch ministers 1575-1815

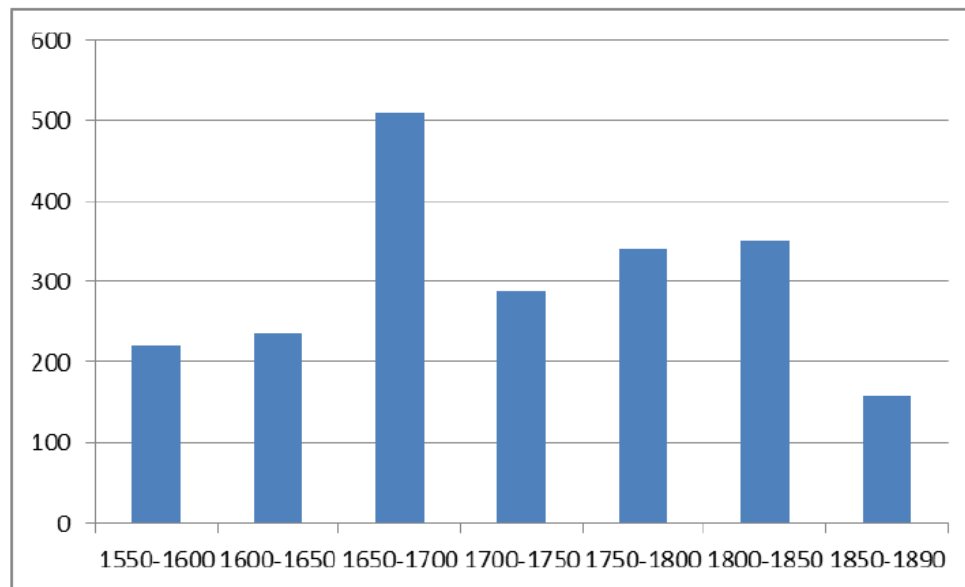


Fig. 3. Mobility in kilometres (distance moved between the known place of birth, work and death) per 50 years (based on dates of death of the ministers)

Acknowledgements

- Many co-workers, from ICES-KIS, MultimediaN E-Culture, CHOICE, CHIP, Agora, MuNCH, BiographyNet, Europeana Connect, PrestoPrime, Dutch Ships & Sailors