Testing the waters: the creation of Global Privateering, an open access database for St Malo's dossiers de prise

Boer, T.W.M. de

IJMH

Research Note

Testing the waters: The creation of *Global Privateering*, an open access database for St Malo's *dossiers de prise*

International Journal of Maritime History
2022, Vol. 34(4) 658–674

DOI: 10.1177/08438714221128160
journals.sagepub.com/home/ijh

Tessa de Boer
Leiden University, Netherlands

Abstract

The relevance of privateering-related archives for maritime history and (early modern) history at large has been increasingly recognized in the past decades. This recognition has gone hand in hand with the prominent role that these archives have assumed in some of the largest projects in the digital humanities, such as the *Prize Papers* project, which facilitates research into these collections and reveals the multiplicity of data that they contain. This article reports on the creation of *Global Privateering*, a (pilot) project that is striving to create a user-friendly open access database for a set of privateering *dossiers* from St Malo, France. It has been a collaboration across disciplines – namely, history and information technology – as well as educational establishments, partnering universities and vocational schooling. The creation of this open access maritime database has been a positive experience for all parties involved, as the complementarity of skills and expertise has been evident, and the benefits have been mutual and multiple. This inspires further exploration of such collaborations and the further expansion of the *Global Privateering* platform.

Keywords

Database, Dutch Republic, France, privateering, St Malo

Introduction

Privateering-related documentation has been widely recognised as containing a wealth of information pertaining to maritime history and beyond. However, the often considerable

Corresponding author:
Tessa de Boer, Institute for History, Leiden University, Doelensteeg 16, Leiden 2311 VL, Netherlands.
Email: t.w.m.de.boer@hum.leidenuniv.nl

size of the archives makes the extraction and analysis of this information a daunting process. Advances in digital humanities and digital archiving promise to significantly ameliorate this situation. They continue to radically transform archival and, by extension, historical research.

The most well-known initiative to facilitate research into these privateering-related collections is undoubtedly the ongoing *Prize Papers* project, a multidisciplinary, multi-institutional international collaboration. It encompasses the in-depth cataloguing, digitisation and open access presentation of the British Prize Papers.[1] The historical and historiographical prominence of (early modern) British privateering made this collection a suitable target. However, equally impressive data can be found beyond the Anglo-Saxon context. Contemporary French privateering in particular was anything but negligible, and its surviving paper trail has the potential to equally contribute to inquiries into 'daily lives around the globe in the time of European Expansion, Colonialism, and Resistance'.[2] Unfortunately, there have been no comprehensive efforts to facilitate access to these privateering collections in the French archives and bring them to a wider audience of researchers.

A pilot project, ambitiously named *Global Privateering*, seeks to take the first steps in this regard.[3] On the *Global Privateering* platform (http://globalprivateering.com/), 176 *dossiers de prise* on Dutch(-related) ships captured by St Malo privateers at the end of the Nine Years' War (1688–1697) will be consultable via a user-friendly open access database. The raw data was manually gathered at the Archives départementales d'Ille-et-Vilaine in Rennes by an independent researcher from the Netherlands. In collaboration with a team of researchers from the Leiden University Institute for History and information technology (IT) students from Nova College, an institute for secondary vocational education (*middelbaar beroepsonderwijs* or MBO), the platform was realised in 2022.[4] This research note offers an evaluation of this collaboration between citizen scientists, university-affiliated research and practical/technical expertise found outside of the university. Whereas university–MBO partnerships are commonly found at faculties of natural sciences or medicine (in the Netherlands), they are much less common in the humanities, including in historical research. Our positive experiences with the development of *Global Privateering* encourage further ventures into these collaborations within the historical field, as many (maritime) archival collections still await digital processing.

1. For the project's main website, see https://www.prizepapers.de/ (accessed 15 July 2022). See also updates on the website of the UK National Archives, https://www.nationalarchives.gov.uk/about/news/prize-papers-research-portal-launched/ (accessed 15 July 2022).
2. See https://www.prizepapers.de/ (accessed 15 July 2022).
3. *Global Privateering* is a joint effort between Siem van Eeten, Cátia Antunes, Tessa de Boer, Wouter van Hezel, Kim Quax, Alex Rasterhoff, Florian de Vries, Céline Bijtenhoorn and Maartje Hids. Additional support and valuable input was provided by Gerhard de Kok (and his students) and Tristan de Boer.
4. *Global Privateering* can be found at http://globalprivateering.com.

Provenance of the *Global Privateering* data

In the early modern era, French privateers were a major force (particularly) in European waters. Several French ports were hotbeds of corsairing activity, with sophisticated institutions to deal with the judicial, administrative and economic aftermath of a capture. Alongside Dunkirk, Calais and Brest, St Malo was one of the major corsair ports. It boasted a strong maritime tradition and readily established commercial links with both the Atlantic and Asia.[5] Privateering from St Malo has been attested to since medieval times, and the city steadily built up a reputation as *la cité corsaire*.[6] The Nine Years' War, flowing into the War of the Spanish Succession (1700–1713), caused French and particularly Malouin privateering to increase in intensity and scope. By outfitting private entrepreneurs (*armateurs*) with letters of marque to take enemy ships on behalf of the Crown, and allocating the *armateur* and the privateer a significant share in the profits of the prize's sale, a *guerre de course* was launched against *les ennemis de l'État* , meaning (primarily) against Anglo-Dutch shipping. This was an effective method of economic warfare, as well as an attractive source of economic gain for the entrepreneur and (local) governments alike.[7]

Despite associations with piracy, swashbuckling and general rogue activity, privateering was subject to strict rules and regulations. For France, the *Ordonnance de la Marine* of 1681 was the central guideline. Each capture was submitted to meticulous administration to ensure that the regulations were followed and the capture was thus indeed 'legal'. For example, members of the captured crew were interrogated not only to establish beyond reasonable doubt that the ship could indeed be classified as an enemy vessel (that is, a legitimate prize), but also to verify that they had not been (excessively) mistreated.[8] The crew of the privateer was similarly subjected to interrogation to establish this. Any perceived infractions of maritime law could quickly provoke retaliation and a general escalation of the international conflict at hand. Establishing the legitimacy of the capture also sanctioned the subsequent auction and sale of the freight, and occasionally the ship, and undermined the jurisdictional bases of possible suits brought forth by the original owners of the ship and freight.

This meticulous administration typically left an extensive paper trail of *dossiers*, constituting a valuable source in maritime history and beyond. For St Malo, these *dossiers de prise* are currently housed in the Archives départementales d'Ille-et-Vilaine in Rennes.[9]

5. Henning Hillmann, *The Corsairs of Saint-Malo: Network Organization of a Merchant Elite Under the Ancien Régime* (New York, 2021), 20–60; Guillaume Lelièvre, 'La course aux épices: Malouins et Vitréens dans l'océan Indien au début du XVII^{e} siècle', *Annales de Bretagne et des Pays de l'Ouest*, 125, No. 3 (2018), 247–84.
6. Hillmann, *The Corsairs of Saint-Malo*, 41.
7. Hillmann, *The Corsairs of Saint-Malo*, 60–122; J. S. Bromley, *Corsairs and Navies, 1660–1760* (London, 1987), 43–72, 213–42, 279–96.
8. Maartje Hids, 'In handen van den viant gevallen: De Franse kapingen vanuit Saint-Malo van Nederlandse schepen tussen 1688 en 1713' (Unpublished Master's dissertation, Leiden University, 2022), 12–14; Hillmann, *The Corsairs of Saint-Malo*, 91–121.
9. Archives départementales d'Ille-et-Vilaine (AD 35), Série 9B: Amirauté de Saint-Malo.

A typical *dossier* consists of several dozen pages. Included are, typically, the interrogations of the crew to extract information on the ship, freight, route, owner and so on; depositions by the privateer, reporting on the circumstances of the capture; detailed inventories of the freight and the ship's outfitting; reports on the auctioning, including the identities of the bidders/buyers and their offers; miscellaneous documents, such as letters or contracts found on board; and pieces of bureaucratic documentation throughout to verify or ratify the proceedings. Whereas the prize papers central to the other aforementioned digitisation projects are often synonymized with the seized papers on board the captured ship (relegating the administrative papers secondary in the collection's historiographical utilisation), the administrative documents are the centerpiece of the *dossiers de prise* as an archival collection.

The historical enquiries that can be based on data derived from the *dossiers de prise* are multiple and varied: questions of small, microhistorical scope go hand in hand with analyses of large-scale, *longue-durée* flows of goods and people. For example, extracting the identities and actions of the *armateurs* aids in the reconstruction of local networks of maritime entrepreneurs, as explored in a recent study by Henning Hillmann on St Malo specifically.[10] Analysing the types of freight found on board the captured ships can fine-tune data on early modern trade flows – ports, seasonality, bottom types, composition and so on. Distilling the personal backgrounds of the interrogated crew adds to recent inquiries into the characteristics and composition of early modern seafaring personnel, which in turn contributes to larger debates on premodern labour migration.[11] Aside from these existing studies and debates, new pivotal and complex questions that are in need of answering arise when closely scrutinising the *dossiers*. Think, for example, of the mobilisation of capital in relatively small-scope markets – a question spurred by the recorded ability of local buyers in prize auctions to bid amounts of over 100,000 *livres tournois* on individual lots.[12] All in all, the data captured in the *dossiers de prise* contributes to a wide range of questions both in maritime history and beyond.

Sourcing the source

As with many other archival collections of a considerable size, enquiries such as the above involve considerable time and effort on the part of the researcher to manually gather data. More often than not, one has to contend with only sampling the collection, as opposed to viewing it as a whole. This makes it difficult to tackle more ambitious research. Advances in digital humanities and digital archiving, however, continue to facilitate this process. Digitisation is often the first step, as online open access availability

10. Hillmann, *The Corsairs of Saint-Malo*.
11. See, for example, the project *Maritime Careers: The Life and Work of Dutch Merchant Marine Sailors, 1700–Present* (Principal investigators : Jelle van Lottum and Lodewijk Petram), https://www.maritimecareers.eu/ (accessed 15 July 2022).
12. See, for example, AD 35, Série 9B, 599, Dossier 6: *Fortune* (108.942 + 1,200 livres tournois for cargo and ship, respectively); 600-1, Dossier 1: *Le Chasseur* (112.123 + 1,300 livres tournois).

drastically increases accessibility: the archive can be consulted *anywhere* and *at any time.* Tools to advance the searchability of the data can be applied next. Typical examples include indexes, transcriptions generated through optimal character recognition (OCR) or handwritten text recognition (HTR; and, the next step, natural language processing) or open access databases.[13]

Despite impressive large-scale initiatives to render accessible the British Prize Papers, no similar initiatives have been taken to unlock the prolific privateering collections in French archives, including the *dossiers de prise* of St Malo. Therefore, their use in research has been limited compared to those archives that have been already subjected to existing digitisation projects. One of the most ambitious attempts to gather and disseminate the data contained in these *dossiers* has been (and is being) made by an independent researcher from the Netherlands. Over the years, he has worked his way through the *dossiers de prise*, photographing the *dossiers* and processing the data in Excel (for metadata on the archival identification numbers, crews, freight and so on) and Word (for transcriptions of the integral text of the *dossiers* in question). In all, he has surveyed a specific subset of the total: 176 *dossiers* pertaining to the capture of Dutch ships during the Nine Years' War. This accounts for a few thousand pages of archival material. He has disseminated his findings in a variety of scientific and popular publications and presentations.[14] This demonstrates the value of citizen science to maritime history.

Recognising that the raw data would be of value to other researchers as well – and thus the importance of making it digitally available for open access – the independent researcher contacted us at the Institute for History at Leiden University to combine resources/networks and help brainstorm ways to bring the raw data as efficiently as possible to academic and independent researchers alike. We enthusiastically received this proposal, and this was the beginning of *Global Privateering*, a project to set up a user-friendly open access database of Dutch ships captured by St Malo privateers, as described in the *dossiers de prise.*

13. Optimal character recognition is generally used to generate transcripts of printed letters; handwritten text recognition is generally used to generate transcripts of handwritten scripts; and natural language processing involves training a computer to 'understand' text and recognise entities such as names, places or dates from the context or structural patterns in the composition of texts.

14. See, among others, Siem van Eeeten, 'Franse kaapvaart: Les corsaires', *Scheepshistorie*, 27 (2020), 50–9; Siem van Eeten, 'De kapersdossiers in Franse archiven', *Tijdschrift voor Zeegeschiedenis*, 39, No. 2 (2020), 75–8; Siem van Eeten, 'Twee Vlielanders in Franse archieven', *Vlielandmagazine*, 32, No. 2 (2021), 34–7; Siem van Eeten, 'De kaping van de *Genoegens van de ongetrouwde Jongeman* door de *Petit Joseph* van de reder de Heer van Lesnaudière, 4 november 1688', https://www.academia.edu/34782484/De_Kaping_van_de_Genoegens_.doc (accessed 15 July 2022); Cátia Antunes and Siem van Eeeten, 'The Saint-Malo Prize Papers: Looking at the Future of Prize Papers in a European Context', 15 November 2021, https://maritiemportal.nl/wp-content/uploads/2021/11/PrizePapers_LunchTalks_Programme_Winter21.pdf (accessed 15 July 2022).

Assembling a team

From the start, it was clear that there were insufficient in-house technical skills to construct an open access database, and that those skills had to be brought in from the outside. A few potential sources were considered but eventually we settled on recruiting IT students from other educational establishments in the Netherlands. Partnering with students (as opposed to working professionals) was an attractive option for two main reasons. First, at the time when we first conceived of the project, internship/work experience opportunities for students were heavily affected and diminished by COVID-related lockdowns. A project such as *Global Privateering* could be conducted (completely) remotely and would offer students a chance to gain professional experience at a time when opportunities were rapidly drying up. Second, we were aware of (and excited by) the prospect of cooperating with other educational establishments beyond the university system. The benefits were more than evident. Beyond the practical benefits (the compatibility of theoretical and technical expertise), the cooperation was promising in societal terms. In recent years, significant advances have been made in collaborations between different educational levels and institutions in the Netherlands.[15] Particularly in faculties of medicine or natural sciences, internships and work experience placements for non-university students are available and institutionally accredited (*erkend leerbedrijf*) as such.[16] Think, for example, of laboratories providing training for (assistant) laboratory technicians. This facilitates the mingling of different types of students, educators and resources that would not as often meet in other settings. However, in the humanities, there are currently markedly fewer of these collaborations, but the societal benefits can be presumed to be equally great.

As there are various levels of IT education on offer in the Netherlands, we first informed ourselves of the general level of technical expertise that the construction of the *Global Privateering* database (with the specific functionality we desired) would require – something as mundane as a fuzzy search option can quickly elevate the required programming skills to a different level, which are only taught at certain levels of schooling. After secondary (high) school, tertiary education in the Netherlands is roughly divided into vocational education (MBO – practically oriented schooling aimed at

15. Examples include the SamenWIJS programme at the University of Groningen, https://www.rug.nl/society-business/science-shops/nieuws/samenwerken-met-mbo-en-hbo-studenten-dat-kan (accessed 15 July 2022); the 'programma Kennisstad' in Leiden – see Cor van der Velden, 'Leiden profiteert van "programma Kennisstad", maar samenwerking met mbo kan beter', *Sleutelstad*, 15 December 2021, https://sleutelstad.nl/2021/12/15/leiden-profiteert-van-programma-kennisstad-maar-samenwerking-met-mbo-kan-beter (accessed 15 July 2022); and the Utrecht University report by Willemijn Schot, Itzél Zuiker and Claudy Oomen, 'Onderzoekssamenwerking tussen PO-, VO- en MBO-scholen met instellingen voor hoger onderwijs: In het kader van de pilot Werkplaatsen Onderwijsonderzoek' (2019), https://www.nro.nl/sites/nro/files/migrate/onderzoekssamenwerking-tussen-po-vo-en-mbo-scholen-met-instellingen-voor-ho.pdf (accessed 15 July 2022).

16. Overview of *erkende leerbedrijven* on https://stagemarkt.nl. Querying 'faculteit' demonstrates the prominence of the (natural) sciences and medicine faculties among them; see https://tinyurl.com/mr22v772 (accessed 15 July 2022).

mastering a specific skilled trade), universities of applied sciences (*hoger beroepsonderwijs* or HBO – higher professional education mixing practical and theoretical elements) and research universities.[17] Both MBO and HBO offer IT courses. We understood that the most complex functions – particularly backend elements – we were envisioning for our database would likely require HBO-level training, and that perhaps a mixed team of MBO and HBO students would best suit our purposes.[18]

We entered into a partnership with Nova College in Beverwijk, an MBO institution with an IT department.[19] We were very fortunate to connect with four talented students who were in need of additional challenges beyond the standard curriculum. With their elevated skills and the continued didactic and technical support of their own teacher to tackle even more complex builds, the need also to recruit HBO students fell through, and we were able to construct the database with an MBO-only team of IT students. In the end, the core team consisted of two members and a recent graduate from Leiden University's Institute for History, four MBO IT students from Nova College, and the independent researcher who gathered the raw data. Important support, furthermore, was provided by the students' supervisor/IT instructor at Nova College, and several Bachelor of Arts and Master of Arts students and other institute instructors, who served as a focus group for the developing database. The entire team is thus a mix of professionals and students of various educational levels, orientations and establishments, with elements of both institutional and citizen science.

The process

The actual construction of the database took place over a time span of little under a year. Our first baseline meeting with the IT students and their instructor (hereafter, the developers) took place in January 2022, setting the autumn of that year as the target launch date. During this meeting, we outlined our wishes for the database's functionality, accessibility and hosting, and informed ourselves of the possibilities. The developers subsequently analysed and categorised our wishes using the MoSCoW method (see Table 1), which is commonly used in software engineering to gain insight into the priorities contained in the wishes of the client (in this case, us).

Early on, there were two themes that emerged as points of contention. These required some significant (but interesting) consideration and the weighing up of the clash between

17. Descriptions of MBO, HBO and university education can be found on the Dutch government website, https://www.government.nl/topics/secondary-vocational-education-mbo-and-tertiary-higher-education/secondary-vocational-education-mbo and https://www.government.nl/topics/secondary-vocational-education-mbo-and-tertiary-higher-education/tertiary-higher-education (accessed 14 September 2022).
18. Simply put, backend software engineering is making sure that a website or application (for example) *works* in the way it is supposed to work. It encompasses the development of the structure of an application. Backend engineering may be contrasted with frontend engineering, which encompasses presentation – for example, how a website or application *looks.*
19. See Nova College's IT department website, https://www.novacollege.nl/opleidingen/allerichtingen/ict/ (accessed 15 July 2022).

Table 1. The MoSCoW method applied to *Global Privateering* (drafted by the developers and edited and adjusted for the final product by the author).

Acronym	Purpose	For *Global Privateering*
M (must have)	Elements that must be present in the final product; the product is not usable without these elements	A database containing the raw data A platform to access the raw data Event-based (capture incident) searchability A user-friendly way to add miscellaneous data to the database later on
S (should have)	Elements that are highly desired by the client; however, strictly speaking, the product is usable without them	(Links to) photographs of the archival material A glossary A page with information on the project, historical background and colophon Standardised spelling of the archival data Fuzzy search A category/place to add miscellaneous pieces of data that fall outside of the main data structure Full-text searchability Category-based searchability A modern user experience A home page with an overview of all data
C (could have)	Elements that it would be 'nice' to have but are not absolutely necessary	Ability to query the database in several languages for common/standardisable terms (for example, querying 'sinaasappel' will also render results for 'orange') Visualisation of location data (harbours, captures) from the data set on an interactive map
W (won't have)	Elements that will not be present in the final product but could be of interest if later add-ons are pursued	Other privateering-related data sets User interface in languages other than English Switch to graph-based (instead of relational) structuring in view of compatibility and cross-referencing with other privateering-related data sets

historical methodology – approaching and processing sources – and technical user-friendliness.

The first theme was language. Our raw data – the privateering *dossiers* of St Malo – is mostly in French. However, its subject matter is mainly Dutch – the captured ships, the crews and so on. The language of international scientific communication (particularly digital tools) is chiefly English. It was important to first distinguish between the language of, on the one hand, the *website* and its user interface and, on the other, that of the *data* it would present. After toying briefly with the idea of making the website available in

French, Dutch and English, we settled on just a single version in English, in line with the notion that *Global Privateering* at this point is still very much a pilot project. The question of what language to present the data in was much more difficult, as it touched on some methodological aspects of historical source work. Here is a practical example: under the heading 'ship name' (in English, as the interface language), would it say *Le Chasseur* (what is literally written in the French-language document) or *De Jager* (its 'actual' name in Dutch)?[20] As a general rule, preserving the original source text and thus language is optimal, as it prevents any information from getting (literally and figuratively) lost in translation. Another problem with translations, especially of names of persons or ships, is that one can rarely be entirely sure about a translation – does *La Dame Anne*, as written by the French administrator, refer to *De Juffrouw Anna*, *De Vrouw Anna* or any one of their possible spelling variants?[21] How were we to deal with spelling variations in general, as early modern spelling was not by any means standardised? Again, by keeping to the source text, the loss of information is minimised and it familiarises the user with the spelling variations on offer. In theory, the data presented in the *Global Privateering* database should thus be entirely in French and directly taken from the literal source text. However, practical arguments against this proved more convincing and made us settle on translating and standardising to English – at least for the time being and within reasonable limits. First, keeping everything in French would arguably diminish the ready accessibility of the data, as we factored in the challenging nature of some French privateering jargon for our target audience (students and casual researchers). Second, the vast majority of the data *could* in fact be translated into English in an arguably unproblematic fashion with minimal loss of information. Whenever this was not the case, we left the term untranslated and included it in a glossary (available on the website). Third, as we received clearance for the actual photographs of the documents to be published alongside the relevant data on the website, a user who is interested in tracing the original phrasing could readily consult the attached photograph. Lastly, the St Malo data set was small enough to standardise by hand the spellings of reoccurring pieces of data (for example, reoccurring harbours) and then, if applicable, translate them. The loss of visible spelling variations was judged to be of less importance than efficient searchability and cross-referencing possibilities (one simple search for 'Lisbon' instead of Lisbonne, Lixbon, Lisboa, Lixboa or Lissabon, for example). Overall, while settling on English due to the relatively straightforward translation and standardisation options, we will reconsider this policy if *Global Privateering* is expanded with other, particularly larger and/or non-western, data sets on privateering.

The second consideration was how to structure and subsequently present the data to the user. Having settled on presenting the user with an information sheet containing all the main pieces of data, we needed to identify *what* exactly was the 'chief' type of metadata contained in the data set and which pieces of data were in essence sub-characteristics of this 'chief' bit of data. As privateering centres around *ships*, without much thought we first centralised the captured ship, categorising all other data points – such as the crew members, the privateer who took them, the cargoes and so on – as 'belonging' to the

20. AD 35, Série 9B, 600-1, Dossier 1.
21. AD 35, Série 9B, 604-2, Dossier 10.

captured ship. The website would then present the user with an information sheet on the captured ship. However, this proved to be somewhat asymmetrical. For example, it did not feel right to delegate the privateer ships (for which the data typically provided numerous subcategorisations of its own) to a 'second-rate' data type compared to the captured ships. We made a breakthrough on realising that the data (and thus the information sheets) should be *event-based* as opposed to *ship-based*: instead of information sheets on the *captured ships*, we were to provide information sheets on *the capture of a ship.* In this way, the captured ship and the privateer would be on an equal level of importance, both involved in the particular event on the information sheet. This would also correlate more closely with the structure of the archive itself, as the data comprises the administrative process files of a capture event – one file per capture, one sheet with the file's information summarised. Event-based databases are also most suited for graph structuring, and it will thus facilitate *Global Privateering*'s possible transition into a graph database if this is judged to be productive in the (near) future.

The structuring and standardisation of the information contained in the privateering *dossiers* was one of the primary duties assigned to us as researchers throughout the project. In terms of database development, this is called normalisation. First, we needed to identify *which* bit of information was subservient to another bit of information. Preferably, one descends through 'levels' of information until the lowest level is reached and no more subdivisions can be made. For the developers, it was important to have this structure in hand in order to programme the database and ensure that it captured/understood the underlying hierarchy. As identifying the structure of the (archival) data requires expertise in terms of the content, this fell to us. Figures 1 and 2 provide examples of the structuring of the data contained in the St Malo privateering *dossiers* .

Figure 1 shows the 'cargo' branch in the structuring of the information in the privateering *dossiers*, reached after 'capture of ship' (main data point) → 'captured ship'. In database normalisation, it is preferable to subdivide the data as much as possible into standardisable data; broadly speaking, the piece of data can be selected from a relatively concise list. Commonly encountered unit types, for example, are not limitless and can thus be considered standardisable. This standardisation is preferable because identical pieces of data (for example, 'barrel' as a unit type – a piece of data that is bound to reoccur) only have to be registered once; it thus prevents the inefficient use of digital data storage space. However, with 'destined for', the prompt is an individual('s) name. This is not standardisable because it theoretically encompasses endless possibilities. The subdivision ends there.

Figure 2 shows the 'provenance' branch in the structuring of the information in the privateering *dossiers*, which is secondary only to the main 'capture of ship' data point. This branch is very straightforward: each bit of data has only one applicable subdivision. The result is the full *cote* or reference/footnote to the information in the archive. This reference will be provided on each information sheet.

We transmitted our findings on the structuring and hierarchy contained in the data to the developers, who built a backend and frontend structure to match it. This structure was constructed such that Excel files could be exported/transferred into it. The next step on our part was thus to construct an Excel sheet with columns for every single type of data (for example, 'archive', 'collection', 'dossier') and put each bit of metadata under the correct

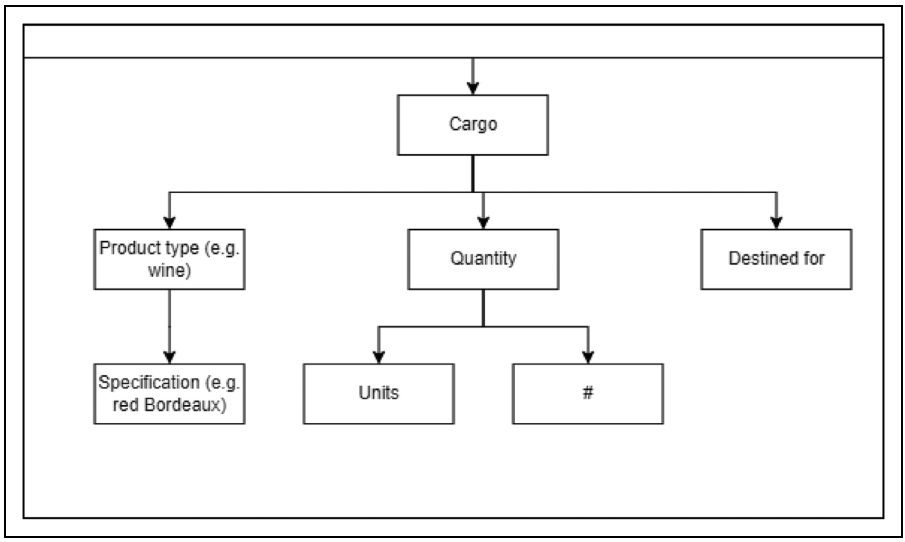

Figure 1. The 'cargo' branch in the structuring of information in the privateering *dossiers*.

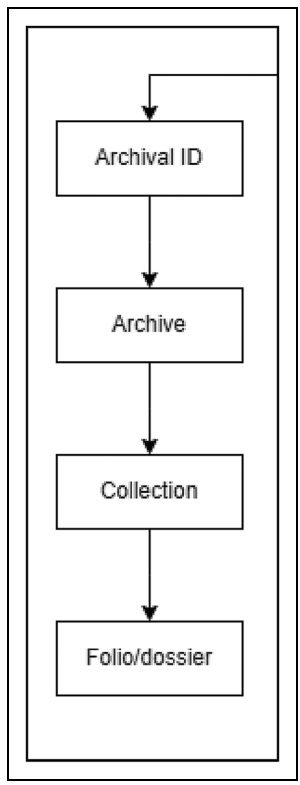

Figure 2. The 'provenance' branch in the structuring of information in the privateering *dossiers*.

denominator. The raw data gathered by the independent researcher was already (largely) in Excel, so we tweaked the existing sheets to fit our structure better and moved or subdivided the information if necessary. The independent researcher took it upon himself to allocate the data to fit the slightly altered structure and to further complete the data where it was still unfinished by going back to the original photographs. To make the Excel file ready for export all in one go, we subsequently went through the data to standardise (for example, spelling variations) and translate it (Figure 3).

As illustrated in Figure 3, using the data structuring that we provided, the developers subsequently translated it into code. Different types of data (numbers, text and so on) require different types of programming variables. The privateering data consists of different types of data and thus requires different types of variables. Here, 'int' refers to 'integer' or a whole number. Closely related to 'int' is 'double', meaning a (64-bit) decimal number. 'Text' or 'string' consists of letter characters and is usually employed for purely text-based information. The developers needed to determine which piece of data fit into which variable and programme accordingly. For example, the ship *La Dame Marie* (text/string) was carrying 80 (int) barrels (text/string) of flour (text/string), accounting for 27.8% (double) of the freight.

Throughout the process, we held a short meeting with the developers every two weeks with the purpose of providing updates on each other's progress and (mainly from the developers' side) asking questions that had arisen during the database design and construction. As the previous section demonstrates, a baseline understanding of the *content* was necessary to design and programme accordingly, as the right elements needed to be connected and put into suitable variables. For example, when the developers were designing the section on auctions, it was unclear to them whether, in early modern privateering, it was customary for the captured freight to be auctioned off as one lot or several lots, and whether there would be multiple buyers or just one buyer. This was important information to have: did the database need to account for just a single 'buyer' category or did it need to be able to accommodate multiple buyers? We

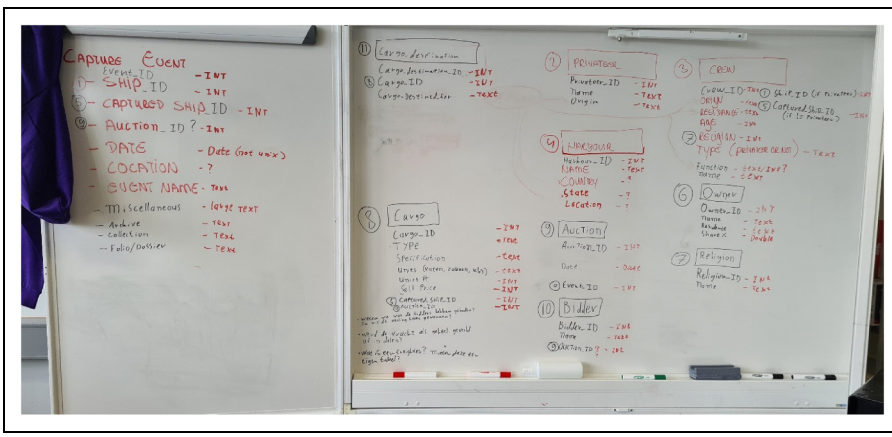

Figure 3. The developers' drawing board at Nova College.

subsequently outlined the relevant historical background for the developers to take into account. Such occasions demonstrate that projects such as this greatly benefit from the continued presence *and* active participation of both content-oriented and technically oriented experts.

Another important meeting was held with a group of Bachelor of Arts and Master of Arts students enrolled in the university's history programme. It was important for us to establish who would be the main audience types for the database: researchers at different stages and from different backgrounds are likely to ask different questions and will thus be interested in different search functions or highlighted data. For our pilot, we determined that our projected user demographic would be students of maritime history and independent researchers, trusting that more advanced researchers could use the relatively simple presentation as a stepping stone to independent enquiry into the accompanying photographic material. During the meeting, we introduced the *dossiers de prise* as a source type, and subsequently asked the students what types of research they would use them for. This focus group provided us with valuable feedback regarding the types of data we should highlight and which search functions should be included in order to help our projected user demographic to answer their questions.

It took a while before there was a 'presentable' product for us to survey, as the backend engineering preceded most of the frontend (for example, the look/presentation of the website) construction. However, a limited proficiency in common programming languages was very helpful for our understanding of the basic characteristics of the backend structure, and to gauge what it could and could not (yet) do. This facilitated the mutual intelligibility of the dialogue between us and the developers. In addition, an increased understanding of the technical limits is, as we will see, one of the key benefits for the historian.

The new world: benefits for the historian

Overall, *Global Privateering* was a very positive and productive collaborative project, and would encourage us to pursue similar partnerships between historians and junior IT developers in the future. One of the particular strengths of this collaboration was the strong mutual benefit: both us as researchers and the developers derived significant and varying benefits from participating. For our part, we have experienced three particular benefits, which we feel are generalisable to similar collaborations beyond this particular project.

First, and perhaps most obviously, it has provided us with a concrete product/tool that significantly aids research into (maritime) archival data. The primary goal from the inception of the project was to transform the raw archival data (captured in Word or Excel) into a user-friendly database that (a) increased the accessibility of the data for searching and analysis and (b) made the data publicly available to all researchers. Increased accessibility and open access availability greatly enable historical research on (large) data sets, to the benefit of the profession at large. Our collaboration with the developers turned this from a vague desire into a reality for the St Malo *dossiers*, which will now be readily consultable for a wide audience, hopefully contributing to further research into (early modern) maritime history.

Second, the close collaboration with the developers, whose background and expertise were very different from our own, was very informative. Continued interaction and dialogue with team members from a different discipline (IT) familiarised us with the basic processes of database design and development (for example, the MoSCoW method), data normalisation and website hosting possibilities. It especially facilitated our understanding of the possibilities and limits with databases such as this. The knowledge of basic backend functionality is broadly applicable and useful when utilising other, similar databases, and can, for example, be used to query more efficiently. In all, it demystified the 'digital humanities' and allowed us a detailed backstage look at how its most useful tools come into being.

Third, we enjoyed being part of a team that extended beyond the university. It demonstrated to us the compatibility of talent/expertise found at universities, on the one hand, and other educational establishments such as MBOs, on the other. Physically welcoming the MBO party on campus was an essential step in bringing down these barriers; of equal importance was our visit to Nova College and visiting their workspace, to fully appreciate and understand what they would bring to the table. Throughout the collaboration, the dialogue was often very interesting; although it was mostly mutually intelligible, it occasionally required skill in communicating certain aspects of the historical craft, such as common methodological approaches in archival research or current debates in early modern maritime history – both of which were relevant to the eventual presentation of the database. The clear and concise conveyance of historical knowledge to an audience beyond the university and outside of the historical discipline is an essential skill to possess and maintain, and this dialogue tested us once more.

Learning the craft: benefits for the developers

On the part of the developers, there were again three main benefits to be had from their participation in the *Global Privateering* team, particularly in aiding their development into fully fledged IT professionals.

First, it provided them with valuable and highly transferable work experience. Part of the Nova College curriculum is working on 'projects' – fictional or real – to familiarise students with common types of assignments that they might encounter later as a working IT professional. The *Global Privateering* project presented the students with a realistic and relatively concise assignment that entailed all aspects that are commonly encountered in a typical commission: creating a product from scratch featuring backend and frontend development and, most importantly, communicating with a client who likely has limited understanding of IT concepts and processes. In our collaboration, the latter element was particularly vital. Throughout the project, we posed as a 'client' and left it up to the developers to translate our relatively ineloquent wishes for the database into workable code: 'We want to be able to sort stuff like you can do on a clothing website' became a fully functional menu where one can filter by harbours, cargo types, privateers or combinations thereof. In addition to practising these communication skills, the construction of the database allowed the students to practise their programming abilities: designing data structures; researching, designing, writing and debugging code; and creating a modern user interface for the website. Their overall work experience gained within the *Global*

Privateering project will be institutionally recognised as European Credit Transfer and Accumulation System (ECTS) credits.

Second, *Global Privateering* did not only provide the developers with 'real' work experience; it also resulted in a real-life product and the opportunity to work with actual data. This is in contrast to other projects that are on offer in the Nova College curriculum, which are often fictional, with fictitious clients, fictitious data and a product that is not intended for actual use. Typical assignments include inventories of shops or customer databases. The developers expressed their enthusiasm for working with a real data set, and particularly one with historical data. They enjoyed working with the St Malo *dossiers*, learning about early modern privateering and the historiographical questions that surround it, and applying their skills to this data. The fact that the resulting database would be live and utilised by researchers was motivational, satisfying and rewarding for the developers, who, in addition, could subsequently add it to their professional portfolios to highlight their skills for future clients.

Lastly, the project offered our particular developers a chance to develop their skills beyond what is normally required within the MBO curriculum. At the beginning of the project, we established that HBO-level training would likely be required to execute some of the elements, thus exceeding MBO-taught skills. As the developers were in need of additional challenging, *Global Privateering* offered them precisely this. They reported that they learned a significant amount throughout the project, and were able to successfully execute things near the end of the project that they had no understanding of at the beginning. It is imperative to acknowledge the developers' instructor for his continuous supervision and dedicated support, demonstrating complex builds to them and helping them solve problems whenever necessary. Together, we are glad to have been able to offer them the opportunity to further elevate their skills and continue learning beyond their expectations.

An example of the developers' work is shown in Figure 4, where we can see a segment of code providing *Global Privateering*'s fuzzy search function through the Levenshtein method. If fuzzy search is turned on, searching for 'Lixboa' will also give the results for 'Lisboa' because the words/spellings are within a set degree of similarity, as stipulated in this code. In dealing with early modern data, such functionality is desirable to account for the non-standardised nature of spelling at the time.

Concluding remarks

Global Privateering launched in the autumn of 2022 in accordance with the timeline we envisioned when beginning the project . Figure 5 shows *Global Privateering*'s user interface (in beta), where the user is presented with an overview of the captures and their basic metadata, as well as filters to narrow down the results. When clicking on one of the captures, the user is presented with a detailed information sheet (Figure 6; in beta) with all of the available metadata from the *dossiers*, such as the captured ship, the privateer, the archival identification number and links to the original scan.

Intended from the start as a pilot, our positive evaluation of the product's utility is encouraging us to explore the possibility of including more data sets on our website, to facilitate comparative research into privateering-related archival matter on a global

```
function levenshteinFilter($results, $search_string, $column_names)
{
    $succesful_results = [];
    $search_string = explode( separator: " ", $search_string);
    foreach ($results as $result) {
        $score = 0;
        //every table has different colums
        foreach ($column_names as $column_name) {
            if (isset($result[$column_name]))
                $score = $score + fieldFilter(explode( separator: " ", $result[$column_name]), $search_string);
        }

        if ($score > 0) {
            $result['score'] = $score;
            $succesful_results[] = $result;
        }
    }

    //sort all elements according to score. a higher score means a more relevant result.
    usort( &array: $succesful_results, function ($first, $second) {
        // use spaceship operator because PHP 8 usort gave a deprecated warning
        // it sorted in reverse, so I switched the variables
        return $second['score'] <=> $first['score'];
    });

    return $succesful_results;
}
```

Figure 4. Segment of code providing *Global Privateering*'s fuzzy search function through the Levenshtein method.

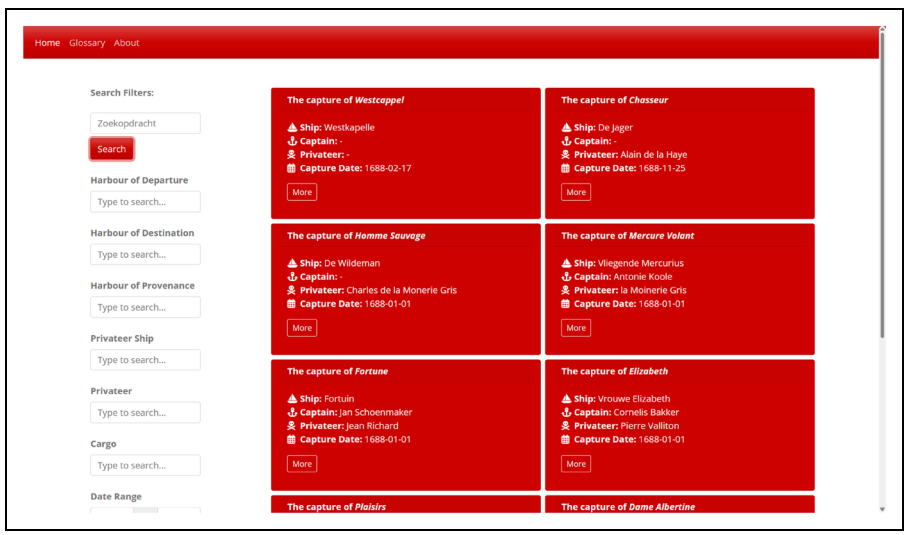

Figure 5. *Global Privateering*'s user interface (in beta).

scale, to increase the amount of data available for online open access, and to combat fragmentation of similar data sets.

The timely launch of the platform is a testament to the productive cooperation within the team, where the diversity of educational backgrounds and disciplines proved to be a

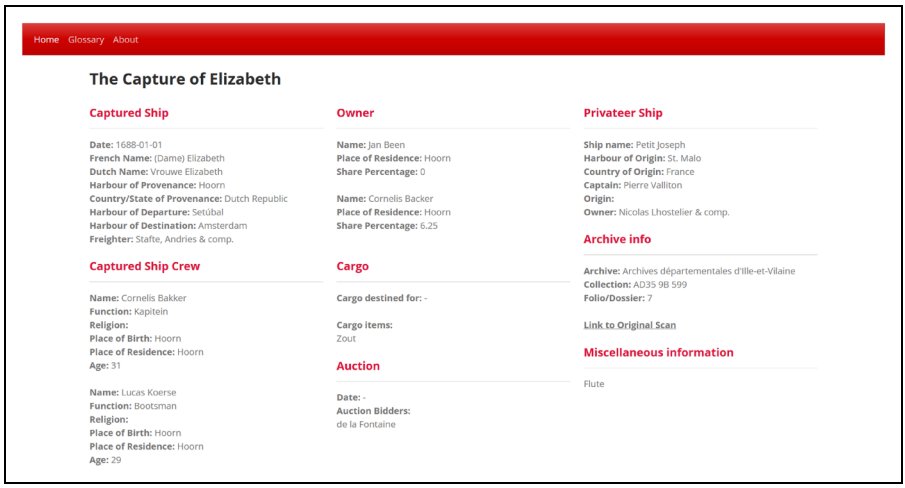

Figure 6. One of *Global Privateering*'s detailed information sheets (in beta).

great asset. The specific collaboration between university-affiliated researchers and MBO technicians is therefore not only fruitful in the natural sciences, but also possible in the humanities. The practical, professional, didactic and societal benefits have been considerable *and* mutual, and therefore deserve further exploration. Many Excel sheets, Word documents or notebooks with treasure troves of archival data are still waiting for their turn in the (digital) limelight and, as *Global Privateering* evidences, there is an abundance of talent out there to turn this into a reality.

Author biography

Tessa de Boer is a PhD candidate in early modern history at Leiden University, the Netherlands. She is one of the co-instigators of the *Global Privateering* project.

Declaration of conflicting interests

The author declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author received no financial support for the research, authorship and/or publication of this article.

ORCID iD

Tessa de Boer https://orcid.org/0000-0002-7359-9945