1. Project Title & Acronym and Abstract

Title	Dutch Ships and Sailors
Acronym	DSS
Abstract	As a sea-faring nation, a large portion of Dutch history is found on the water. However, much of the digitized historical source material is still scattered across many databases and archives. This *curation and demonstrator* project aims to bring together the rich maritime historical data preserved in the many different databases. We propose a (semantic) *web-based infrastructure* that will house various maritime-historical datasets. We will provide a tool chain and methodology for *converting legacy datasets*. The infrastructure includes common vocabularies to normalize and enrich existing data. Links are established between the datasets and to other relevant datasets on the Web. Although the infrastructure will be set up to facilitate 25+ identified datasets, we initially populate the infrastructure with four selected datasets. These will allow us to investigate two case studies in order to answer the historical research question "*To what extent did patterns of shipping and recruitment in the Dutch maritime sector change over the course of the 18th and 19th centuries?*"
Target Start Date	01-04-2013
Target End Date	28-02-2014
Type	Demonstrator & Curation
Call	Open call

2. Coordinator

Name	Prof. dr. Lex Heerma van Voss
Function	Head Research group Political and Institutional History
Organization	Huygens Institute for the History of the Netherlands
Address	Postbus 90754, 2509 LT Den Haag
E-mail	
Tel	
Fax	
Role(s)	Project-coordinator – User / Data Provider

3. Composition of the Project Team

Name	Dr. Victor de Boer
Function	Post-doctoral researcher
Organization	VU University Amsterdam – Dept. of Computer Science – Web & Media (VU-WM)
Address	De Boelelaan 1081, 1081 HV Amsterdam
E-mail	
Tel	
Role(s)	Technology provider

Name	Matthias van Rossum, Mphil
Function	Post-doctoral Researcher
Organization	VU University Amsterdam – Faculty of Arts – Economic and Social History
Address	De Boelelaan 1105, 1081 HV Amsterdam
E-mail	
Tel	
Role(s)	User / Data provider

Name	Dr. Jur Leinenga
Function	Post-doctoral Researcher
Organization	Huygens Institute for the History of the Netherlands
Address	
E-mail	
Tel	
Role(s)	User / Data provider

Name	Prof. dr. Karel Davids
Function	Professor of Economic and Social History
Organization	VU University Amsterdam – Faculty of Arts – Economic and Social History
Address	De Boelelaan 1105, 1081 HV Amsterdam
E-mail	
Tel	
Role(s)	User

Name	Prof. dr. Jan Lucassen
Function	Professor International and Comparative Social History
Organization	International Institute of Social History and VU University Amsterdam
Address	De Boelelaan 1105, 1081 HV Amsterdam
E-mail	
Tel	
Role(s)	User

Name	Dr. Joost Schokkenbroek
Function	Head Curator and Manager Academic Programs
Organization	Het Scheepvaartmuseum
Address	Postbus 15443, 1001 MK Amsterdam

E-mail	
Tel	
Role(s)	User

4.   CLARIN centre

Huygens ING will host the finalized web-infrastructure as well as the individual converted datasets. Long-term sustainability will be secured through the curated resources at DANS-KNAW.

5.   Requested Budget: Just the requested amount in Euro

The requested budget is: € 98.436,25

6.   Description of the Proposed Project

6.1   Research Question(s)

The maritime industry has been central to regional and global economic, social and cultural exchange (Unger 2011). It is also one of the best historically documented sectors of human activity. Many aspects of it have been recorded by shipping companies, governments, newspapers and other institutions. In the past few decades, much of the data in the preserved historical source material has been digitized. Among the most interesting data are those on shipping movement and crew members. The Dutch Republic in the 17th and 18th centuries had to rely to a large extent on immigration to man its fleet (Van Lottum 2007). Especially in Asian waters, it also relied on Asian crews (Van Rossum 2011).

Often information that deals with the same shipping movements and crew composition is spread over several historical sources and hence over several databases. The data often refers to the same 'places', 'ships', 'persons' and 'events'. By linking the different available databases, the data can complement and amplify each other, and new research possibilities open up. Ideally, we would want to be able to follow a ship from port to port, and crew members pursuing their careers from ship to ship.

With that aim, this project brings together the rich maritime historical data preserved in different databases. It will do so by enriching and linking a selection of available datasets, and developing an infrastructure that will be open for future links to more maritime historical data and databases. The curation aspect of the project is that available datasets are brought together, and in some cases made available on the Internet for the first time. Metadata and links will be added, to identify and link identical places, ships, persons and events in different databases. The demonstrator aspect is that the project will show that added research possibilities are opened by linking the data across databases. Data and metadata sets will be made available to other researchers and database compilers in the discipline. The portal will offer an opportunity to make additional related data sets available and link them to each other.

One way to enrich the data will consist of linking available maritime historical data on 'ships', 'places' and 'persons' to textual historical data adding information on 'ship movements' and 'events'. For the maritime historical data, we will use for the 18th century a collection of internally linked databases concerning the shipping and personnel of the Dutch East India Company (Verenigde Oost-Indische Compagnie) and for the 19th century the Monsterrollen database of the Noordelijk Scheepvaartmuseum constructed by Jur Leinenga. Both sets of data will be linked to the textual historical source of Historische Kranten of the Koninklijke Bibliotheek, which offers news on 'ship movements' and (overseas) 'events' for both periods.

The Monsterrollen databases contains elaborate data on the crew composition of ships from the Northern Netherlands (c. 1800-1930) and provides information on the sailors involved, such as the places of origin, wage and age. Through the information of the Historische Kranten this information can be linked to 'ship movements' (arrivals and departures) and other 'events' mentioned in the historical news papers, especially in the category 'Scheepstijdingen' and 'Scheepsnieuws'.

The data concerning the Dutch East India Company are available in the databases VOC Opvarenden, providing extensive data on crews of VOC ships leaving the Republic, the database Dutch-Asiatic Shipping, providing data on all inter-continental voyages of VOC ships, and the database Generale Zeemonsterrollen, providing data on the crew composition and sometimes location of VOC ships stationed in Asia and not engaged in inter-continental shipping. Most of the 'ships', 'places' and (intercontinental) 'ship movements' have already been standardized and linked between these datasets. The information of the Historische Kranten will be mined for still unknown 'ship movements' in Asia and 'events'. For example, the VOC ship Kroonenburg is mentioned in the Dutch-Asiatic Shipping database, arriving from the Republic in July 1762 in Batavia, and leaving from Batavia to the Republic again in November 1766. In this period, the ship was active in Asian waters, and a yearly account of its crew size is provided in the Generale Zeemonsterrollen database. The shipping movements and possible events occurring in the intra-Asiatic shipping are reported on in the Historische Kranten. For example, the Middelburgsche courant of 17 December 1765 mentions that on 10 May 1765 the ship was sent from Batavia to Persia.

The project makes available data to answer the research question: how did patterns of shipping and recruitment in the Dutch maritime sector change over the course of the 18th and 19th centuries? The users involved in the project will investigate two separate instances: Dutch East-Indies shipping in the 18th century and Northern-Netherlands shipping in the 19th century. This will supply information on added value of this research approach through linked datasets and these data for different periods and types of shipping.

This project will enable research on the specific patterns of shipping and recruitment in these two sectors. How did these patterns change, and how was this reported in the shipping news? The infrastructure and method created for linking the data through 'places', 'ships', 'persons' and 'events' can, after this project, be extended to other maritime historical data.

6.2 Research Data

The Historische Kranten have been made digitally available by the KB (http://kranten.kb.nl). The metadata is in XML-format, and accessible via OAI-PMH-protocol. The Monsterrollen database has been made available by the Noordelijk Scheepvaartmuseum (http://www.noordelijkscheepvaartmuseum.nl). It will be made available in this project in a re-usable and interoperable format. The VOC Opvarenden and Dutch-Asiatic Shipping databases are available via DANS and for a broader public via an online interface

(http://vocopvarenden.nationaalarchief.nl/ and http://www.historici.nl/Onderzoek/Projecten/DAS). The Generale Zeemonsterrollen database has been constructed by one of the participants and will be made available through this project.

6.3 Technology

Data format The resources will be curated using open (Semantic) Web standards, specifically the Resource Description Framework (RDF). Uniform Resource Identifiers (URIs) will be assigned to resources, acting as persistent (and Web-dereferenceable) identifiers. The RDF can be serialized as XML to allow for XML-based storage and retrieval. The use of Semantic Web standards and principles allows for the light-weight integration of data sources. Specifically, these principles do not impose a singular database schema, but rather enable ad-hoc integration through mapping of fields and values (Domingue 2011) This has made earlier successful data integration efforts such as MultimediaN E-culture possible.[1]

Existing data schemas will be used for the curated sources. These include Dublin Core and SKOS (Simple Knowledge Organization System). As part of the project, reusable data schemas will be identified and imported in the infrastructure. RDF and SKOS are CLARIN-recommended standards, for data categories (e.g. Dublin Core), we provide mappings to ISOcat.

(Semantic) Web-based infrastructure. The common infrastructure will be realized using Semantic technologies, specifically an RDF triple store with SPARQL querying functionality . The RDF data, metadata and links will be created using ClioPatria[2]. ClioPatria is the award winning, SWI-Prolog-based platform for Semantic Web Applications. It joins the SWI-Prolog RDF and HTTP infrastructure with a SeRQL/SPARQL query engine, interfacing to the The Yahoo! User Interface Library (YUI) and libraries that support semantic search. ClioPatria has been used successfully in the cultural heritage domain in for example the MultimediaN E-Culture and EuropeanaConnect projects. The Poseidon project uses ClioPatria for exploring piracy attacks and other events related to martime safety.[3] ClioPatria includes support for converting legacy datasets as well as establishing links between data sources. The latter is done in an interactive fashion focusing on quality and transparency of the enrichments using the Amalgame alignment platform (de Boer 2012). All tools are Web-accessible. All software and data formats used within this project will employ open licenses. To ensure compliancy with Huygens infrastructure, the resulting RDF triples will be transferred to a JAVA-based triple store (Sesame, Jena) located at that institute.

6.4 Description

The currently fragmented historical data provide only partial views on the development of Dutch shipping and recruitment patterns from the 17th century onwards. By connecting the various sets of data it will be possible to address new sets of questions concerning changing recruitment patterns and patterns of shipping. This will be analyzed in combination with specific historical events mentioned in the Historische Kranten database. The method and infrastructure created can in the future provide an infrastructure for linking more and broader maritime historical data. As this can serve to map the mobility of ships, people and also goods, it is relevant not only to maritime historical research, but also to economic, social and migration history research.

---

[1] http://e-culture.multimedian.nl/
[2] http://cliopatria.swi-prolog.org
[3] An (incomplete) list of projects can be found here: http://e-culture.multimedian.nl/software/ClioPatria.shtml

6.5 Plan

Type: Demonstrator & Curation project

Analysis and collection of maritime datasets. Within a pilot study, over 25 maritime historical datasets have been identified.[4] As part of the proposed project, we will expand this list and provide a first web portal to these datasets. This portal will provide information about the location, type, content and format of the datasets as well as any IPR and accessibility issues. For the project, a number of the most promising data sets will be cross-linked and linked with an external data set (the newspapers). This will serve as a demonstrator for the possibilities of linking other datasets related to maritime history. Thus, the portal will serve for the dissemination of the results reached in the project.

Task	Participant	Effort	Related Deliverable
Completion of identification of datasets; overall requirements	VU-Hist	1,5 PM	D1
Initial portal setup	VU-WM	0.5 PM	D1

Design and setup of common infrastructure

The infrastructure based on semantic technologies (RDF, SKOS) and realized using the ClioPatria platform. Specific features will be based on the requirements provided by (potential) users. Common vocabularies and schemata will be identified in a collaboration between the technology provider and the user.

Task	Participant	Effort	Related Deliverable
Requirements for infrastructure	Huygens ING	1 PM	D2
Data curation databases	Huygens ING	2 PM	D3
Historical thesaurus places	Huygens ING	4 PM	D6
Design and setup of Semantic infrastructure	VU-WM	1PM	D5
Identification of common schemata	Huygens ING	1 PM	D2
Identification of common vocabularies	Huygens ING	1 PM	D2
Implementation of common schemata	VU-WM	0.5PM	D3
Implementation of common vocabularies	VU-WM	0.5PM	D3

Enrichment of three datasets

In the pilot study, a selection of datasets (*Monsterrollen*, *VOC Opvarenden, Dutch-Asiatic Shipping, Generale Zeemonsterrollen* and *Historische Kranten*), were identified that will be used to validate the common infrastructure. They will be used to answer the research question as stated in section 6.1. In these enrichment tasks, we convert the datasets, use the common schemas and vocabularies and establish internal and external links. Specifically, we establish links regarding 'places', 'ships', 'persons' and 'events' between the databases. Persistent identifiers are established in the form of Uniform Resource Identifiers (URIs).

---

[4] Jan Lucassen & Matthias van Rossum, 'Uitslag enquête Maritiem Historische Databases', (intern rapport, 7 september 2012).

Task	Participant	Effort	Related Deliverable
Conversion and enrichment of *Monsterrollen*	VU-WM	1 PM	D3
Conversion and enrichment of *VOC Opvarenden, Dutch-Asiatic Shipping, Generale Zeemonsterrollen*, and *Historische Kranten*	VU-WM	1PM	D3
PIDs for curated resources	VU-WM	0.25PM	D 3

Documentation and tool consolidation

The tool-chain used for the conversion will be consolidated into a comprehensible package. and documented for future use. Specifically, a step-by-step conversion and enrichment guide will be delivered to ensure that future expansion of the datasets is enabled.

Task	Participant	Effort	Related Deliverable
Consolidation of tool chain	VU-WM	0.5 PM	D11
User documentation	VU-WM	0.5PM	D10, D11

CLARIN centre and curation plan:

All components are built using open source technology that runs on every LinuX system. The complete system will be developed on a dedicated server at VU Amsterdam. For future access we distinguish between a live environment (hosted by Huygens Institute) and an archival environment (DANS). VUA will transfer the infrastructure to Huygens' servers. The curated data and metadata will be stored at DANS. The metadata will be made harvestable via the OAI-PMH protocol.

Task	Participant	Effort	Related Deliverable
Long-term storage and access	Huygens/ DANS	0,5 PM	D 9
Infrastructure migration	Huygens	1PM	D 9

Metadata/ mapping

Metadata for the datasets will be using Semantic Web standards as well as CMDI categories and data-values. RDF and SKOS are CLARIN-recommended standards, for data categories (e.g. Dublin Core), we provide mappings to ISOcat.

Task	Participant	Effort	Related Deliverable
Metadata mapping	VU-WM	1PM	D 12

7. Deliverables and Milestones

Deliverable	Title	Type	Responsibility	Month of completion
D1	Maritime portal	SW	VU-WM, VU-HIST	M2
D2	List of common schemas, vocabularies	DOC	Huygens-ING	M4

D3	All datasets converted	DATA	VU-WM	M5
D4	Metadata of the resources dealt with in the project	DATA	VU-WM	M5
D5	Infrastructure set up and tested	SW	VU-WM	M6
D6	Internal links established	DATA	VU-WM	M8
D7	External links established	DATA	VU-WM	M8
D8	Metadata made available on a recognized CLARIN server	DATA	Huygens-ING	M9
D9	The curated resources on a recognized CLARIN server	DATA	Huygens-ING	M9
D10	Documentation of curated resources	DOC	VU-HIST	M10
D11	Documentation of tool chain	DOC	VU-WM	M10
D12	Requirements and desiderata for the CLARIN infrastructure	DOC	VU-WM	M10

8.   IPR and Ethical Issues: Risks

Monsterrollen is built by Jur Leinenga, He is willing to cooperate in the proposed project and to make his dataset available. The Generale Zeemonsterrollen database is built by Matthias van Rossum. He is willing to cooperate in the proposed project and to make his dataset available. The VOC Opvarenden and Dutch-Asiatic Shipping databases are available through DANS. Historische Kranten is available through http://kranten.kb.nl in an open and standardized way. (http://kranten.kb.nl/about/Ontsluiting)

9.   Expertise of the applicant(s)

Huygens

Prof. dr. Lex Heerma van Voss has research experience as a historian in maritime history and with databases of shipping and crews. He has been involved in and supervised a number of historical data projects, including Clio-Infra. He heads the research group on Political and Institutional History at Huygens ING.

VU WM

Dr. Victor de Boer has extensive research experience in using and developing semantic technologies in the cultural heritage domain and has worked on the MultimediaN E-Culture, EuropeanaConnect and CLARIN-VK projects. He is currently employed as a post-doctoral researcher at the Web & Media group at the VUA.

VU HIST

Matthias van Rossum MPhil has expertise on economic, social, migration and maritime history. He has knowledge of maritime historical databases, has built the *Generale Zeemonsterrollen* databases and organized a recent workshop on maritime historical databases (7-9-2012).

Prof. dr. Jan Lucassen, prof. dr. Karel Davids en dr. Joost Schokkenbroek are renowned experts in maritime history. They will act as advisors for and users of the proposed infrastructure.

10. Project budget details

11. Literature

Boer, Victor de, Jan Wielemaker, Judith van Gent, Michiel Hildebrand, Antoine Isaac, Jacco van Ossenbruggen, Guus Schreiber. Supporting Linked Data Production for Cultural Heritage institutes: The Amsterdam Museum Case Study. In: Proceedings of the 9th Extended Semantic Web Conference (ESWC 2012) Heraklion, Greece. May 27 – 31.

Bruijn, J.R., F.S. Gaastra, I. Schöffer, *Dutch-Asiatic shipping in the 17th and 18th centuries* (Den Haag 1979-1987)

Domingue, J., Pedrinaci, C., Maleshkova, M., Norton, B., Krummenacher, R.: Fostering a relationship between linked data and the internet of services. In Domingue, J., Galis, A.,Gavras, A., Zahariadis, T., Lambert, D., Cleary, F., Daras, P., Krco, S., Müller, H., Li, M.S., Schaffers, H., Lotz, V., Alvarez, F., Stiller, B., Karnouskos, S., Avessta, S., Nilsson, M., eds. The Future Internet. Volume 6656 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2011) 351–364.

Heerma van Voss, L., J. van Lottum, J. Lucassen, 'Sailors, National and International Labour Markets and National Identity, 1600-1850', in: Unger, *Shipping and Economic Growth*, 309-351.

'Forum: New database projects in maritime history, *International Journal of Maritime History*,'24 nr. 1 (2012) 253-359.

Lottum, J. van, *Across the North Sea. The impact of the Dutch Republic on labour migration, c. 1550-1850*, (Amsterdam 2007)

Van Lottum, Jelle, and Jan Lucassen, 'Six cross-sections of the Dutch maritime labour market: a preliminary reconstruction and its implications (1607-1850), in: Richard Gorski (ed.), *Maritime labour: Contributions to the history of work at sea, 1500-2000* (Amsterdam, 2007), 13-42.

Parthesius, R., *Dutch Ships in Tropical Waters* (Amsterdam 2010).

Rossum, M. van, 'De intra-Aziatische vaart. Schepen, "de Aziatische zeeman" en ondergang van de VOC', *Tijdschrift voor Sociale en Economische Geschiedenis* 8, nr. 3 (2011) 32-69.

Unger, R. (ed.), *Shipping and Economic Growth 1350-1850* (Leiden 2011).