

Digitizing, Curating and Visualizing Archival Sources of Maritime History: the case of ship logbooks of the nineteenth and twentieth centuries

Kostas Petrakis

Centre for Cultural Informatics,
FORTH-ICS, Heraklion, Greece

Georgios Samaritakis

Centre for Cultural Informatics,
FORTH-ICS, Heraklion, Greece

Thomas Kalesios

Centre of Maritime History,
FORTH-IMS, Rethymno, Greece

Enric García Domingo

T.I.G./University of Barcelona/
Museu Marítim de Barcelona, Spain

Apostolos Delis

Centre of Maritime History,
FORTH-IMS, Rethymno, Greece
(apostolosdelis@ims.forth.gr)

Yannis Tzitzikas

Centre for Cultural Informatics, FORTH-ICS,
Heraklion, Greece
Department of Computer Science, University
of Crete, Heraklion, Greece

Martin Doerr

Centre for Cultural Informatics, FORTH-ICS,
Heraklion, Greece

Pavlos Fafalios

Centre for Cultural Informatics, FORTH-ICS,
Heraklion, Greece
(fafalios@ics.forth.gr)

Data de recepció: 26 d'octubre de 2020

Data d'acceptació: 7 de desembre de 2021

DOI: <https://doi.org/10.51829/Drassana.28.649>



■ ABSTRACT

A vast area of research in historical science concerns the analysis of historical archival sources. This involves activities such as digitizing the historical sources, usually using spreadsheets or simple relational databases, and then analyzing the transcribed data using a range of methods depending on the kind of data and the type of research question that needs to be answered. In this paper, we describe the process of digitizing, curating and visualizing original archival sources of maritime history, a process done in the context of a European (ERC) project called *SeaLiT*. In particular, we present a set of innovative tools that have been implemented for supporting historians in transcribing the original sources and curating the transcribed data as well as a web application that visualizes the curated data on an interactive map. The overall process is demonstrated for the case of 16 original *ship logbooks* from the nineteenth and twentieth centuries kept by seven archives in Greece and Spain.

Keywords: Archival Data Transcription, Data Curation, Map Visualization, Digital Humanities, Digital Seafaring, Maritime History, Ship Logbooks.

■ INTRODUCTION

Research in historical science is often concerned with the study and analysis of archival sources in order to describe, examine, question and analyze a sequence of past events and also investigate the patterns of causes and effects that are related to them. Current practice is nearly exclusively to use spreadsheets or simple relational databases to organize the data as rows with multiple columns of related parameters. This is appropriate for direct quantitative analysis under varying parameters which may consequently be used for the scholarly interpretation of causes and impacts. However, it also has some typical problems including:

- The need for faithful, fast and collaborative data entry by historians working on the same project but from different countries.

- The difficulty of representing nested tabular structures, which is quite common when transcribing archival sources.
- The need to curate the transcribed data without having to change the information as it appears in the original transcripts (and thus spoil the data).
- The difficulty in revisiting the original sources of transcribed facts for verification, corrections or improvements.
- The difficulty in combining and integrating data from multiple and diverse data sources and thus perform quantitative analysis of facts coming from different information sources.

These problems cannot be tackled effectively using spreadsheets or simple relational databases and there is currently no solution available. To this end, in this article we describe an approach to cope with these problems. Specifically, we introduce a general methodology and a set of tools that have been developed for i) transcribing a set of historical archival sources, ii) performing corrections and enrichments on the transcribed data without altering the information as transcribed from the original sources, and then iii) making use of the curated data in an external application that visualizes the data on an interactive map. The general objective of these tools is to support historians in collaboratively creating, curating, maintaining and exploring a rich database of archival data for both current and future research. To demonstrate the tools, we consider the case of original ship logbooks from the nineteenth and twentieth centuries gathered from seven archives in Greece and Spain.

The context of this work is the *SeaLiT* project¹, a European (ERC) maritime history project where multiple researchers from different organizations in five countries collaborate towards the project objectives. *SeaLiT* explores the transition from sail to steam navigation and its effects on seafaring populations in the Mediterranean and the Black Sea between the 1850s and the 1920s. Historians in this project investigate issues including the maritime labour market, the evolving relations among ship-owners, captain, crew and local societies, and the

development of new business strategies, trade routes and navigation patterns during the transitional period from sail to steam. The archival sources that are studied in SeaLiT range from handwritten ship logbooks, crew lists, payrolls and student registers, to civil registers, business records, account books and consulate reports. These sources have been gathered from various authorities and are written in various languages including Spanish, Italian, French, Russian and Greek.

Below, before detailing the tools we have developed for data transcription, curation and visualization, we discuss the main challenges with respect to data management in historical science and describe the archival sources of our case study, i.e., the ship logbooks from the nineteenth and twentieth centuries. Finally, we show how the transcribed and curated data and their exploration through the map application can facilitate research in maritime history.

■ PROBLEM DESCRIPTION

The general data management challenge faced by historians is the ability to catalogue historical archival sources accurately and then use the transcribed data as a primary source for historical research. In this context, we have identified three main data management activities that need to be performed by historians, each one having its own requirements and challenges: i) *data digitization*, ii) *data curation* and iii) *data exploration*.

In *data digitization*, historians need the means to faithfully and collaboratively transcribe the archival sources into a database, aiming at both short and long-term use beyond the objectives of a particular research problem or project. In particular, historians need easy-to-use tools that enable them to transcribe as much relevant information as possible and as fast and exactly as possible. This also involves support for nested tabular structures for data entry as well as a strategy to handle uncertainty in the original data such as difficulty in recognizing the characters in a piece of text, a common problem when transcribing historical sources. In SeaLiT, for instance, historians have decided to put a question mark when a

specific character in the original archival source is not recognizable, or put the text inside square brackets when they are not sure about a set of characters.

In *data curation*, historians need to curate the transcribed data from the different sources, probably enrich them with additional information, and finally integrate them into a common form from which historical research and quantitative analysis can be carried out correctly and efficiently. This involves i) applying corrections in the transcripts (e.g., correction of typos), ii) normalizing values for enabling comparisons (e.g., normalization of quantities and periods), iii) adding missing information or enriching with additional data (e.g., adding location coordinates for enabling map visualizations), iv) maintaining vocabularies of terms (e.g., 'ship types' or 'professions'), and v) dealing with varying identity assumptions (a problem known as *instance matching*). With regard to the management of vocabularies, the same term (e.g., the type of a ship) might have been written in different ways even in the same language (e.g., 'barque', 'barc', 'bark'), or a broader/narrower term might have been used. Thus, handling vocabularies or thesauri of terms in this context is very important for supporting more accurate analysis services and more effective data exploration services. As regards the varying identity assumptions, when applying quantitative analysis in the transcribed data, such as finding the average number of sailors embarked from European ports in ships of a specific type, it is important to have given the same 'identity' (identifier) to different person references that correspond to the same person (e.g., giving the same identifier to the person references 'Achille M. Micheletti' and 'Achille Marius Michelletti' who are, according to the historians, the same person). In this way, we can avoid errors in the analysis results such as wrong or misleading statistics.

As regards the enrichment of the data known about an entity, historians need to be able to fill in missing information through indirect evidence with some likelihood in order to improve data coverage with known parameters and thus enable more accurate quantitative (statistical) analysis. For example, when the nationality of a ship is

unknown, a good guess is where it was built. Such enrichments require the ability to play with different assumptions, albeit without spoiling the transcribed data.

In general, an important requirement is that all these curation steps must not alter the data as transcribed from the original sources; historians must be able to 'go back' at any time and check how some particular information appears in the original source(s). This is crucial for the reliability of research results as well as verification that may also lead to knowledge revision. However, there is currently no system available that supports these requirements, in particular support for nested structures in data entry (e.g., nested tables), embedded instance matching and vocabulary maintenance processes, and provenance-aware data curation.

Finally, with regard to *data exploration*, the curated data is made available for exploration by historians and other interested parties through various data access interfaces and visualizations such as search systems, browsing interfaces, map applications, charts, etc. Here historians need to be able to find answers to both simple and complex information needs that might also require combining complementary information from different sources. The challenge is how to make these interfaces intuitive in use, user-friendly and trustworthy.

To cope with these challenges, in this article we mainly explore the possibilities of a new system called FAST CAT. FAST CAT is a web-based system that supports collaborative data digitization through the creation of 'records' belonging to specific 'templates'. A template represents the structure of a single data source, ship logbooks for instance, and organizes the data into tables, supporting advanced functionalities such as nesting tables and selection of terms from vocabularies. Once the historians have finished the transcription of the archival sources into FAST CAT records, they can start the data curation process through FAST CAT TEAM, a special environment in FAST CAT that allows the collaborative curation of the main entities and terms that appear in the records. After the curation process has been completed for one or more of the transcribed records, the curated data can be then exploited by external sys-

tems. For the case of logbooks, we have developed (and present in this article) a web application called *Ship Voyages*, which visualizes on a map the transcribed and curated ship logbooks of the SeaLiT project.

■ CASE STUDY: SHIP LOGBOOKS FROM THE NINETEENTH AND TWENTIETH CENTURIES

Ship logbooks are an invaluable type of source, yet largely unexplored by historians who have only used them in a secondary and supplementary way². Other scientists, however, have first appreciated and made extensive and systematic use of them. Climatologists have worked most with logbooks and they pointed to their usefulness as early as 1970³. Along with their aim of creating a world ocean meteorological database of the period before 1850 through the study of logbooks, climatologists have also contributed to historical studies. They have extracted information and analyzed important aspects and questions in maritime topics such as the history of navigation, trade routes, naval warfare, medical history, environmental history and the social history of work and life on board. Indeed, they were the first to highlight the importance of logbooks for historical research and suggest the need for further analytic exploration of this type of sources⁴. Along with them, albeit to a lesser extent, bio-scientists and ecologists explored whaling logbooks of sailing ships which offered them "valuable information on whaling grounds, whaling practices and the responses of whale stocks to over four centuries of persistent human predation⁵". Historians in some cases used logbooks as the principal source but did not attempt a critical analysis of the multifarious data integrated in their narrative⁶. Based on different historical desiderata, others used the logbooks as the main source to study the material culture and assets of sailors⁷. Only lately have historians focused on detailed analysis of the information from logbooks, reaching conclusions on issues such as ship technology and performance and the role of environmental factors in major historical developments (in this case also making good use of the work already done by climatologists⁸). A very important step towards the fur-

ther exploration of this source has also been made by the digitization of British and American ships' logbooks and their free online access⁹.

In Greece, interest focused mainly on the logbooks in the period of the Greek War of Independence and to a lesser extent merchant sailing ships in the nineteenth century. In both cases we are dealing largely with the transcription of the manuscript accompanied by a historical commentary and explanations on terminology and other difficult parts of the text¹⁰. In the academic field, the only paper to our knowledge consists again of the transcription of the manuscript at postgraduate level¹¹.

In the SeaLiT project, the exploration and analysis of the information from ships' logbooks is among the prime research objectives aiming to contribute to the principal questions on the effects of the transition from sail to steam navigation in the Mediterranean. The sample focuses only on merchant ships, excluding any sort of naval vessel since the project does not address naval/military aspects of the transition to steam. Up to now, the data of fourteen ships have been entered, nine Greek and five Spanish (Table 1). However, data from the logbooks of Italian sailing and steamships are expected to be added by the end of the project. The list of Spanish ships includes two *corbetas* or barques (sailing ships), one passenger and cargo steamship, and two passenger steamships. The Greek ships consist of four sailing ships, three brigs and one barquentine, and five cargo steamers. The examined period of the logbooks stretches from the mid-1860s to the 1920s, which corresponds to the period of study of the SeaLiT project. Thus there is the opportunity to explore the evolution of and changes in many aspects on board different types of Mediterranean vessels (sailing ships, cargo and passenger steamships) across the entire period of the transition.

In Greece, starting in 1836 ship logbooks were the necessary legal documents a captain had to keep on board along with the property certificate of the ship, the crew list, the charter party, the bill of lading, the guarantee of payment of custom duties, the permission for departure, the declaration of cargo, the sanitary certificate and the maritime loan booklet. In case of a court

dispute, the ship logbook was the *sine qua non* accepted written evidence upon which testimonies had to coincide and legal decisions had to be made. This is why the logbook had to be compiled at least twice a day as the responsibility of the captain or another deck officer, and also had to be validated after the end of the voyage by the port authority or by a Greek consulate if the ship arrived in a foreign port¹².

The examined ship logbooks, of both sailing and steamships during the whole period, have the exact same standardized format in printed pages with instructions for the correct way to fill it in at the beginning of the book. On the first and the last pages of the book, the identification data of the ship are mentioned: name, type, tonnage, owner, captain, port registry and the unique telegraphic code, validated by the authorities with the note of the exact number of the pages of the book. The categories of information included the date, the time, before and after noon, the wind direction, the course, the distance covered in nautical miles and any events, which was provided with the largest space to fill in. In this space, apart from the daily incidents at sea and ashore, information on speed, coordinates, crew and the goods carried was recorded (Image 1 and Image 2). The only exception to this type of format is the logbook of the brig *Coundouriotis* in the years 1866-74 which was compiled in a plain book divided in handwritten form into sections that include the date and day of the week and a larger part for everything else (Image 3). However, it is very interesting that the logbook of the *Asimoula* is compiled in a Spanish logbook format bought in Tarragona before departure from its port (Image 4), and the *Demetrio S. Schilizzi* in a British *Nautical Diary and Day's Work Book* bought in Cardiff before departure on the Cardiff-Rio de Janeiro voyage (Image 5).

The examined logbooks were all compiled by hand by more than one captain and officer. The understanding and the extraction of information from this type of source presents various difficulties along with the palaeography of the text. First is the language of the period and especially the nautical terminology which modern researchers are often unfamiliar with¹³. Then another dif-

TABLE 1. LIST OF SHIPS OF THE EXAMINED LOGBOOKS

Ship Name	Ship Type	Tonnage	Year Built	Place Built	Port Registry	Ship-owner	Logbook period of voyages
Camila Primera de Tossa	Barque (<i>Corbeta</i>)	210	1859	Tossa de Mar, Catalonia	Barcelona	Viuda Moré e Hijos and Josep Mestres	1882-1884
<i>Habana</i>	Barque (<i>Corbeta</i>)	684.92	1868	Blanes, Catalonia.	Barcelona	Balcells y Subirán Cía.	9/4/1892-29/6/1892
<i>Tordera</i> (ex <i>Lualaba</i>)	Passenger and cargo steamship	1850 grt / 1170 nrt	1878	Port Glasgow	Barcelona	Línea de Vapores Tintoré	13/1/1916-12/2/1917
<i>Leon XIII</i> (ex <i>Isla de Cuba</i> , ex <i>Taroba</i>)	Passenger liner steamship	4938 grt / 3235 nrt	1888	Pointhouse, Glasgow	Barcelona	Compañía Transatlántica Española	12/11/1896-29/8/1897
<i>Infanta Isabel</i>	Passenger liner steamship	8170 grt / 4118 nrt	1912	Port Glasgow	Cadiz	Naviera Pinillos	20/2/1922-13/7/1922
<i>Coundouriotis</i>	Brig	352 and 86/94 tons	1853	Capo d'Istria	Hydra	G. Kalogiannis	1866-1874
<i>Stratigos Favieros</i>	Brig	362	1874	Skopelos	Skopelos	G.D. Zacharias, from 1880 Panayotis Siskos	1876-1889
<i>Eleni Koupa</i>	Brig	276	1878	Syros	Syros	G. Kalogiannis	1878-1888
<i>Asimoula</i>	Barquentine	307	1880	Galaxidi	Galaxidi	D.E. Katsoulis	1882-1885
<i>Thiresia</i> (ex <i>Charles Tennant</i>)	Cargo steam ship	874 grt / 580 nrt	1869	Low Walker, Newcastle upon Tyne	Piraeus	A. Mango & D. Foscolo	1888-1890
<i>Demetrio S. Schilizzi</i>	Cargo steam ship	2034 grt / 1277 nrt	1893	West Hartlepool	Piraeus	A. Mango & D. Foscolo	1/3/1896 - 20/6/1896
<i>Leonidas</i>	Cargo steam ship	2751 grt / 1750 nrt	1896	Sunderland	Andros	A. Embiricos	1904-1906
<i>Andriana</i>	Cargo steam ship	2958 grt / 1769 nrt	1906	Sunderland	Andros	A. Embiricos	1908-1909
<i>Constantinos</i> (ex <i>Strathearn</i>)	Cargo steam ship	4419 grt / 2845 nrt	1906	Greenock, Glasgow	Piraeus	Panaghos C. Lemos	1924-1927

Source: We would like to thank Dr. Annita Prassa, Director of the General States Archives of Magnesia, Volos, and Eleni Spilioti for having sent a copy of the logbook of the *Stratigos Favieros*. We would also like to thank Minas Antypas, PhD candidate, Univ. of Crete, for having sent an electronic copy of the logbooks of the *Coundouriotis* and the *Eleni Koupa* from Hydra. We additionally acknowledge the help of Daniel Muntané and Claudia Mateo for their work with the logbooks kept in the Archival collections of the Maritime Museum of Barcelona; Private Archive of Evangelos Rafalias, Hydra; Maritime Museum of Galaxidi; Private Archive of the Family of Nikos Zachariadis, Skopelos; Hellenic Literary and Historical Archive (ELIA), Athens; Konstantinidis, "Διο Ναυτικά Ημερολόγια"; Association of Friends of Oinousses Island.

*Image 1. Brig Eleni Koupa
(Private Archive of Evangelos
Rafalias, Hydra)*

Ημερομηνία		Ω ρ α :		Ανεμοί	Πορεία	Μέγεθος	Σ υ μ β α ν τ α
Η. Μ.	Μ. Μ.	Η. Μ.	Μ. Μ.				
8	9			Β. Δ.			Ανταρcticus - αὐτὸς ὅσον τὸ Κρόνος οὐρανῷ, γαλαξίας ὡς ὅτι αὐτὸς
"	9	8					Ἥλιος γνωστὸς ὡς ὁ ἥλιος, γαλαξίας αὐτὸς
"	10			Β. Δ.			Ἰσχυρὸς ὡς ὁ ἥλιος, αὐτὸς γαλαξίας
"	11	4		Β. Δ.			Μεγίστος δ' αὐτὸς ὡς ὁ ἥλιος, αὐτὸς γαλαξίας
"	12						ὡς ὁ ἥλιος, αὐτὸς ὡς ὁ ἥλιος, αὐτὸς γαλαξίας
"	13			Γ. Δ.			Ἥλιος, γαλαξίας αὐτὸς ὡς ὁ ἥλιος, αὐτὸς γαλαξίας
"	14			"			ὡς ὁ ἥλιος, αὐτὸς ὡς ὁ ἥλιος, αὐτὸς γαλαξίας
"	15			"			ὡς ὁ ἥλιος, αὐτὸς ὡς ὁ ἥλιος, αὐτὸς γαλαξίας
"	16	8		Β. Δ.			ὡς ὁ ἥλιος, αὐτὸς ὡς ὁ ἥλιος, αὐτὸς γαλαξίας
"	17	6		Β. Δ.			ὡς ὁ ἥλιος, αὐτὸς ὡς ὁ ἥλιος, αὐτὸς γαλαξίας
"	18	8		Β. Δ.			ὡς ὁ ἥλιος, αὐτὸς ὡς ὁ ἥλιος, αὐτὸς γαλαξίας
"	19	8		Β. Δ.			ὡς ὁ ἥλιος, αὐτὸς ὡς ὁ ἥλιος, αὐτὸς γαλαξίας
"	20			Β. Δ.			ὡς ὁ ἥλιος, αὐτὸς ὡς ὁ ἥλιος, αὐτὸς γαλαξίας

drassana
número 28 | 2020 | 60 - 87

67

Image 3. Brig Coundouriotis (Private Archive of Evangelos Rafalias, Hydra)

drassana
número 28 | 2020 | 60 - 8769

Image 5. Cargo steamer Demetrio S. Schilizzi
(Archive, Hellenic Literary and Historical
Archive [ELIA], Athens)

NAUTICAL DIARY & DAY'S WORK BOOK,		from	to
10th Week.]		THURSDAY, 5th MARCH, 1896.	[65-301.
NAUT. EPHEMERIS,	Sun's Declination, { Mean Noon, 8 5° 41' 28" } Var. 1 Hour, 58 09"	Equation of Time, decreasing, { Sub. from Mean Time, 11m 30' 27s. } Var. 1 Hour, 0' 587s.	Sun's Semi-Diam., 16' 8" Sidereal Time, 22h 54m 55' 11s.
<i>ὁ ἀνεμὸς μεταβόρρως, ἀνέμω, τὴν νύκτα γεννημένον, καὶ μὴ τὴν ἑλίου τοῦ Ὁρίου ὁ Ὁρίου ἀπὸ τοῦ Ὁρίου, τὴν τὴν νύκτα ὁ Ὁρίου ἀπὸ τοῦ Ὁρίου. Ὁρίου ἀπὸ τοῦ Ὁρίου. Ἐλασσοειδὲς Ὁρίου τοῦ Ὁρίου καὶ, μὴ ἀπὸ τοῦ Ὁρίου Ἐλασσοειδὲς τὴν Ὁρίου τοῦ Ὁρίου ἀπὸ τοῦ Ὁρίου Ἐλασσοειδὲς, τὴν Ὁρίου τοῦ Ὁρίου ἀπὸ τοῦ Ὁρίου Ἐλασσοειδὲς Ὁρίου Ὁρίου 58, τὴν Ὁρίου τοῦ Ὁρίου Ἐλασσοειδὲς ἀπὸ τοῦ Ὁρίου 9. — Ἐλασσοειδὲς τὴν Ὁρίου τοῦ Ὁρίου ἀπὸ τοῦ Ὁρίου ἀπὸ τοῦ Ὁρίου τὴν Ὁρίου τοῦ Ὁρίου ἀπὸ τοῦ Ὁρίου ἀπὸ τοῦ Ὁρίου.</i>			
		Lat. by Obs., 13° 52'	
		Long. by Obs., 22° 28'	
		Course by Obs., 215 mi.	
		" by D. R., N 28° E	
		Distance by Obs., 215 mi.	
		" by D. R.	
		" by Log, 205	
		Current, (205) (105)	
		Var. used, 20 S.	
		Dev. used, 19 S.	
		Wind, Direct, N.E.	
		" Force, (adverse)	
		Sea, μικροὶ κύμα	
		Weather, ὁρατὴ	
		Barom., 29° 80'	
		Thermom., Air, 59	
		" Sea,	
		Hygrometer,	
10th Week.]		FRIDAY, 6th MARCH, 1896.	[66-300.

difficulty concerns the variety of measurements and standards of the period in the various geographical areas the ships sailed. The calculation of the weights and measures of the cargo and the different monetary systems are among the most frequent. Third, the calendar used in the Greek ships is the Julian since Greece adopted the new (Gregorian) calendar as late as 1923 after many other countries had already embraced it. In terms of navigation, a principal problem is the position of the coordi-

nates due to the multiple types of meridians used by the captains of different nations. Before 1884, when the International Meridian Conference held in Washington DC established the Royal Observatory in Greenwich, UK, as the 0 degrees point of the world's measurement in time and space, many different types of meridians were in use¹⁴. British ships before 1884 used as a meridian the position of St. Paul's Cathedral in London, 30 s west of the Greenwich meridian¹⁵. However, it seems that even

after the adoption of this international convention, there was not an immediate response by other countries and thus the old meridians were still in use for practical and political reasons¹⁶.

In fact, Spanish ships used the Cadiz Meridian long after the adoption of Greenwich at the Washington Conference in 1884. Even before taking Cadiz as a meridian, it was popular to use a meridian running through Hierro Island in the Canary Islands or, more rarely, Tenerife¹⁷. To make things even more confusing, there were two different Cadiz meridians. The first was in the Cerro de Torre Alta, an old castle site since 1753. However, in 1798 a new observatory was built on Isla del León in San Fernando, close to Cadiz outside the old town. This meant that after 1799, the Cadiz Meridian was tied to the Real Instituto y Observatorio de la Armada. One author claims the Madrid Meridian was used as well, but there is no proof of this in navigation material¹⁸. Of course, Spanish charts had Cadiz as a reference point, though it was also common to find uses of foreign charts with the Greenwich or Paris meridians¹⁹.

Most of the Spanish logbooks used the Cadiz Meridian up to 1907. This is something that it is never specified in the log, but can be discovered by positioning a ship using another source, say an Admiralty chart. Thus, in order to standardise Spanish sailing with that of other countries, the difference between Cadiz and Greenwich has to be calculated by subtracting 6° 17' 15" when sailing eastward and by adding the same coordinates if sailing westwards from Greenwich. Indeed, the logbooks of Spanish ships examined in the period after 1884 followed the Cadiz Meridian.

In the case of the Greek ships, it is not mentioned which meridian was adopted in the calculation of the coordinates. However, data about the position and course of the steamships that operated after the 1884 convention or sailing ships that operated before that year indicate that most probably they used British charts. The only exception to this is the barque *Asimoula* whose captain when sailing to the Black Sea and Azov used Russian maps that had Saint Petersburg as a meridian. Another difficulty is the identification of places mentioned

in the logbooks with today's toponyms that have altered after many years due to political and geographical changes and to policies of nationalization of place names. In addition to this, often the captain and other officers recorded foreign geographic locations as they had heard them or thought them correct, making it a very difficult task for the researcher to identify them. Then there is also the habit of some Greek captains to use often ancient Greek names for certain locations across the Mediterranean and the Black Sea²⁰.

Despite these difficulties, the content and type of information in the logbooks contribute greatly to the specific questions and topics regarding the transition from sail to steam in the Mediterranean. One of the major as well as relatively understudied topics is the history of navigation, which so far has mostly focused on the scientific and technical evolution of ocean navigation but not on empirical aspects of navigation²¹. Similarly, very little has been produced²² about navigation in the Mediterranean and the Black Sea. One of the very important aspects of navigation that the data from the logbooks make a significant contribution to is the evolution of trade routes along with the technological transition to steam. The pattern of evolution that emerges in the case of the Greek sailing ships and early steamers up until the late nineteenth century is that they were confined to the Mediterranean and the Black Sea. From the early twentieth century, cargo steamers expanded their routes to Northern Europe and from then until the 1920s they sailed completely outside the Mediterranean in the Atlantic, demonstrating the gradual expansion of Greek steamers from regional to almost global trade routes. In the case of Spanish ships, data from the logbooks show that the trade routes are linked mostly to Spanish overseas possessions or ex-colonies. The two sailing ships *Habana* and *Camilla Primera de Tossa* sailed to the Spanish Caribbean and the latter also to South American Atlantic ports as well as to southern US ports. The *Tordera* sailed around the Iberian coast (Portugal included) and the *Infanta Isabel* along the Spanish coast but also to the Caribbean. The *Leon XIII* operated the longest routes, connecting Britain, the Spanish coast

and the Indian Ocean ports of Aden, Colombo and Singapore up to Manila in the Pacific Ocean²³.

Linked to the trade routes is the sailing course the ships followed during their voyage. Strongly affected by the prevailing winds and currents, sailing ships in particular had to sail on a specific course to each enclosed sea or ocean in order to have a quicker and safer journey. It is very remarkable that Greek steamers on routes to the southern Atlantic followed almost the same course, close to the coast, as sailing ships had done previously. The seasonality of voyages is closely connected to weather conditions and the system of winds and currents. Sailing ships such as the *Asimoula* and the *Countouriotis* were more vulnerable to weather conditions and they usually sailed in the spring, summer and autumn. Another very important aspect of navigation is the duration of the voyage, which was affected not only by the distance or the weather during the voyage but also by special sailing conditions, such as the shallow waters in the Sea of Azov or the difficult entrance to the Kerch Strait in the Black Sea. The duration of the voyage was additionally impacted by the time the ship is under sail but also especially by the time it spent in different ports. Sailing ships used to stay more ashore due to their dependence on a fair wind to set sail. Another factor critically affecting time ashore was harbour facilities. In a big port protected from the weather and with mechanical means which can load and unload a ship quickly and safely, ships did not have to stay very long. By contrast, even steamships such as the *Konstantinos* as late as the 1920s had to stay ashore for almost a month due to repairs and servicing work as well as alterations to the hull to accommodate a different type of cargo (e.g., cattle). The average time Greek steamers spent ashore ranged from 14 to 20 days.

The entries in the logbook during the time that a ship was ashore also provides information about ship operation services²⁴, supplies, crew enlistment, repairs and sometimes local social customs or other external factors which might prolong the duration of stay. Examples are days off such as Sundays, or local festivities as happened to the *Konstantinos* in Buenos Aires in 1925 which delayed the loading or unloading of the ship²⁵. However, oth-

er external, often political, factors could also bring about much more serious problems as was the case with the *Stratigos Favieros* at Marseille in 1877. The ship could not find any freight for almost five months as a consequence of the constitutional crisis in France²⁶. Fifty years later, the captain of the *Konstantinos* protested to his agent because he felt that the dock workers in Poti (then in the Soviet Union) were not working as hard as he wanted. The agent informed the captain that they could not do anything about it because the workers were communists!²⁷

Aside from the information on navigation, trade routes and the duration of the voyage, the events that took place on board and ashore are also of paramount importance in the working and living conditions as well as the social relations and behaviours of *Seafaring Lives*. In the logbooks of the Greek ships examined, however, very few cases of work accidents, illnesses or deaths are reported. Similarly, very few instances of desertions, stowaways or quarrels between the captain and the crew or among the sailors are mentioned on either sailing ships or steamers. The only exception is the logbook of the *Konstantinos* in which there is plenty of evidence of all the above-mentioned types of events. Especially in the case of accidents and injuries, the different way of treating patients on this ship is very interesting. Some were sent to hospital and others, often officers, to a hotel under medical supervision. Another type of information that is perhaps rarely found in logbooks concerns the psychology of the people on board. The captain and owner of the *Asimoula* was overwhelmed by the difficulties he faced during his voyages, and in the margin of the logbook he twice wrote about his disappointment and frustration²⁸. Also, ancient Greek proverbs are written on the cover of the book as well as some very sentimental descriptions of the troubles that the *Asimoula* had during the voyage, which reflects the anxiety and the psychological strain on some people working at sea²⁹.

■ APPROACH OVERVIEW

Figure 1 provides an overview of the overall data management methodology we follow and the tools we pro-

Figure 1. The overall data management approach and the offered tools

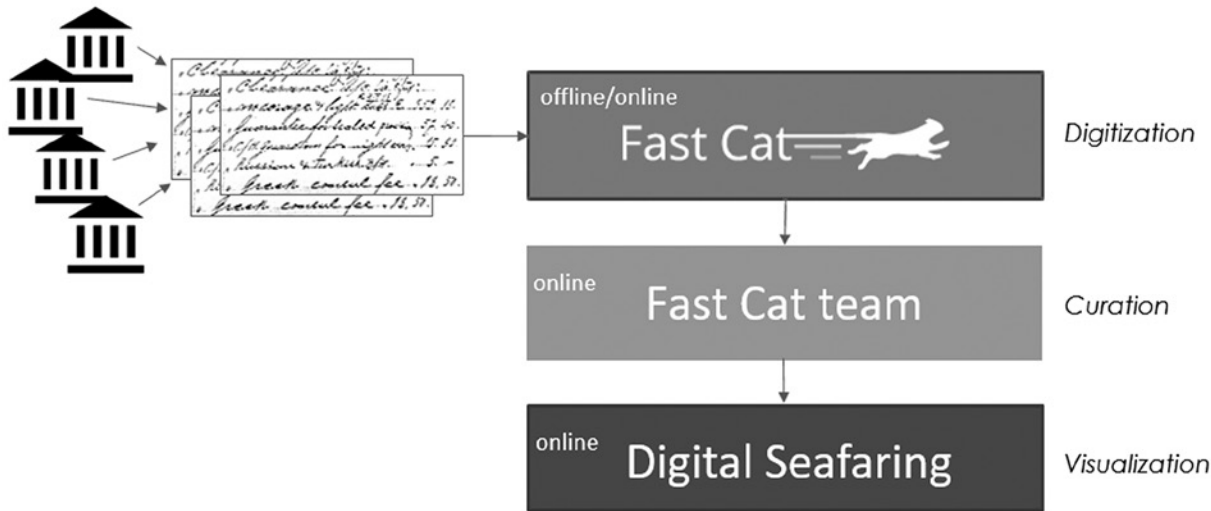
vide for supporting historians in digitizing, curating and exploring their information sources. Below, we briefly describe the tools and their functionalities, while further details are provided in the subsequent sections.

At first, the archival sources of interest (ship logbooks in our case study) are collected from the various providers and digitized using the FAST CAT system. FAST CAT is a web-based system that supports collaborative data entry through the creation of 'templates' and 'records'. A template represents the structure of a single data source, e.g., of ship logbooks, and organizes the data into tables, supporting advanced functionalities such as nesting tables and selection of terms from vocabularies. The advantage of using tables for data entry is twofold: i) historians are familiar with this tabular form due to the widespread use of spreadsheet programs such as Microsoft Excel, and ii) tables facilitate fast data entry and easy detection of errors since the process is iterative and the user can easily inspect the previous entries (of the previously-filled rows). After having defined the templates (in a pre-processing step), historians can start the

digitization process by creating records, where each record belongs to a particular template.

Once the historians have finished the transcription of the archival sources into FAST CAT records, they can start the data curation process through FAST CAT TEAM, a special environment in FAST CAT that allows the collaborative management of *entities* and *vocabulary terms* that appear in the records without needing to alter the original transcribed data, thus maintaining the transcribed data as close to the original information sources as possible. This is very important for historians since data curation is in general an ambiguous process that might require long-term research and repeated revisions. It further enables other historians to work with the original transcribed data and apply their own curation and analysis processes. Thus, decoupling data entry from data curation makes the transcribed data reusable beyond a particular problem or research project.

With respect to the management of entities, in SeaLiT there are four types of entities that can be curated through FAST CAT TEAM: *persons*, *locations*, *ships*, and *legal enti-*



ties. Historians can correct a specific property of an entity, such as the name of a ship or the date of birth of a person, provide additional information, for example coordinates for locations, or indicate that two or more entity instances are actually the same entity and thus must have the same identity (instance matching). Regarding the vocabularies of terms, historians are allowed to provide a preferred term in English (since the source language can be in any language) and a broader term, thus enabling the creation of a hierarchy (thesaurus) of terms in a particular language (English in our case). An important feature here is the capability to check and directly visit the FAST CAT record(s) in which an entity instance or vocabulary term appears, which is important for validation as well as for finding additional context information.

After curation, the data is ready for analysis and exploration by historians and other interested parties. In general, there is a plethora of ways to visualize the data and support their study, e.g., through tables or interactive charts that provide aggregated information. Of course, this greatly depends on the kind of underlying data. In our case study (ship logbooks), we visualize the data on an interactive map, allowing historians to inspect the voyages of ships as transcribed from the original logbooks and curated in FAST CAT TEAM. As we will detail below, users are able to interact with the map, e.g., by filtering the displayed voyages based on ship, by getting more information about a ship being in a particular location, or by visiting the corresponding FAST CAT records.

■ DATA TRANSCRIPTION AND CURATION

Data Transcription

The first step before starting the digitization of the original archival sources is the creation of the ‘templates’, each one representing a distinct archival source. This is performed in a pre-processing step in close collaboration between historians and data engineers. This collaboration is necessary for better designing the structure

of the data entry tables in a template in a way that enables historians to accurately and quickly digitize the archival data. After having created the templates, historians are able to start the digitization process by creating ‘records’ through the FAST CAT system.

Figure 2 shows the homepage of FAST CAT. The interface is quite simple: there is a menu on the left and a table with all the templates on the right. For each template, the table provides some basic information, in particular the *title* of the template, its related *categories* (keywords that provide information about the type of the included information), its related *organizations* (responsible for the data entry process), and the *source languages*. The user can select a template and create a new record to start digitizing a particular archival source. For creating the record, the user first needs to provide some basic information by filling in a small form. This information is different for each template. Figure 3 shows the form for the case of the template “Logbook”, where the user has to provide the following information: *Type of Ship*, *Name of Ship*, *Date of Document (From/To)*, *Author of Record* (the name and surname of the historian performing the data entry). After filling the table and clicking the “Create” button, the record is created and the user can start filling the tables, as shown in Figure 4.

The first two tables of each record contain some basic metadata information about the record and the archival source. The first table has the title “Fast Cat Record Information”. The user can provide the name and role of each author (historian performing the data entry), while the ID, the creation date and the last modification date of the record are automatically filled by the system. The second table has the title “Source identity”, allowing the user to provide information about the archival source such as the name and location of the archive or library, the collection title, the book title and date, the name and location of the issuing authority, etc. (the information requested depends on the type of the archival source). The remaining record tables are different for each source. For instance, for the case of logbooks, the template contains two additional tables: ‘Ship Identity’ and ‘Voyage Calendar’. In the ‘Ship Identity’ ta-

Figure 2. The homepage of FAST CAT

ble, the user can provide the following information: ship name, other ship name(s), ship type, telegraphic code, registry port(s), registry number and year, tonnage, captain name and surname, owner(s) name and surname (if the owner is one or more persons), organization(s) name and headquarters location (if the owner is one or more organizations). In the 'Voyage Calendar' table, the user can provide the following information: digital source page, ashore/at sea (selection between the two), route (from location/to location), in port, duration (calendar type, from date, to date), analytic calendar. The last column (analytic calendar) is a nested table where the user can provide a lot of information about events related to a particular voyage such as weather conditions, course of ship, speed value and type, coordinates (latitude, lon-

gitude), route (movement type, from location, to location, distance from shore), type of event, event description, time of event, related person(s)/organization(s), information about loaded goods/things, and others.

A column in a table accepts values of different types, in particular: i) *entity* (the value is the name or an attribute of an entity, e.g., of a person), ii) *vocabulary term* (the value is a term from a controlled vocabulary), iii) *literal* (the value is a literal, e.g., a free text, a number, or a date).

All the records created by a particular user can be viewed by clicking "My records" in the left menu on the FAST CAT homepage (Figure 2). From there, the user can edit a record at any time, export it in JSON format (as stored in the database), delete it, or share it on FAST

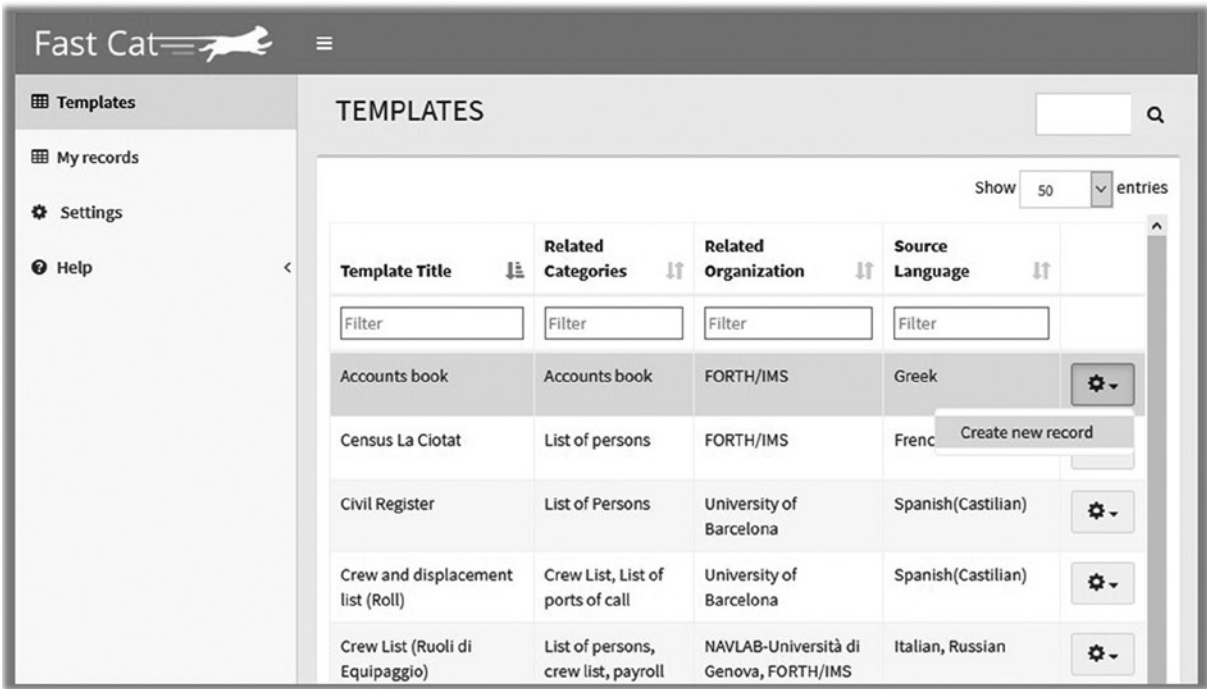


Figure 3. Data requested for creating a record of the template 'Logbook'

Figure 4. A record belonging to the template 'Logbook'

The screenshot shows the 'Fast Cat' application interface. On the left is a sidebar with navigation links: 'Templates', 'My records', 'Settings', and 'Help'. The main area displays a form for creating a record. The form has a 'Template Title' field set to 'Logbook'. Below it are several input fields: 'Type of Ship' (a dropdown menu), 'Name of Ship' (a text field), 'Date of Document (From)' (a date picker), 'Date of Document (To)' (a date picker), 'Author of Record (Name)' (a text field), and 'Author of Record (Surname)' (a text field). At the bottom right of the form are 'Create' and 'Cancel' buttons.

CAT TEAM for curating its entities and vocabulary terms (more below). Furthermore, the user can at any time export the transcribed data of a record in Excel or XML format (Figure 4; gear icon at the upper right corner).

Data Curation

After completing the data transcription process in FAST CAST, users can share the records in FAST CAT TEAM and start the *collaborative* data curation and enrichment process. Figure 5 shows the FAST CAT TEAM homepage, where the user is shown a table containing all the publicly shared FAST CAT records as well as a left menu with further options. For each record, the table shows the following information: ID, template title, record title, record author(s), related template categories,

The screenshot shows the 'Fast Cat' application interface. At the top is a header bar with the 'Fast Cat' logo and a gear icon. Below the header is a title bar for the record: 'Logbook, Κουντουριώτης, Ιστοφόρο, 1990-10-01_1900-11-30, Pavlos Fafalios'. The main area is divided into three sections: 'FastCat Record Information', 'Source Identity', and 'Ship Identity'. Each section contains a table with data for the record.

FastCat Record Information

Use english to fill in the fields of this table

	Date		Authors		
Id	Creation date	Last Modified	Name *	Surname *	Role
1	2020-10-02		Pavlos	Fafalios	

Source Identity

Use the source language to fill in the fields of this table

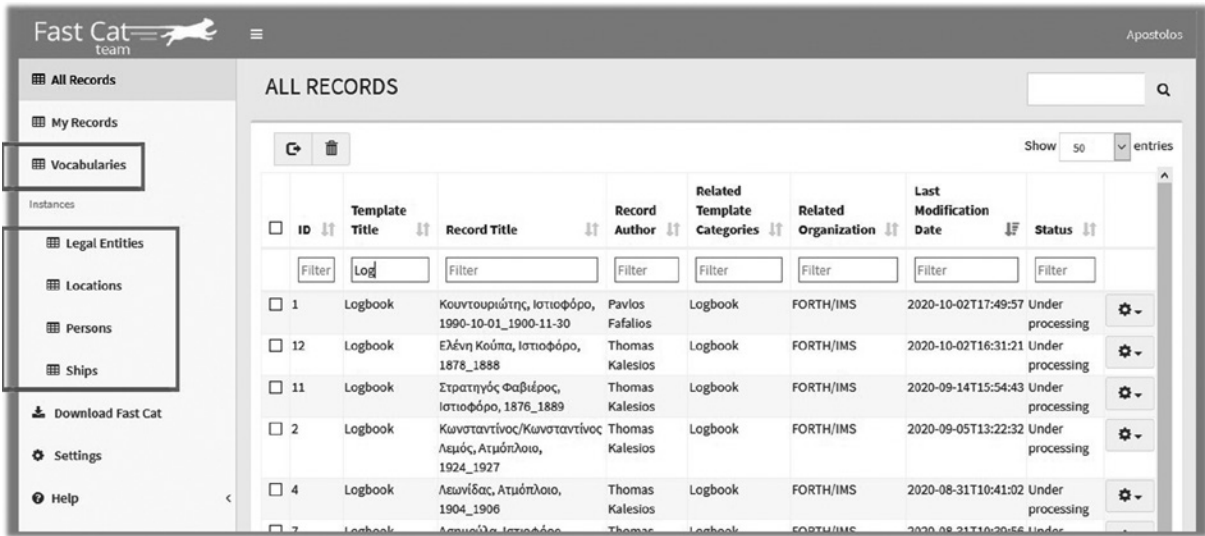
Archive / Library		Book					Issuing authority				
Name	Location	Collection Title	Original Title	Archival Title	Number	From *	To *	Within	Name	Location	Note
						1990-10-01	1900-11-30				

Ship Identity

Use the source language to fill in the fields of this table

										Owner				
				Registry		Captain		Person		Organization				
Ship name *	Other Ship Name	Ship type *	Telegraphic Code	Port	Number	Year	Tonnage	Name	Surname	Name	Surname	Name	Headquarters Location	Note
Κουντουριώτης		Ιστοφόρο												

Figure 5. The homepage of FAST CAT TEAM



related organizations, last modification date and status. The status of a record can take one of the following values: *under processing* (data entry is still in process), *ready for review* (data entry has been finalized but review and/or curation is still in process), *reviewed and ready for publishing* (data entry and curation has been finalized and the data is ready for 'publication', i.e., exploitation by other services), and *published* (the data has been made available to other services for exploitation). The user can select one or more records and view them (without edit rights), export them, delete them or change their status. Since the records might be numerous, the user can also filter the shown records by writing some text in one of the shown table columns or use the search box in the top-right corner.

From the left menu (Figure 5), the user can click 'My Records' and view a table with only the records authored by the user, click 'Vocabularies' and start curating the vocabulary terms that appear in the records, or click one of the entity categories ('Legal Entities', 'Loca-

tions', 'Persons', 'Ships') and start curating the entity instances that appear in the records.

On clicking the 'Vocabularies' menu item, the user is shown a table with all the vocabularies (Figure 6). For each vocabulary, the table shows its title, its source language(s), the template(s) in which it is used, and the related organization(s) (responsible for the templates). Next to each vocabulary title, the table also shows the approximate number of vocabulary terms. The user can select a vocabulary from the table and edit or export it (in JSON format). By selecting to edit a vocabulary, the user is redirected to a webpage showing a table with all the terms of the selected vocabulary (Figure 7). For each term, the table shows its value as it appears in the transcribed record(s), its preferred value (in English) and its broader term. The preferred term value and the broader term can be filled by the user by clicking 'Edit term' (Figure 7; gear icon at the right of the last row column). The user can also inspect and visit the FAST CAT records in which the term appears by clicking the infor-

Figure 6. Management of vocabularies
in FAST CAT TEAM

Figure 7. Inspecting the terms of a vocabulary
in FAST CAT TEAM

FAST CAT TEAM Apostolos

VOCABULARIES Search

Show 50 entries

Vocabulary Title	Source Language	Used in template	Related Organization
Ship	Filter	Filter	Filter
Course of ship	≈ 1311 Greek	Logbook	FORTH/IMS
Part of Ship	≈ 157 Greek	Logbook	FORTH/IMS
Ship type	≈ 76 Greek Spanish(Castilian) Italian Russian	Accounts book Logbook Payroll Crew and displacement list (Roll) General Spanish Crew List Naval Ship Register List Crew List (Ruoli di Equipaggio) Register of Maritime workers (Matricole della gente di mare) Notarial Deeds Payroll of Russian Steam Navigation and Trading Company	FORTH/IMS University of Barcelona NAVLAB-Univ NAVLAB-Univ

Ship type (Expanded View):

- Go to vocabulary
- Export vocabulary

FAST CAT TEAM Apostolos

VOCABULARIES > Ship type

Show terms used in template type: All and record: All Show 50 entries

Export to Excel

Original Source Term	Preferred Term (English)	Broader Term	Status
Filter	Filter	Filter	Filter
Nave	Ship / Full-rigged ship		Under processing
goletta	Schooner		Under processing
schooner	Schooner		Under processing
scooner	Schooner		Under processing
Scouner	Schooner		Under processing
Clipper	Clipper		Under processing
Brigantino	Brig		Under processing

Nave (Expanded View):

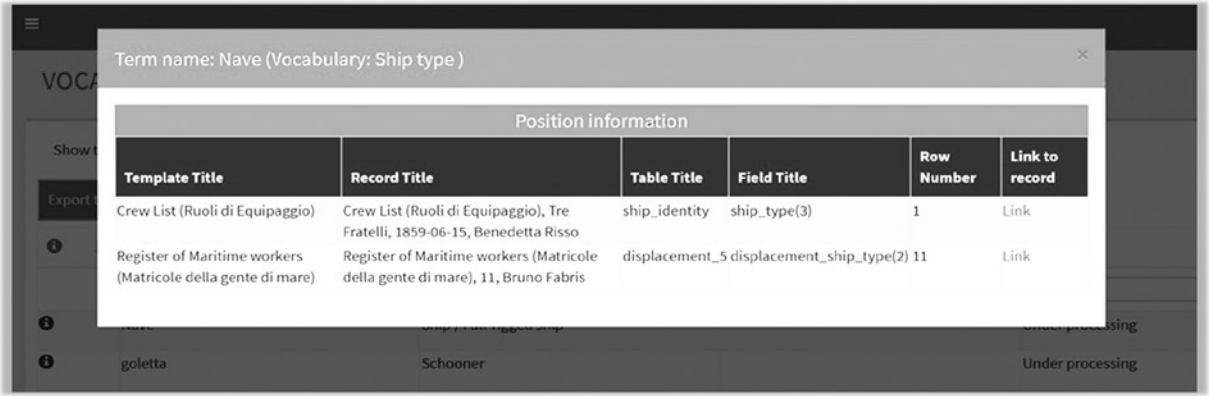
- Go to vocabulary

Figure 8. Inspecting the FAST CAT records in which a vocabulary term appears

information icon to the left of each term (Figure 8), thereby getting additional context information that can help the user to disambiguate or better understand the term.

For the curation of the entities that appear in the records (Legal Entities, Locations, Persons, Ships), the user can click on the corresponding menu item (Figure 5) and inspect a table with all the instances of the selected entity type. The shown entity instances can be filtered by template type and/or record, as well as exported to Excel allowing for further external analysis. For each different entity type, the table displays different information and the user has different curation options.

is a conflict, or mark as different one or more matched instances (Figure 9). For Ships, the user is shown the following data: *name, previous name, type, call signal, construction location, construction date, telegraphic code, flag, owner company, registration list, registration number, registration location*. Similar to the case of person entities, the user can correct one of the attributes or select two or more ships and indicate that they correspond to the same ship (manual instance matching). For Legal Entities, the system only shows the *name* of the legal entity as it appears in the record(s), allowing the user to change its value and set a preferred one.



Template Title	Record Title	Table Title	Field Title	Row Number	Link to record
Crew List (Ruoli di Equipaggio)	Crew List (Ruoli di Equipaggio), Tre Fratelli, 1859-06-15, Benedetta Risso	ship_identity	ship_type(3)	1	Link
Register of Maritime workers (Matricole della gente di mare)	Register of Maritime workers (Matricole della gente di mare), 11, Bruno Fabris	displacement_5	displacement_ship_type(2)	11	Link

For Persons, it shows the following data/attributes: *name, surname A, surname B, maiden name, father's name, place of birth, date of birth, date of death, registration number, status/capacity/role*. Note here that some of these attributes might be empty if the information is not in the transcribed records. The user can select a person and change the value of one of their attributes, select two or more persons and indicate that they correspond to the same person (manual instance matching), and also indicate the preferred value if there

Finally, for Locations the system shows the following data: *source location name, corrected/vernacular location name, other location name, location type, broader location name, ID (TGN or Geonames ID), coordinates*. The user can select a location and correct or provide one of its attributes, as shown in Figure 10. Here the system offers the capability to directly query external geolocation services, in particular the Getty Thesaurus of Geographic Names (TGN)³⁰ and Geonames³¹ (ASK GETTY and ASK GEO buttons), and get the unique ID of

Figure 9. Curation of Person instances in FAST CAT TEAM

the location as well as its coordinates. When querying these services, the system retrieves and shows a list with the relevant locations, allowing the user to select the correct one. When selecting one of the retrieved locations, the system shows the location on a map (helping the user to directly check whether the location is correct) and the coordinates are automatically filled. If the user is not sure about the exact location, they can tick the checkbox 'Location certainty' and optionally include a comment. Linking the locations to coordinates is very important in our case since it enables the visualization of the transcribed data on a map (as we will see below).

■ DATA VISUALIZATION: THE 'SHIPS VOYAGES' APPLICATION

After the curation process has been completed for one or more of the transcribed records, the curated data can be exploited by external systems. Below we describe such a system, called *Ship Voyages*, which visualizes on a map the transcribed and curated ship logbooks of the SeaLiT project.

The system is a publicly-available web application³², which is part of a set of web applications with the umbrella name 'Digital Seafaring'³³, developed as part of the SeaLiT project. The user interface is an *interactive map*

PERSONS ?

Show instances used in template type: All and record:

Mark as same ResetTable Export to Excel

Name	Surname A	Surname B	Maiden name	Fathers name	Place of Birth	Date of Birth	Date of Death	Registration number	Status	Capacity	Role	Status
Virgilio	Biso					1843		4222	Under processing			
Pasquale	Cichero					1840		11068	Under processing			
Michele	Fortunato											
Bertolotto	Gerolamo											

Mark as same

Name: ☒ Virgilio ☐ Pasquale

Surname A: ☒ Biso ☐ Cichero

Date of Birth: ☒ 1843 ☐ 1840

Registration number: ☒ 4222 ☐ 11068

selection of preferred value

Save changes Cancel

Add/Remove preferred term

Name: ☒ Pasquale ☐ Virgilio

Surname A: ☒ Cichero ☐ Biso

Date of Birth: ☒ 1840 ☐ 1843

Registration number: ☒ 11068 ☐ 4222

change of attribute value

Save changes Cancel

Position information

Template Title	Record Title	Table Title	Row Number	Link to record	Mark as different
Logbook	Logbook, Κωνσταντίνος/Κωνσταντίνος Ατμόπλοιο, 1924_1927, Thomas Kalesios	ship_identity 1	Link	Mark as different	
Logbook	Logbook, Κωνσταντίνος/Κωνσταντίνος Ατμόπλοιο, 1924_1927, Apostolos Delis Thomas Kalesios	ship_identity 1	Link	Mark as different	

Unmatch an entity instance

Figure 10. Curation of Location instances in FAST CAT TEAM

annotated with ship routes, where a ship route is defined as a set of ship locations as documented in the logbooks.

Figure 11 shows a screen dump of *Ship Voyages*. On the left of the user interface, the user can see a list with all ship routes grouped by ship name. By clicking the name of a ship, all routes corresponding to that ship are selected and displayed on the map. The user can also select and visualize one or more individual ship routes or filter the list of the displayed routes based on some keywords. When the user moves their mouse over a ship icon, the route of the ship is highlighted. Here different background colours are used for explaining how the location coordinates have been derived. The white back-

ground colour means that the location coordinates have been found in the original source. The yellow background means that the location coordinates were written in the original source using a different coordinate system that is not supported by the current map applications (e.g., Cadiz Meridian), so the historians had to manually add the coordinates through FAST CAT TEAM. Finally, the red background means that there are no location coordinates in the original source and only the location names are given, so the historians added them through FAST CAT TEAM.

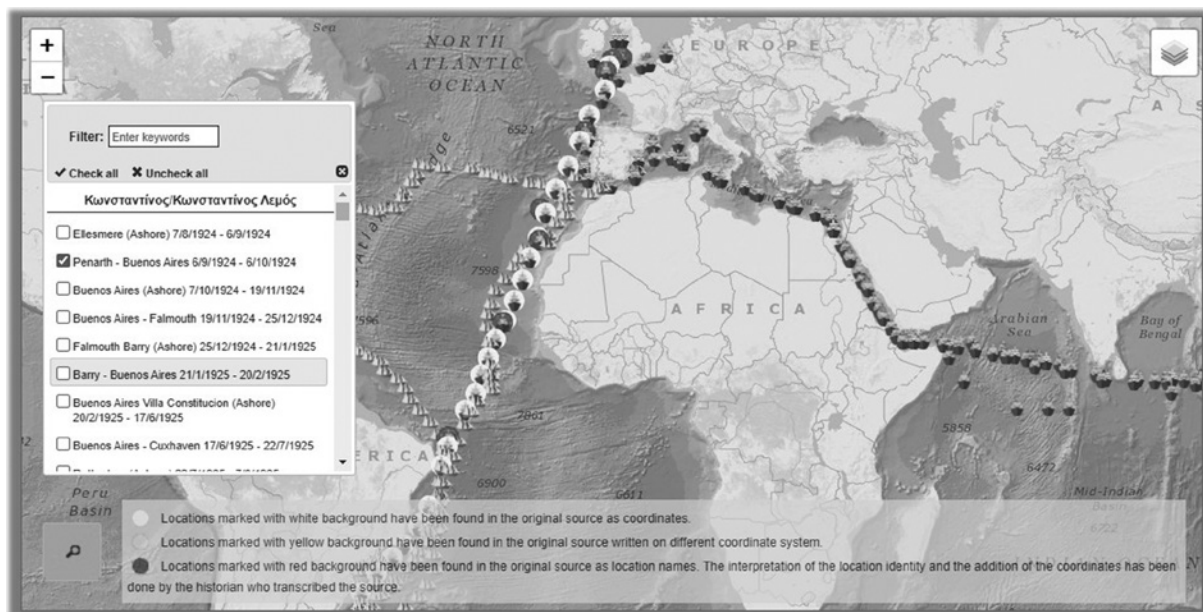
By clicking a ship icon, the user gets more information about the selected ship location from the tran-

The screenshot displays the FAST CAT TEAM interface with a modal window titled "Location: Santa Pola" open. The modal contains the following fields and options:

- Corrected/Vernacular Location name:** "Santa Pola" with a dropdown arrow. A note below states: "this will replace the source location name."
- Buttons:** "Ask Getty" and "Ask Geo".
- TCN ID (Getty Location database):** "santa pola".
- Geonames ID (Geonames):** "2511102".
- Coordinates:** "38.19165 -0.5658".
- Secondary optional location name (e.g. historical name):** A text input field with the placeholder "Fill in second name for the location (optional)".
- Location certainty:** An unchecked checkbox with the text "I am not 100% sure about this location identity".
- Buttons:** "Save" and "Cancel".

In the background, a table of location entries is visible. It has columns for "Coordinates" and "Status". The "Coordinates" column contains values like "-36.35894 -56.72351" and "38.19165 -0.5658". The "Status" column contains the text "Under processing". Each row has a gear icon for settings.

Figure 11. Visualizing the transcribed and curated data on a map



scribed logbook such as ship name and type, an image of the ship, location date and time, weather conditions, ship course and related events, e.g., change of the ship's course (Figure 12). By clicking the ship image, a pop-up window shows the image in higher resolution, while the user can also directly visit the corresponding FAST CAT record by clicking the 'View Source' hyperlink, thereby allowing the user to inspect additional context information as transcribed from the original sources.

■ APPLICATION IN HISTORICAL RESEARCH

Ships Voyages is the first, and only to our knowledge, map visualization based on data from ship logbooks in an online open access form. It is unique because it allows the user to follow the ships' routes as a whole and at each separate observation point of the voyage (usually recorded once a day, around noon, in the logbook). It al-

so allows the user to visit the transcribed metadata from the source, to check the information or to read other types of information from the logbooks that do not appear in the map visualization, such as events, and also to retrieve the information. Therefore, *Ships Voyages* is a highly sophisticated platform that is designed to enable a deep level analysis of metadata and of the content of the source. Thus the application is not only informative, educative and entertaining for the broader public, but also very useful to academic users for research purposes. It is also unique in terms of historical subject, since it is about the voyages of merchant ships from Mediterranean countries during the transition from sail to steam navigation. Hence it provides the opportunity to study the evolution of the routes and of several aspects of navigation in two different technological systems of sea transportation vehicles, the sailing ship and the steamship. It

Figure 12. Inspecting a ship location on the map

also encompasses trade routes and various aspects of navigation in enclosed seas, including the Mediterranean, the Black Sea, and the English Channel, as well as in oceans, such as the Atlantic, the Indian and the Pacific.

One of the first and extremely important advantages that *Ships Voyages* offers is a very clear visualization of the ship's routes. Although this might seem self-evident or trivial, visual effects are of paramount importance for understanding fundamental aspects of navigation. First, it operated as a correcting tool for the observation points and routes entered into the database. Points that were shown on land instead of at sea, due to different meridian systems adopted by the captains and other related problematic issues that arose after the entry of data, were corrected and solved thanks to the map visualization. Then on a single route, for instance between a

mainland Spanish and a Spanish Caribbean port, it shows exactly the course the ship had followed. If this observation is applied on several voyages on the same route, it enables us to build up the navigation pattern and any variations followed by sailing ships and/or steamers on the specific trade route over time (e.g., Spanish sailing ships and steamers in the Caribbean). Third, the selection of several or all the routes of a specific vessel or a category of vessels (e.g., Greek cargo steamers) illuminates the evolution and perhaps the extent and the type of expansion of trade routes from regional to global waters over time. Fourth, the meteorological information contained in the logbooks and visualized in the mapping of the *Ships Voyages* may function as a historical database on climatic information. Especially in major projects on climatology such as CLIWOC that created a climatic

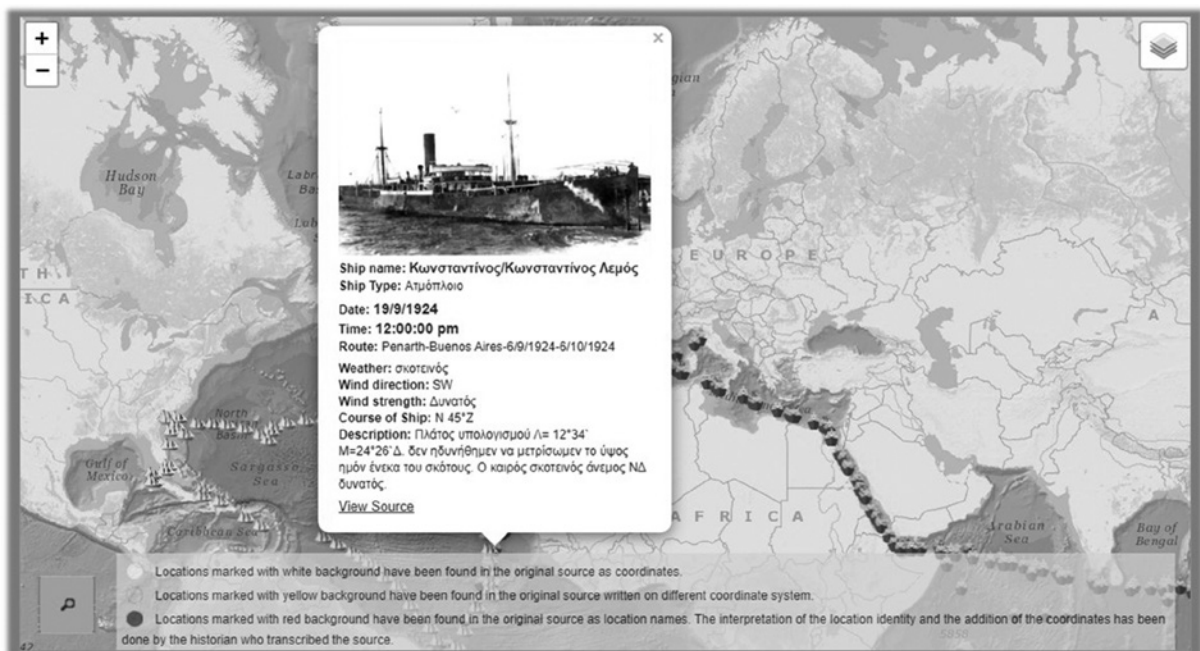
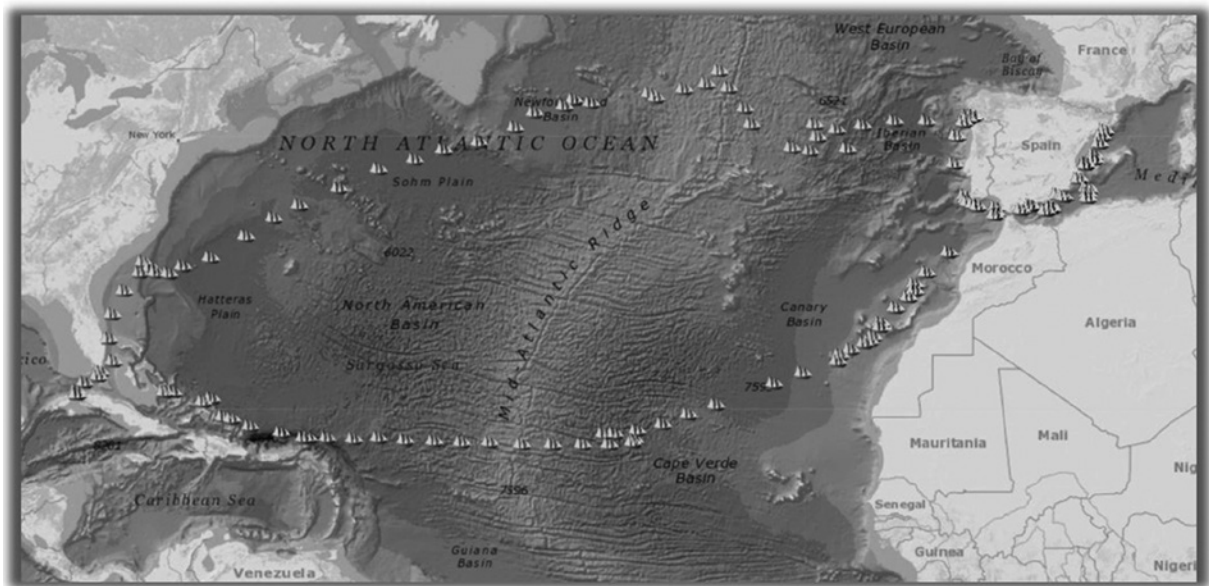
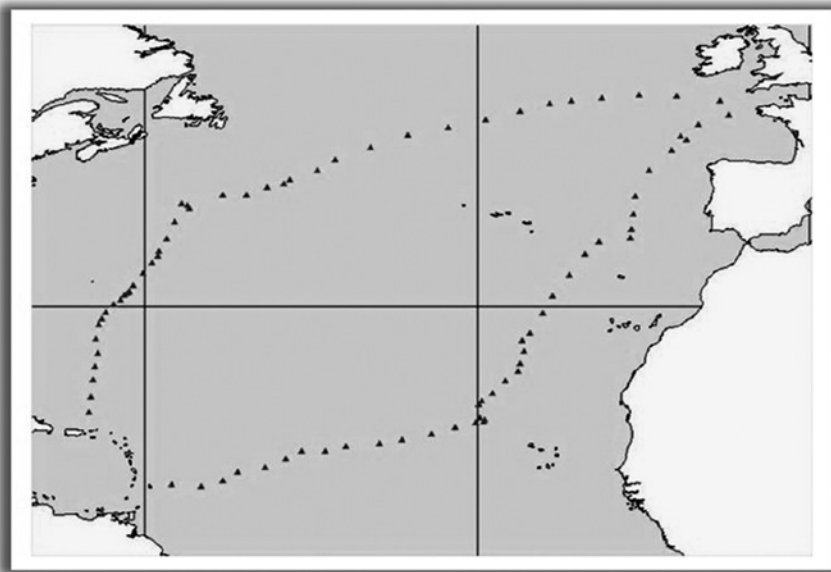


Figure 13. The West Indies route compared (sources: Wilkinson, "British Logbooks", 21, 83; Ship Voyages, https://isl.ics.forth.gr/FastCatTeam/templates/ship_map.html)



data series based on ocean routes in ships' logbooks for the period 1750 to 1850, *Ships Voyages* covering the immediately following period in both enclosed seas and oceans may be useful as an expansion of historical climatic data in time and space³⁴.

The visualized results of *Ships Voyages* can be also compared with results from other research projects and may corroborate findings on some topics. For instance, Figure 13 compares the routes to the West Indies and back of HMS *Acasta* between 1807 and 1809, examined by Wilkinson as a result of the CLIWOC project, and of the merchant barque *Habana* between April and September 1892 as shown in *Ship Voyages*. The similarity of the sailing course is remarkable in both cases, on the outward leg of the voyage taking the route south of the 30th parallel north and sailing with the advantage of the NE trade winds, and on the homeward leg, the route north of the Bahamas and the Azores.

■ CONCLUSION

We have presented an approach and a set of tools that support historians in digitizing, curating and then visualizing original archival sources of maritime history. We described the common data management challenges faced by historians in the context of a European research project, called SeaLiT, and demonstrated the tools for the case of ship logbooks from the nineteenth and twentieth centuries. The general objective is to support historians in collaboratively creating, curating, maintaining and exploring a rich database of archival data transcribed from original historical sources through user-friendly (web-based) interfaces.

The described tools FAST CAT and FAST CAT TEAM are innovative in the sense that they support features such as nested tabular structures for data entry, embedded instance matching and vocabulary maintenance processes as well as provenance-aware data curation, important features that are not currently supported by existing solutions. In addition, they are configurable which means that they can easily be used for digitizing and curating other data sources beyond the case of mar-

itime history. Further information on how to use these tools is available at <https://www.ics.forth.gr/isl/fast-cat>.

The *Ships Voyages* application, accessible at <http://www.sealitproject.eu/digital-seafaring>, is to our knowledge the first open access map visualization based on data from ship logbooks, allowing users to follow the routes of ships but also visit the related transcribed data and check the information or find other information that does not appear in the map visualization, such as events that took place during a particular ship voyage.

Currently, we are studying additional methods on how to explore the curated data, focusing on how to support historians in expressing complex information needs through intuitive user interfaces and visualize the results in different ways, such as tables or charts.

■ ACKNOWLEDGEMENTS

The authors would like to thank Prof. Gelina Harlaftis for her review of this paper. It has received funding from the European Union's Horizon 2020 research and innovation programme under i) the European Research Council (ERC) grant agreement No 714437 (Project "*SeaLiT - Seafaring Lives in Transition. Mediterranean Maritime Labour and Shipping during Globalization, 1850s-1920s*"), and ii) the Marie Skłodowska-Curie grant agreement No 890861 (Individual Fellowship, Project "*ReKnow - Research Knowledge Documentation, Analysis and Exploration in Empirical and Descriptive Sciences*").

■ NOTES

1. SeaLiT - Seafaring Lives in Transition. Mediterranean Maritime Labour and Shipping during Globalization, 1850s-1920s. ERC Starting Grant (ID: 714437), 1 February 2017 - 31 January 2022, <http://www.sealitproject.eu/>

2. Probably one of the first to methodically study logbooks as a source of raw information was Matthew Fontaine Maury (14 January 1806 -1 February 1873), a US naval officer who studied thousands of ships' logs and charts. Among other works, he published the *Wind and Current Chart of the North Atlantic*, a book that showed sailors how to use the ocean's currents and winds to their advantage, drastically reducing the length of ocean voyages.

3. Oliver, J., & Kingston, J. A., "The usefulness of ships' logbooks in the synoptic analysis of past climates", *Weather*, (December 1970), 520-28.

4. Ricardo García-Herrera, Phil Jones, Dennis Wheeler, Gunter Können, Maria Rosario Prieto (eds.), CLIWOC: Climatology of the World's Oceans, 1750-1850, *Climatic Change* 73, no. 1-2, (November 2005) (10 articles in this issue); Barriopedro, D., Gallego, D., Alvarez-Castro, M.C., et al. "Witnessing North Atlantic westerlies variability from ships' logbooks (1685-2008)" *Climatic Dynamics* 43, (2014): 939-55; García-Herrera, R., Barriopedro, D., Gallego, D., Mellado-Cano, J., Wheeler, D., & Wilkinson, C. "Understanding weather and climate of the last 300 years from ships' logbooks.", *WIREs Climatic Change* 9, (2018):e544, <https://doi.org/10.1002/wcc.544>; Clive Wilkinson, "British Logbooks in UK Archives 17th -19th Centuries. A survey of the range, selection and suitability of British logbooks and related documents for climatic research" *Climatic Research Unit Research Publication* 12; Wheeler, D., "Understanding seventeenth century ships' logbooks: An exercise in historical climatology", *Journal for Maritime Research*, (March 2004), published electronically at www.jmr.ac.uk; Wheeler, D., "The weather vocabulary of an eighteenth century mariner: the log books of Nicholas Pocock", *Weather* 50, (1995), 298-304.

5. Brown, S., Credland, A., Savours, A., & Stonehouse, B., "British Arctic whaling logbooks and journals: A provisional listing" *Polar Record* 44, 4, (2008): 311-20; Lesley Evans Ogden, "New Data from Old Treasures: Whaling Logbooks", *BioScience* 66, no. 7, (July 2016) 620.

6. Anne Farrow, *The Logbooks: Connecticut's Slave Ships and Human Memory*, (Middletown CT: Wesleyan University Press, 2014).

7. Meaghan Walker, "In the inventories of deceased British merchant seafarers: exploring merchant shipping and material culture, 1860-1880", *International Journal of Maritime History* 31, no. 2, (2019):330-46.

8. Morgan Kelly & Cormac Ó Gráda, 2014. "Speed under Sail, 1750-1850," Working Papers no. 14/10, School of Economics, University College Dublin; Dagomar Degroot, *The Frigid Golden Age: Climate Change, the Little Ice Age, and the Dutch Republic, 1560-1720*, (Cambridge: Cambridge University Press, 2018).

9. <https://blogs.bl.uk/asian-and-african/2020/05/digitised-east-india-company-ships-journals-and-related-records.html?fbclid=IwAR1JCJJB9xNH0xGqfu2lZDgEvUXq2N96HVwRuJK0ViDrOuBB7a0UgSZGOOM>; <https://www.whalingmuseum.org/explore/library/logbooks-journals> (accessed on 4 November 2020)

10. Georgia P. Koulikourdis, *Ο Αλέξανδρος του Χατζή Αλεξανδρή, ένα πολεμικό καράβι των Ψαρών ημερολόγιο και δράση* (1821-1838)

[The Alexandros of Hadjialexandris, a fighting vessel of Psara. Journal and action (1821-1838)]. (Athens, 1972); Foula Pispirigou, *Το Ημερολόγιο του «Κίμωνος»* [The Logbook of Kimon], (Athens: Potamos 2012); Goergios Sacthouris, *Ιστορικά ημερολόγια των ελληνικών ναυμαχιών του 1821. Εκ των ημερολογίων του ναυμάχου Αν. Τσαμαδού*, [Historic Journals of the Greek Battles of 1821. From the Journals of Captain Anastasios Tsamados], (Athens: Karavias, 1978); G. Vlassopoulos, *Φουρτούνες και μπουνάτσες, τα καράβια μας στη Μεσόγειο του 19ου αιώνα, ανέκδοτα ημερολόγια τριών εμπορικών καραβιών*, [Storms and Calms. Our ships in the nineteenth century Mediterranean. Unpublished logbooks of three merchant ships], (Athens: Papazisis 2003) and, *Οδυσσέας, ένα καράβι της Ιθάκης 1837-1841* [Odysseus, a ship from Ithaki, 1837-1841], (Athens: Melissa, 1992).

11. Maria Konstantinidis, 2007, "*Δυο Ναυτικά Ημερολόγια του τέλους του 19ου αιώνα. Του Ιστιοφόρου «Αλέξανδρος Γ'»* (1884-1888) από το Ναυτικό και Ιστορικό Μουσείο Γαλαξειδίου και του Ατμοπλοίου «Θηρεσία» (1888-1890) από τη Βιβλιοθήκη «Ο Κοραΐς» της Χίου" [Two ship logbooks of the late 19th century. The sailing ship Alexandros the Third (1884-1888) from the maritime and historical museum of Galaxidi and the steamship Thiresia (1888-1890) from the Korais Library of Chios.], MA Thesis, Ionian University.

12. G. A. Rallis, *Ερμηνεία του Ελληνικού Εμπορικού Δικαίου*, [Interpretation of Greek Commercial Law], (Athens: 1863), 32-35, 89, 118, 159-60.

13. Clive Wilkinson, "The Non-Climatic Research Potential of Ships' Logbooks and Journals" *Climatic Change* 73, no. 155 (2005): 157; Credland, Savours, & Stonehouse, "British Arctic Whaling", 313.

14. Charles W. J. Withers, *Zero Degrees. Geographies of Prime Meridian*, (Cambridge Mass., London: Harvard University Press, 2017), 5-14; García-Herrera, et al., "Understanding weather", 4.

15. García-Herrera, R., Wilkinson, C., Koek, F.B., et al. "Description and General Background to Ships' Logbooks as a Source of Climatic Data", *Climatic Change* 73, (2005), 25.

16. Withers, *Zero Degrees*, 13.

17. The divergence of the Hierro Meridian is 18° 09' west of Greenwich. In the case of the Tenerife Meridian, the divergence is 16° 39'14" west of Greenwich.

18. The divergence is 3° 41' 16" west of Greenwich.

19. France used the Paris Meridian until 1911.

20. In the logbooks of the steamship *Andriana*, entry of 22.6.1909, and the *Leonidas*, entry of 26.9.1905, the port of Cadiz is mentioned with its ancient Greek name Gadeira (Γάδειρα). However, the most common use of ancient Greek names concerns place names around Istanbul, the Black Sea and the Adriatic. In that latter sea, chaos awaits the researcher as the same

place might have an Italian, German, ancient Greek or a variety of Slavic names.

21. Willem F.J. Mörzner Bruyns, "Research in the History of Navigation: Its Role in Maritime History", *International Journal of Maritime History* 21, no. 2 (December 2009): 216-86.

22. Apostolos Delis, "Navigating perilous waters: routes and hazards of the voyages to Black Sea in the nineteenth century", in *Linkages of the Black Sea with the West. Navigation, Trade and Immigration*, eds. M.C. Chatziioannou & A. Delis (Rethymnon, 2020), 1-33.

23. https://isl.ics.forth.gr/FastCatTeam/templates/ship_map.html (accessed on 4 November 2020)

24. For an example about ship operation services through processed data from Greek ships of the nineteenth and twentieth centuries, the reader can see and explore the data base: https://isl.ics.forth.gr/FastCatTeam/templates/ship_chandlers.html.

25. Logbook of the steamship *Konstantinos/Konstantinos Lemos* (1924-1927), entries of 22.2.1925 to 24.2.1925; 22.2.1925 was a day off because it was Sunday followed by the next two days. At that time there was the big celebration of carnival that cost the ship three days without work, but also on the other hand gave a unique opportunity for the crew to enjoy the celebration.

26. Logbook of the sailing ship *Stratigos Favieros* (1876-1889), entries of 2.5.1877-13.10.1877.

27. Logbook of the steamship *Konstantinos/Konstantinos Lemos* (1924-1927), entry of 5.8.1927.

28. Logbook of the Greek sailing ship *Asimoula* (1882-1883): entry of 3/15.9.1882 (in the margin of the page) and 7/19.9.1882. Afterwards the new captain also writes in the entry of 16/28.7.1883: "[...]μου έσπασε η βίδα και έμεινα στα κρύα του λουτρού όθε ειδείς το ένα καλό περίμενε και το άλλο. Ο Θεός να τα φέρει ολίγον ευνοϊκότερα διότι είμαι επόμενος να χάσω το κεφάλι μου." [The screw broke and I don't know what to do, one difficulty after the other. God help me because I am going to lose my mind!].

29. Logbook of the Greek sailing ship *Asimoula* (1882-1883): "Ουδέν! αγαθόν! αποκτάτε! χωρής! κόπον!"; a rough translation might be 'no pain, no gain'. 8/20.9.1882: "Είμαι νήσος του Αιγαίου και ανέθρεψα συνάμα τον υιόν του Αχιλλέως, αν με αποκεφαλίσεις όνομα βασιλέως θα σχηματίσεις» Ν. Ευθ.Κ.- Η νήσος Σκύρος" [a riddle for the island of Skiros], and also in the same entry: "Όταν θυμώνεις ούτε λέγε ούτε πράττε-Πλάτων" ['When you are angry it is better not to act or to speak"- Plato]. Entry: 30/11.12.1882 "Σας λέγω κύριοι ότι εις αυτό το ταξίδι υποφέραμε τόσα όσα δεν υποφέραμε εις όλα μας τα ταξίδια" [I say gentlemen, that on that journey we suffered the most from all our journeys].

30. Getty Thesaurus of Geographic Names - <http://www.getty.edu/research/tools/vocabularies/tgn/>

31. GeoNames geographical database - <https://www.geonames.org/>

32. https://isl.ics.forth.gr/FastCatTeam/templates/ship_map.html (accessed on 4 November 2020).

33. <http://www.sealitproject.eu/digital-seafaring> (accessed on 4 November 2020).

34. <http://webs.ucm.es/info/cliwoc/object.htm> (accessed on 4 November 2020).