

The efoto-project: Narrative construction of the past and semi-automated data curation

Mareike Katharina Schumacher, University of Hamburg

***Abstract:** The efoto project is a cooperation between the Cultural Ministry of Hamburg, the University of Hamburg and numerous archives, museums and libraries of the city. Its overall aim is to create a large image database of the city and make it accessible to the public. Access is provided via a mobile application which has been developed as part of the project. In the content narratives are seen as the central link between image data and users, and the images in the database can be enriched with narrative elements at different levels. Partner institutions can create city walks with them and users can comment on them or record audio notes with their personal narratives in addition to the images.*

In this article we want to show how scientific insights and workflows are used to create or initiate narratives that supplement the historical photographs. In the first place images and metadata are provided digitally. Narratives in the form of comments and audio notices will be added by users once the application goes public. During the efoto-project it became apparent that in our partner institutions, there are no ready-made city walks already available for mobile application. As city walks are time-consuming to create or to convert into the right format for a mobile application, we developed the model of semi-automated digital data curation presented in this article.

In our attempt to develop the model we used strategies from narrative and cultural theory, digital humanities and historiography. This article outlines the process of semi-automated data curation through a case study about the German pirate Claus Störtebeker.

Keywords: Digital Humanities, Data Curation, Cultural Heritage, Narrative Theory, Störtebeker, Text Mining

1. Conceptual foundation of efoto-Hamburg

In efoto at least three different interests coincide. On the one side there is a political perspective from which it is most important to realize certain strategies and guidelines. The most important political signposts are the ‚eCulture Agenda 2020‘ (Freie und Hansestadt Hamburg) and the European inspire guideline (European Union 1995-2017). Transparency and accessibility of electronically stored data are two of the ideals which efoto wants to meet. On the other side there is a scientific interest in analyzing how users react to and interact with the provided data. A third field of interest is a pragmatic approach to cultural heritage. Institutions like museums and archives want to follow their educational mandate by bringing culturally relevant data to users, who are not usually their

typical target audience. From all three perspectives there is an additional interest in digital data.

Taking this network of interests as our starting point, we needed to develop a pragmatic conceptual foundation. This would help us to develop technical features for our mobile application and provide us with a methodology for analyzing and processing data, to increase the probability of user interaction.

The three most important factors inside the application are the representation of Hamburg as a city on a map, the image data and the users. We have gathered about 45,000 images from five partner institutions, and located more than 6,000 images on a map. We assume that people will first look for images showing places where they are or have already been. But we do not want to stop here. In order to avoid creating a one-way communication system through which cultural institutions, museums and archives feed their audience, we seek a dynamic process in which users decide in social interaction whether a picture is of cultural value or not.

Culture

Our understanding of culture has to include at least the three main factors that determine the efoto application. The basic assumption is that “culture” is a counterpart to the term “nature”, which includes everything human-made and the process of making (Müller-Funk 2006: 8; Eagleton 2001: 7f; Ahrendt 2010: 16): this is part of the foundation of our conceptual framework. This broad perspective on culture is very object-bound (although it accepts that the making of something is also a cultural act, yet there nevertheless has to be some kind of outcome), and can be closely connected to the image data, which is the starting point of efoto. But only using this conception of culture would not bring us any further, because it does not give us enough details on the role of people. In the first place, people have to cultivate nature for their living (Ahrendt 2010: 16), thus forming “civilization” (Eagleton 2001: 19ff). At the same time there seems to be the necessity to cultivate the self (Eagleton 2001: 13). Finally, culture also means regulation in social interaction or, as humans always live in a political plurality (Ahrendt 2010: 17), there seems to be an urge for cultivation in groups or even in whole societies. Culture can also be understood as a system of socially approved behaviour, of practices and rituals (Müller-Funk 2006: 8) and that whole nations, countries or governments can also be instances of cultivation (Eagleton 2001: 14). These three basic functionalities of culture coincide with our set of variables within efoto; the city of Hamburg as the cultivation of a living environment; the self, which is cultivated inside this environment; and the group which forms itself on the mobile application platform, where a process of cultivation will be initiated.

There are many other notions of culture. Some of these are discussed here to explain why they are less important for efoto and thus not included in our conceptual framework. The term culture can refer to an elitist system of art (Müller-Funk 2006: 8; Eagleton 2001: 33). Some of the pictures in efoto might meet this claim, but most of them were not initially intended to be artistic. They were made for various purposes, from documenting urban development to illustrating political statements. As we argue that they can all be culturally relevant today, it is clear that the efoto project is not confined to an elitist idea of culture.

We stated above that we see one functionality of culture as the cultivation of the self. However this does not mean that we understand culture as a means to distinguish between those who are “cultivated” (groups, classes or whole societies) and those who are not (Eagleton mentions this understanding of culture e.g. in 2001: 39).

Neither do we focus on cultural phenomena, which could be examined too widely at macro level, nor do we use theories, which privilege the cultural sentiments of individuals too much. This means that we set aside the fundamental idea of cultural pessimism, in which cultures follow a certain scheme of rise and fall (Müller-Funk 2006: 77; Müller-Funk 2006: 87).

We also exclude the psychoanalytical aspects of culture (Müller-Funk 2006: 22ff) from this study, because we do not focus on the problems that the urge for cultivation we named above may cause for individuals. In addition to that we exclude economic aspects of culture as elaborated by representatives of critical theory (Adorno, Horkheimer 2003; Benjamin 1963).

Our basic understanding of culture in the sense of cultivation of the environment, self and other people in a group mostly answers the questions of why there is culture and how culture creates itself in the first place. In addition to that, we need to know more about cultural heritage.

The term cultural heritage is closely bound to UNESCO and their aim to preserve structures and practices of cultural value (Vecco 2010, Dormaels 2013). Beginning with a very normative understanding of cultural heritage (Vecco 2010: 322), it becomes clear that cultural heritage objects can be made valuable through social interaction (Vecco 2010: 323). The interpretation of cultural heritage followed a similar development. From monological traditions it changed to a more discursive process in which personal, collective, regional and national identities play an important role (Silbermann: 2017). Objects can be included in a process of ‘heritization’ and thus become a source of cultural knowledge (Dormaels 2013, 108). The distinctions between tangible and intangible cultural heritage were founded on ideas of a dynamic process of cultural value development (Vecco 2010: 323). So from a very object-bound view of cultural heritage objects became considered as triggers for cultural actions, and

cultural heritage itself became accepted as existing without materiality. Nevertheless the separation of tangible and intangible cultural heritage in a strict either/or distinction can exclude some of the insights on the dynamics of culture outlined above. The cultivation of environment, the self and social surroundings is not fully captured, for example. Additionally, including aspects of time using this either/or distinction seems to be problematic. From our point of view the linguistic paradox of ‚Heritization‘, which Dormaels describes (Dormaels 2013: 112), is exactly what cultural heritage is about: something from the past is included in an actual current and potentially ongoing cultural process. This “something” may be material as well as immaterial or a combination of both. In any case cultural relevance is not seen as a finite intrinsic quality.

The assumption that culture is something ongoing which does not and cannot have an ending, any kind of fulfilment or constant value meets Luhmann’s ideas on culture. A cultural artefact is constructed as such by two levels of reflection. A reflection on an object ascribes a certain value to it, but only if there is reflection on a second level is a cultural process initiated, and the object becomes a cultural artefact (Luhmann 1995 and 1997). In this understanding culture is radically bound to a semantic element (reflection) and a specific moment in time. Only objects which are part of the second level of reflection are of cultural value (at the moment of reflection) which means that all other objects may be forgotten (at least until somebody starts reflecting on it again).

Other theorists see the power of culture in its discursive quality. This might become a negative connotation as in Foucault’s approach to culture as a discursive system of inclusion and exclusion (Seitter 2004: 166; Müller-Funk 2006: 184ff). A more neutral understanding of cultural dynamics between the actions of subjects and the interpretation of those acts as cultural is provided by Geertz with his understanding of *thick description*. As an anthropologist Geertz has described how thick description is used in everyday life to grasp cultural meaning in human interaction and how it is central for the interpretation of culture, as in ethnological work (Geertz, 1973). Geertz even goes as far as seeing every element of culture as a potential text which can or even must be interpreted in order to understand it (Geertz, 1973: 448).

A recent line of research in the theorization of culture highlights narration as absolutely central to culture (Fisher 1984; Müller-Funk 2008; Nünning 2014). This tradition specifies Geertz’s extended concept of text and states that culture is constructed narratively (Müller-Funk 2008). This does not mean that the importance of objects is denied. As Nünning points out, the underlying semiotic term of culture has three dimensions: material, mental and social (Nünning 2014: 28). All three dimensions can have different forms. The material dimension can include texts (Nünning 2014: 28), it can also mean artefacts or art objects (Müller-Funk 2008: 9) or the body of an acting and narrating subject (Müller-Funk 2008: 13). The mental dimension is a complex system of narrative

codes and structures which is created collectively and handed on from generation to generation (Müller-Funk 2008: 13). These narratives are of central importance for individual identity construction (Ricoeur 1991; Georgakopoulou 2007; Müller-Funk 2008: 13) as well as for collective identification (Bracker 2013). They are used for positioning the individual as well as for the creation of coherent meaning from surrounding elements and events (Bracker 2013). Narrative codes and structures are manifested in the social interaction of telling. In this understanding cultures are held together by the act of telling and are thus understood as narrative communities (Müller-Funk 2008: 14).

The images in efoto can be seen as part of the material dimension of culture. They are embedded in cultural codes which are expressed visually by the pictures they show or in textual form in detailed metadata. As Georgakopoulou (2007) has pointed out, individuals constantly work on their narrative identity by telling small stories. This is why we assume that users will also look for the possibility of constructing narrative identity while using the application. This kind of identity work can be done by aligning oneself towards specific images as well as to other users.

Collective identities can be created in small groups for example by narrating alternative stories (Bracker 2013: 5). The narrative can be seen as a dynamic link of significance between nodes (Bracker 2013: 6), in our case users and image data. Users who share certain narratives thus build a (narrative) community. On the macro-level sharing a whole reservoir of narratives is what forms a society (Müller-Funk 2008: 14).

While analysing the different understandings on culture outlined above, we noticed that they include three-dimensional models, which can all be refined to a common denominator. We started with the need for the cultivation of surroundings, for the self and for others, and went on to Luhman's ideas of a cultural system formed by objects, first-level reflections and second-level reflections. Lastly, we turned to the material, mental and social dimensions of individual and collective identity creation. So in our theoretical basis for the efoto project we also consider culture to be a process which happens on three levels. The first is the object level (which can also be seen as the cultivation of a living environment or the material dimension of culture), the second is the individual level (also the first level of reflection or the mental dimension of culture), and the third is the collective level (also the second level of reflection or the social dimension of culture):

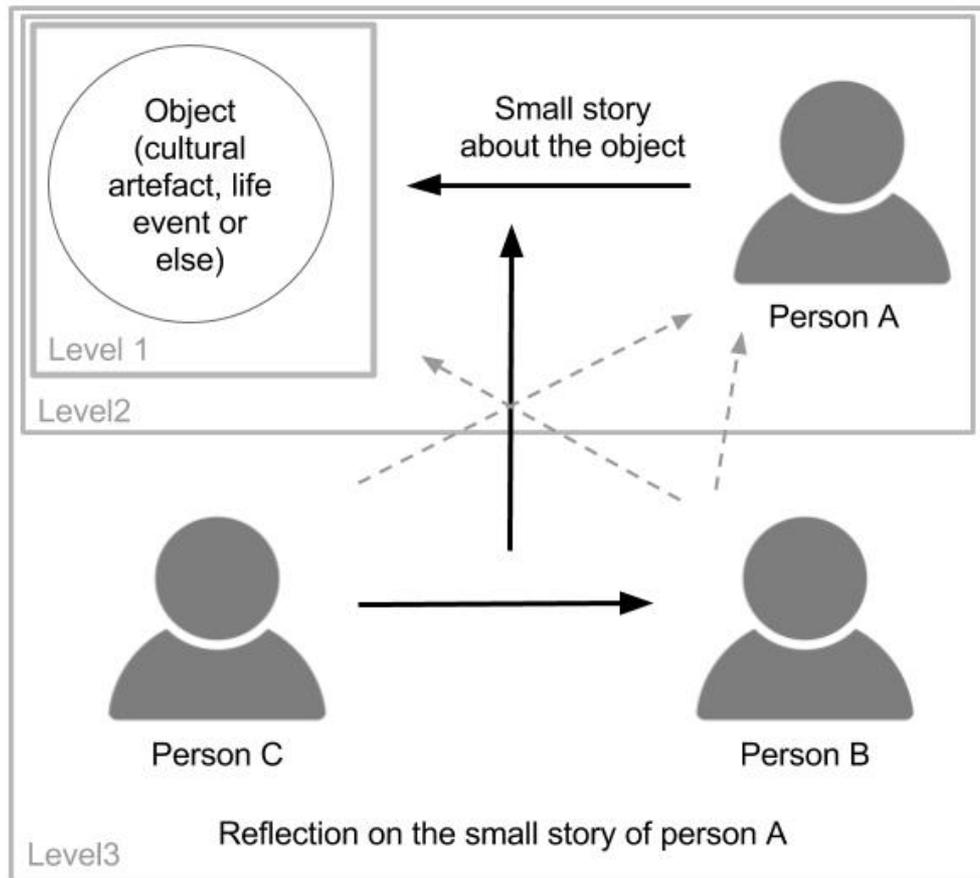


Figure 1: Model of three-levelled cultural heritage process.

As explained above, all links between levels have narrative qualities and levels can be crossed. Culture is formed when all three levels are combined. This means that objects that are not narratively bound to at least two individuals who communicate about them in a reflective, discursive or interpretive way may be discarded, because they are not bound to all three levels of the cultural heritage process. Objects which are connected to personal and interpersonal usage are automatically preserved for future generations.

2. City walks as triggers for cultural interaction

City walks are constructed by connecting different objects at the first level—images, audio files and locations—and then adding narratives at the second level. Photographs of the city are enriched with historical data and stories of historical personalities. In interviews with contemporary witnesses tonality, emotionality and linguistic aspects emerge. Using this variety of formats we try to give a starting point for the collective identification at the third level of the cultural process, thus encouraging the preservation of the data as part of the cultural heritage of Hamburg. Assuming that it is more likely that someone will

reflect at the third level if the object in question has already been reflected upon at the second level by many people, we have developed a method to integrate text mining tools in the construction of a historic city walk.

2.1 Case Study

We chose the legend of Stoertebeker as a case study, because it is an important part of the cultural narrative identity of Hamburg. However, there are no original images of the pirate and his life is still considered to be controversial in historiography (Rohmann 2007; 2012). Taking Stoertebeker's story as an exemplary case study is an attempt to supplement the images in efoto with a subject which cannot be presented in purely visual representations. In addition to narrativity this specific case implies fictional elements as well as factual incidents, and is connected to earlier history as well as to the contemporary period. Thus Stoertebeker's story includes a multiplicity of factors which increases the probability of it being reflected upon by users in a cultural heritage process.

2.1.1 Claus Stoertebeker

Claus Stoertebeker is said to have been a German pirate and freebooter in the late medieval period. As Puhle (2012) points out, there is no consensus about whether "Stoertebeker" really refers to a historical personality or rather stands for a certain type. Together with other individuals who were formerly of the same or even higher importance (Bracker 2005: 47) but today are less remembered, he was known as one of the 'victual brothers' (Puhle, 2012). This loosely bound brotherhood of privateers made their living in the Baltic Sea, fighting in wars on the side of the Hanseatic League (Rohmann 2011: 257f). Later on they were classified as pirates, be it because of political strategies or because of legal changes concerning privateering in general (Rohmann 2007: 96ff). Stoertebeker is said to have lived in the second half of the fifteenth century, and could have died on 21 October 1401 in Hamburg (Puhle 2012: 150). Several invoices in the city records, including the bill for the burial of 73 victual brothers in 1401, were analysed to find the exact date of Stoertebeker's execution (Koppmann 1877: 45). The one undeniable fact about Stoertebeker is that he is part of numerous narratives transmitted orally or in textual, audiovisual or musical adaptations and he is the object of historiographic research as well as the subject of museum exhibitions (Rohmann 2011). One of the regions he is associated with is the city of Hamburg, which makes him interesting for the efoto project.

In our case study, we want to find out which events from Klaus Stoertebeker's life are most often retold in literary adaptations and to which places they are connected. We want to investigate which events from his life are reported in journalistic texts, whether they can be located in the city of Hamburg, and if they can be linked to images in the efoto database. Finally we look at how people connect to Stoertebeker and the legends told about him.

2.1.2 Setting

To answer these questions we set up three corpora. The first corpus contains literary texts about Stoertebeker. Using this corpus we try to identify places and cultural artefacts (e.g. buildings) associated with the pirate in Hamburg. The life events that emerge from this corpus are most probably (in some cases certainly) fictional. The information found from this corpus is used in three ways: 1. the places are identified on the map of Hamburg using the efoto application, 2. names of objects and their locations are used to search the efoto database for matching images, 3. the images found in the efoto database are connected to the passages of the narrative source.

The second corpus includes journalistic texts that discuss Stoertebeker. This corpus is of a different quality than the first one, because the starting point for these texts is always factual events in the present. Nevertheless there will be fictional events mentioned in these texts as well. We analyse this corpus using the same techniques as for the first corpus, namely the extraction of places named in the texts and their connection to events.

In the third corpus we assemble texts written by authors of literary fiction that contain some autobiographical aspects or are fully autobiographical. It is a rather small corpus but of some interest to our research focus, because this genre of texts is key for exploring why and how people associate their own life stories with that of Stoertebeker.

Corpora one and three were set up using the Textgrid Repository, the German Text Archive (DTA) and the archive of project Gutenberg. In these free online textual archives literary sources from the public domain are stored in TEI, txt or html format. Using these sources meant that we had access to digital text copies of high quality, but we had to limit our study to texts from the public domain, which means that the publication date of the literary texts in our corpora stop around 1940 due to EU copyright law. To compensate for this, the second corpus was established to populate the available texts with journalistic articles from online periodicals, but excluded articles dealing with events and tourism that used Stoertebeker's story. Instead we concentrated on articles about recent studies on Stoertebeker and on a spectacular crime from 2011 when the alleged skull of Stoertebeker was stolen from the Hamburg Museum. The articles were published from 2010 to 2016.

2.1.3 Semi-automated data curation of a legend based city walk

Our model of semi-automated data curation exploits a workflow in which three Digital Humanities Tools are used to process textual data. As we do not always use the tools in the designated way and combine them rather unconventionally, we had to do some preliminary work on our data in order to optimize the results to adjust the performance of the tools on all three different corpora.

Step 1: Setting up the corpora

We used a very simple method to set up our three corpora. We searched the archives named above for Claus Stoertebeker (for all variant spellings of his name) via full text search. We excluded fictional texts that merely referred to Stoertebeker in only a few passages. The first corpus contains ten texts, the third corpus merely four. The second corpus contains twelve texts.

Step 2: Training the Named Entity Recognition Tool

The Stanford Named Entity Recognition Tool (NER) uses machine learning techniques to automatically tag entities (persons, locations, organizations) in texts. Working with NER threw up one problem concerning the setting of our case study. The classifiers provided by the Stanford University are trained on journalistic texts only and use a linguistic concept of place. Thus they arrive at a high level of correct identification of places in German (factual) texts, which lies at around 70%. While this fits our second corpus quite well, it turned out to perform very poorly on the fictional and autobiographical texts. The main reason for this seems to be that the seemingly simple category of place is used so differently in fictional texts that we needed another model in order to achieve a similarly reliable outcome. We constructed a corpus of training data which included extracts from one hundred fictional texts from the same temporal range as those in corpora one and three (dating from around 1850 to around 1950), randomly chosen from the TextGrid Repository. From each text in the training corpus, we extracted a randomly chosen passage of exactly the same length. We manually tagged all identifiable locations using the narratological model of place provided by Dennerlein and Hühn (2009). In the end we came up with an F-Score of 66-76%.¹

Step 3: Analysing the NER results in CATMA (Computer Aided Textual Markup and Analysis)

We completed the tagged texts to xml files and uploaded them into the textual markup and analysis tool CATMA in order to analyse the corpora. We ran a collocation query on the first corpus thus searching for places in proximity to the name of Stoertebeker (and its spelling variations). Afterwards we were able to look at the frequencies of place names and the context in which they were named. We extracted all place names referring to places in Hamburg. We repeated the analysis on the second and third corpora and in the end compared what kind of outcome could be made out for the different corpora.

¹ Around 97% precision and 50-60% recall.

Step 4: Mapping life events in Carto.com

The last step in our workflow brings the findings of the data mining analysis explained above into a form which can easily be transferred into the efoto application. All place names identified in step 3 were located on a map using the georeferencing tool carto.com. The efoto database was searched for images matching the place names to illustrate the locations in the city walk. Both geolocations and images are connected in the efoto application to form a "station". Life events or contemporary events connected to the places are used as subjects of these stations in the city walk.

Our model of semi-automated data curation can be visualized as follows:

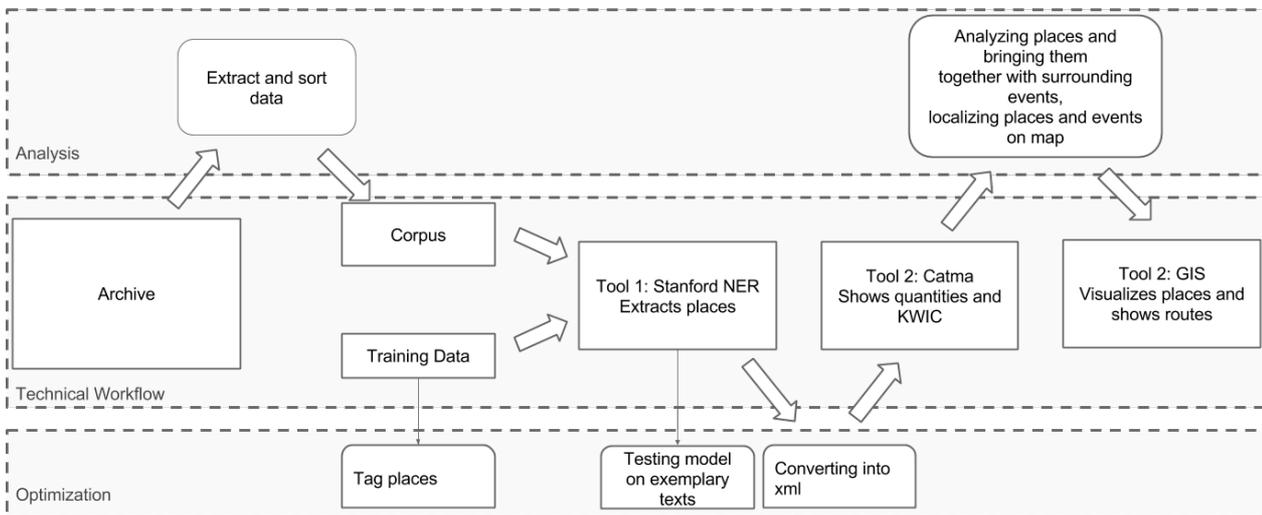


Figure 2: Model of semi-automated data curation.

The digital humanities tools we used were supplemented with analytical methods which cannot be performed automatically. The process of data extraction is in itself an interpretive act, because the decision as to which texts are included in a corpus and which are excluded is made here. Additionally, the findings have to be interpreted before they can be implemented in a city walk. Identifying events in a text is not (yet) a task that a tool can undertake because the category of “event” in a narratological sense is too complex (Dunn/Schumacher, 2016). Some of the tasks mentioned are merely necessary because tools are not yet compatible. In the end this model of semi-automated data curation should be seen as a first proposition for using different Digital Humanities tools in data curation workflows, but it is of course possible that some steps will become obsolete with the further development of tools and methods.

2.2 Findings

2.2.1 Creation of a Stoertebeker city walk

We found nine possible stations for a Stoertebeker city walk. Four of them are situated right in the city centre, one in the direction of St. Pauli, and the others towards the Elbe riverside and even further down in direction of the North Sea. The data mining revealed that one source we included in the first corpus was actually not about Stoertebeker himself but about a man whose nickname was Stoertebeker but who lived some hundred years later (Fock 1913). Thus we excluded this source and the possible city walk station it included retrospectively.

The most interesting outcome for the creation of a Hamburg city walk following events in Stoertebeker's life was that stories about his death were most commonly connected to places in Hamburg. In addition to that, his death story is also narrated in detail in most of the sources. This is not really a surprising outcome as the narrative of Stoertebeker's beheading is spectacular. The legend tells that before he was executed, he asked the city representatives for a favour. After his beheading, he wanted to try to walk along the line of his comrades. Those of whom he passed without his head, should be released. The legend tells that he did walk past a couple of his fellow pirates, before somebody tripped him. These are the common elements of Stoertebeker's death story told in all of the sources we included in the first corpus. There are variable elements, as to how many pirates were sentenced with him, how many of them were passed by the headless Stoertebeker, who made him fall and whether or not those he passed were indeed freed. In most sources the officials of the city broke their word and beheaded all of the pirates regardless.

Stoertebeker's death is always bound to an area of Hamburg called "großer Grasbrook", a peninsula where, today, a quarter called "HafenCity" is built as an urban regeneration project. It is likely that this has been retold uncontroversially, because this area was the common place for execution in Hamburg in the medieval period. We chose the monument of Störtebeker on the "Grasbrook" peninsula to illustrate the story of his beheading. There also is only one version of what happened to the other pirates' heads after their executions. They were hung up as deterrents along the riverside of the Elbe, so this is another place included in the citywalk. The riverside of the Elbe and its beach are a common place for picnicing and sunbathing today and have been a leisure area for some time. As such it is represented in various images in the efoto database, including historical postcards. For the city walk we chose a picture from 1925 showing a group of people in their bathing suits, as a contrast to the site's former usage.

Agreement over other legends about Stoertebeker are less consistent in the sources of the first corpus. There are three different versions of where and how he was imprisoned before his death. The most probable one tells that he was held in a cell in the medieval city hall of Hamburg (Grässe 1854; Rölleke 2001).

This brief legend recounts that the cell became known as “Stoertebeker’s hole” (a literal translation) and that Stoertebeker bargained about his impending execution. The story tells that he tried to free himself by offering a chain of gold long enough to surround the whole city of Hamburg. Another story tells that Stoertebeker spent his last night on earth in a cage in front of the St. Nikolai church (Klabund 1926). In this version he was exposed to very bad weather conditions that made him lose all his clothes. Although a beautiful young lady comes to his rescue he does not take the chance to escape but faces his fate heroically. In a third version (Engel 1920) the pirate awaits his death underneath the altar of St. Catherine’s church. This legend says that he got drunk with his teacher who was imprisoned with him. While drinking they talk about their approaching death. In the efoto database there is one image of the former townhall in which “Stoertebeker’s hole” must have been. There also are some pictures of the “Hopfenmarkt”, a market, which was situated in front of the St. Nikolai church, where the literary source positions Stoertebeker's cage. Unfortunately there are no fotos of the altar of St. Catherine’s but there is a historic image of the church from the outside.

Two versions about what happened to Stoertebeker’s treasure were found in the sources. In both versions people made a crown from the pirate’s gold. This crown was so huge that it could fit around a church’s tower. In one version it was the tower of St. Nikolai (Herzog 1928). It is said in this legend that this church fell in love with the nearby St. Catherine’s church and passed the crown to her. In the other version it was St. Catherine who got the crown right away (Storm 1877). Then there is again disagreement as to what became of the crown. One source said it can still be seen at St. Catherine’s today, the other one states that it was lost to the French (Müllenhoff 1845). Another legend states that the treasure was so huge that the crown was only built after those citizens who had been robbed by the pirate were compensated (Rölleke 2001). There are images in efoto showing both St. Catherine’s and St. Nikolai’s church. For the first one we chose a contemporary photo of its tower, still showing a golden “crown”. As the St. Nikolai church has a special history of its own (being destroyed in the Second World War, its ruin is a memorial today), we chose a picture which shows the whole church in a total view, a perspective that cannot be recreated today.

Whereas Stoertebeker’s death in either version is narrated in most sources, far fewer texts tell how he became imprisoned. Two sources reveal that a fisherman from Blankenese discovered the pirate’s ship and played him a trick (Bechstein 1930; Jensen 1924). He went on board unseen at night and put lead into the oar sockets. He went to the Hanseatic League and when they came to capture the pirates in the morning, Stoertebeker and his men found themselves unable to flee. Another source text states that the fisherman was sent out by the Hanseatic League first of all (Grässe 1854). However, because the geosition

for the ship's berth is not reachable by foot or public transport, we did not include this narrative into our city walk. In addition, there are no images which explicitly show the Elbe in front of Neuwerk Island in the efoto database.

Most stories in our source documents concentrate on Stoertebeker, but in a few stories there is an anecdote about another figure. Although the exact numbers vary, there seem to have been around seventy pirates executed on the day of Stoertebeker's death. The legends say that they were all executed by one man. As the executions were carried out by sword this must have been an exhausting task, so stories say that the hangman was asked afterwards if he felt weary. He answered that he was not and joked that indeed he could go on and kill all the present officials of Hamburg, too. This was too much humour for them and they sentenced him to death as well (Grässe 1854; Rölleke 2001). This anecdote is added to the geosition of Stoertebeker's beheading on the "Grasbrook" peninsula. We chose an image of the "HafenCity" to go with it, as this contrasts very well with the historical story from former times, because during the construction of this quarter people were playing many jokes on some officials of the city.

As intended, the second corpus shows some connecting elements between Stoertebeker's life and contemporary events. It also reveals an interesting archaeological finding from the nineteenth century.

Stoertebeker's death was the frequently retold life event of the pirate in the second corpus as well, compared to his last hours on earth, which played a minor role. The crown that was allegedly made of his gold to decorate either St. Nikolai's or St. Catherine's tower played no role at all in this corpus. However, Stoertebeker's treasure was mentioned in one source as hypothetically being buried on the islands of Rügen or Heligoland (SVZ.DE 2016).

There are more politically significant events connected to Hamburg in this corpus. Some sources explain that Stoertebeker was one of the 'victual brothers' before becoming a pirate (Leipold 2011; Frey 2014; SVZ.DE 2015; Schellen 2010; SVZ.DE 2016; Mittelacher 2015; Welle 2016). Some texts say he robbed only rich merchants and spared the poorer ones thus being some kind of "German Robin Hood" (Welle 2016; Schellen 2010; Mittelacher 2015; SVZ.DE 2015; Frey 2014). One source implies that the city of Hamburg tolerated Stoertebeker's raids even after his time as a freebooter, because his prizes were important for the economy of the city. However, the day came when they could not close their eyes any more (Leipold 2015). None of these findings are especially interesting for the creation of a city walk, because they are not connected to specific places or took place in places too far away from the city to include them in a walk. Nevertheless they are informative about why Stoertebeker is still part of the cultural heritage of Hamburg. He seems to be an ambivalent character, oscillating between a fighter against the rich and an infamous pirate, who turned his back on his former employer. He is compared to Robin Hood,

and to the Taliban. This continual ambiguity makes it impossible to bring the Stoertebeker case to an end. Even if a well respected historiographer states that there was no pirate called Stoertebeker at all, there will always be some kind of doubt about this. One reason for this doubt are the traces and cultural artefacts connected to the medieval pirate. The second corpus shows us how this comes about.

The journalistic texts in the second corpus reveal an interesting fact about the place where pirates were executed in Hamburg. In the 1880s the area had become the historical “Speicherstadt” (literally translated: storage city), still in existence today. During the construction works two reasonably well-preserved skulls were found in the ground (Welle 2016). Both had a huge iron nail in them which led to the conclusion that they had been the heads of pirates executed here (Mittelacher 2015). After their beheading, the heads of pirates were usually nailed to pillars at the river Elbe close to the port, so that everybody entering the city from the sea would see them. We connected the narrative about how the skull, which could have been Stoertebeker’s, was found to the place where the Stoertebeker monument stands today. In our database we found an image showing the construction of the Speicherstadt from above, which we linked to this narrative.

The two most important narratives about the skull found in the second corpus are how a forensic scientist gave it a modern autopsy to find out whether it really could have belonged to Stoertebeker (Mittelacher 2015), and how the skull was stolen from the Hamburg Museum (Schellen 2010; Frey 2014; Focus 2012; Herder 2012; Focus 2012). In the first story, found in one of the source texts, the coroner tells that two facts led to the deduction that this could have been Stoertebeker’s skull. First, and most importantly, is that it was so well preserved, in comparison to the second skull which had been found with it. The large nail was inserted very carefully. It seems that the head of this pirate was to have been recognizable on the pillar for as long as possible (Schellen 2010; Mittelacher 2015). Second, radiocarbon dating has shown that the skull once belonged to a man in his forties, which matches the stories about Stoertebeker’s age when he was executed (Mittelacher 2015). However the genetic material they extracted was too old to be compared with probable descendants of Stoertebeker. So again this narrative does not solve the most important ambiguity about Stoertebeker which is whether he existed or not.

The second narrative found in this corpus, the theft of the skull from the museum and how it was returned, has another interesting function. Not only is there a farcical, even fantastical aspect about stealing the skull of a pirate from a museum, there is also the question of the value of this cultural object. Because the thieves believed the skull to be Stoertebeker’s and because the police and the court invested effort in solving the case, the object as such became more valuable. If it had been just any skull without a chance of being Stoertebeker’s,

neither the press nor the citizens of Hamburg would have been so interested. Belief that this skull was indeed Stoertebeker's increased through this story. As the events in these narratives happened quite recently there were no pictures of them in the efoto database, which holds mostly historic image data.

As the texts of this corpus are factual rather than fictional representations there were not as many variations of the stories as there were in the first corpus, at least not when it came to events connected to the city of Hamburg. There were some reflections on whether Stoertebeker was no pirate but a merchant in Danzig (Frey 2014) and a hypothesis about him being the invention of monks in Lübeck (SHZ.DE 2015), but as these narratives are connected to different places, they are not the focus of our case study.

In the end we identified nine events, which we connected to five places in our city walk. In order to create an attractive route we visualized all ten places we found in the sources in the online geoinformation system carto.db:

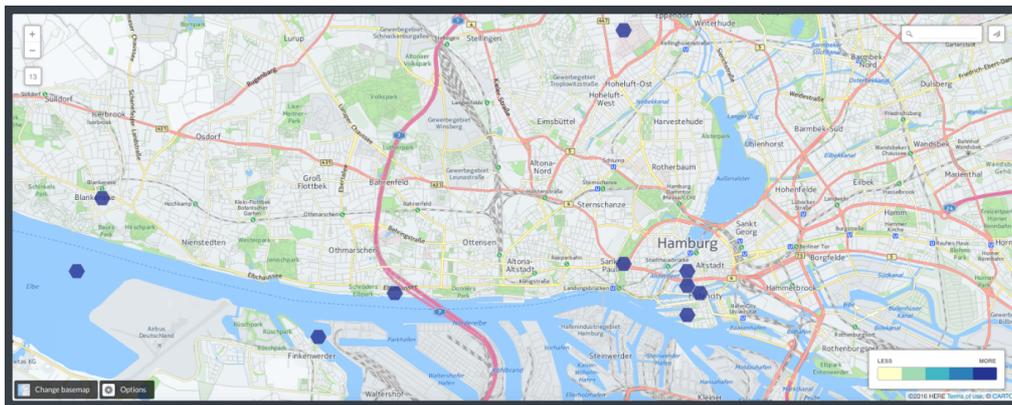


Figure 3: Visualization of Places in Hamburg, which are connected to Stoertebeker.

The table below shows that some places are connected to more than one event and some events to more than one place. Some of the events are told in various narratives. The structure of the city walk, which includes stations with subordinated chapters, offers the option of keeping all the variations of the legends found in the first corpus, thus underlining the multiplicity of Stoertebeker's myth.

Events	Place(s) and images	Variation
<p>Imprisonment</p>	<p>historic town hall</p>  <p>St. Nikolai</p>  <p>St. Catherine's</p> 	<ol style="list-style-type: none"> 1. Cell of the town hall, bargaining for his life 2. Cage in front of St. Nikolai's defying weather and faith 3. Cell underneath the altar drinking and talking with his teacher
<p>Execution</p>	<p>Peninsula Großer Grasbrook</p> 	<ol style="list-style-type: none"> 1. Walking alongside his comrades without his head thus rescuing at least a few of them 2. Walking alongside his comrades without his head until being stopped, comrades are executed anyway

<p>Hanging up the heads</p>	<p>Riverside of the Elbe (“Elbstrand”)</p> 	<p>Heads were hung up along the riverside</p>
<p>Using the treasure</p>	<p>St. Nikolai</p>  <p>St. Catherine’s</p> 	<ol style="list-style-type: none"> 1. Crown for St. Nikolai who passed it on to St. Catherine’s 2. Crown for St. Catherine’s, lost to the French in war 3. Compensation for the people of Hamburg, then crown for St. Catherine’s
<p>Finding the treasure</p>	<p>St. Catherine’s</p>	<p>Carpenter found the pirate’s gold inside the mast of his ship.</p>

Joke by the hangman	Peninsula Großer Grasbrook 	Hangman was asked whether he was exhausted after beheading around 70 pirates. He denied, stating that he could still go on and kill all the officials of the city, too.
Finding the skull	Großer Grasbrook 	During the construction work of the “Speicherstadt” two probable pirate skulls were found
Deducing that one of the skulls was Stoer-te-beker’s	Universitätsklinikum Eppendorf	A forensic autopsy was made
Theft of the skull	Hamburg Museum	Skull was stolen and brought back to the museum

Table 1: Events, places and variations of narratives found in corpus 1 and 2.

Due to the fact that we wanted to create a city walk that could be done by foot and public transport in about two hours, we decided to exclude places from the city walk that could not be easily reached (such as the location on the Elbe near Neuwerk Island), were too far away from the centre of the walk (Blankenese, Eppendorf), or were inside a building (Hamburg Museum). In the end the city walk was implemented in the mobile app as follows:

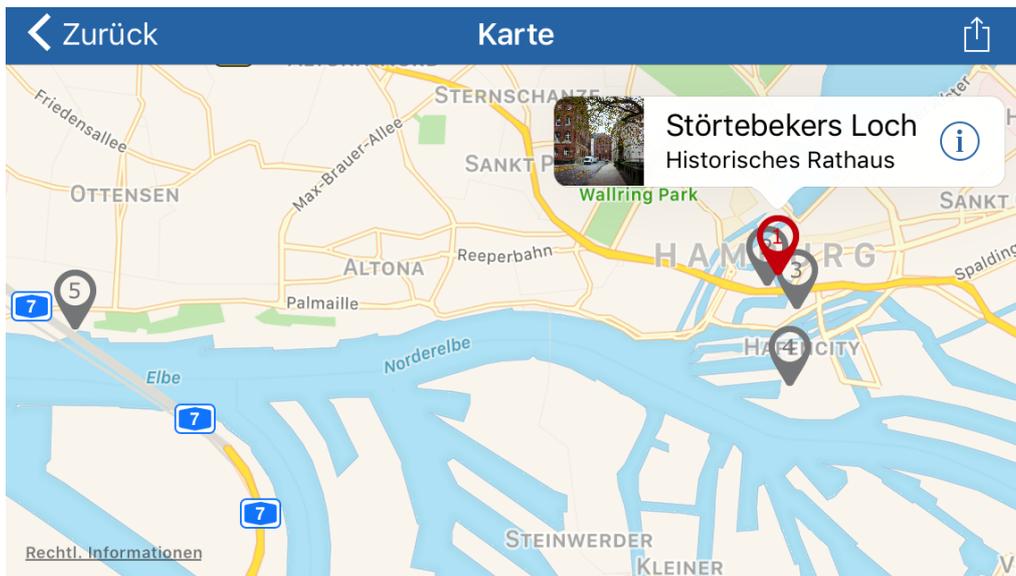


Figure 4: Screenshot of the Stoertebeker citywalk in the mobile app efoto.

2.2.2 Identification with Stoertebeker

Our third corpus revealed how people might connect to Stoertebeker and use narratives about him to create new stories or develop a narrative identity. In Gorch Fock's anthology "Nach dem Sturm" (literally: after the storm) Stoertebeker is used prototypically and allegorically. One sequence shows the protagonist on a ship and wanting to write something. But then he gets haunted by a ghostly vision of Stoertebeker. In this vision the pirate talks to his comrade Godeke Michels who gives him the advise to stay ashore and settle down to an honest family life. But Stoertebeker, and most probably the protagonist as well, is addicted to the sea. He says, he loves it more than he could ever love a woman.

Another passage contains a comparison between the protagonist's ship and the one Stoertebeker used to have, stating that not even the pirate's ship could have been faster (Fock 1941: position 265.315). Using Stoertebeker as comparison here shows that he is some kind of prototype for a man of the sea who is known to have the best and fastest ships. There is a third passage in this source where Stoertebeker's name is used allegorically. It is about a ship which might have been white in the times in which Stoertebeker lived but now looks grey (Fock 1941: position 43.080). Stoertebeker's time is used thus synonymously for a very long time ago.

In two of the other source texts the protagonists are nicknamed Stoertebeker. In one case the nickname is used to mock the first-person narrator, alluding to the fact that he does not seem to distinguish between what is his and what actually belongs to someone else very well (Ringelnatz 1931: position 238.720). In the other case the mistress of the protagonist ironically calls him Stoertebeker

when they are just about to eat during a sea voyage. While the woman does not mind the state of the sea at all, the protagonist is not able to look at food without feeling sick (Tucholsky 1956: position 21.968). Once again the nickname Stoertebeker is used with a mocking tone or at least makes the addressed person feel ashamed. In this case the shame rather seems to be caused by a feeling of not living up to a certain role model. But how did a medieval pirate and freebooter become a role model in the first place? The fourth text in the third corpus sheds some light on that.

In Theodor Fontane (1959) the author narrates an episode from his childhood. He used to play with his friends in a place called “Stoertebeker’s well” (author’s translation). That was a place in the dunes on the Baltic coast, where the protagonist used to go camping. There is a legend saying that Stoertebeker used this place to hide with his comrades when not at sea. Being there, camping with his friends just like the pirate probably did hundreds of years ago, causes a real sense of delight in the young boy. Doing the same thing as the pirate makes him feel very strong and brave. But then the feeling changes as he reminds himself that Stoertebeker was executed in Hamburg. This thought is rather unpleasant to him and he has to remind himself that it is not very likely that he would be pursued and caught by the court of Hamburg. So we end up with an ambiguity with which the legendary person of Stoertebeker is regarded. There is admiration for his bravery and strength, but in the very next moment thoughts about his violent death, which was caused by the fact that he was a pirate and no honourable man. This results in an unsettled feeling. The autobiographical narrator in this source uses a vivid term, literally translated as “sweet shudders”, to describe how he feels comparing himself to Stoertebeker.

To sum up this corpus shows different aspects of identification with Stoertebeker. He seems to incorporate a certain role model of a sailor who is especially brave and fearless. Nevertheless it seems to be hard to come to a judgement about him. On the one side his courage is underlined by his defiance of all authority and not even being intimidated by a sentence of death. On the other hand he was undeniably dishonest, which is not an admirable trait. Nevertheless Stoertebeker seems to stand for a prototype of masculinity, which little boys as well as grown men would like to match. The name Stoertebeker is used allegorically or as a projection surface. The ambiguity of his person strengthens the legend about him and makes people connect to him in different ways.

3. Conclusions

Starting from very high-level and merely abstract models of culture and narrative identification, we developed a practical model of semi-automated data curation. We tested this model on the case of Stoertebeker and his legends showing that the results could be implemented into a mobile application as well as led back to further investigations on the relevance of Stoertebeker’s myth today.

In our practical attempt we showed that interdisciplinary scientific workflows, methods and techniques of data mining can be used for digital content curation. We showed that investigations on two aspects of data curation could be pursued. Digital Humanities tools and methods can be used to create a city walk, and to decide which places and events should be included in a narrative representation of certain aspects of cultural heritage. Multiplicity, as one probable result of the mostly oral tradition of legends, turned out to be an especially enriching feature in our case study. Further some facets of the ongoing fascination with a mythical personality as Stoertebeker's were revealed.

Our model of semi-automated data curation is a first attempt to try out a workflow of Digital Humanities tools and methods in a process of data curation. This article is not a contribution to the discussion of big data analysing techniques, but it shows that a set of corpora, exceeding in size what one person could read closely in a time span appropriate for data curation, could be used to create a narrative representation of a historical topic. We do not try to exclude the expertise of cultural heritage curators, but instead want to highlight the importance of interpretation in data mining.

It is beyond the scope of this paper to prove whether the resulting city walk really functions as a trigger for the emergence and identification of cultural heritage. This will show when the implementation of the mobile application is completed and users test the Stoertebeker city walk. The model of semi-automated data curation should be tried out on other subjects in order to show whether it is fully or merely partly transferable. We hope that our model of semi-automated data curation can function as a starting-point for further investigations on data mining processes in cultural heritage.

Works cited

- Adorno, Theodor W.; Horkheimer, Max (2003): *Dialektik der Aufklärung*. Frankfurt am Main: Fischer Verlag.
- Ahrendt, Hannah (2010): *Vita Activa oder Vom tätigen Leben*. München: Piper.
- Bamberg, M. (2011): "Who am I? Narration and its contribution to self and identity". In: *Theory & Psychology*. 21 (1), p. 3-24, DOI: 10.1177/0959354309355852.
- Bamberg, Michael (2012): "Identity and Narration - the living handbook of narratology". *Wikis.sub.uni-hamburg.de*. View date 17.11.2016 http://wikis.sub.uni-hamburg.de/lhn/index.php/Identity_and_Narration.
- Benjamin, Walter (1963): "Das Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit". In: *Drei Studien zur Kunstsoziologie*. Frankfurt am Main: Suhrkamp.
- Bracker, Jacobus (2013): "Einleitung: Homo pictor meets homo narrans". In: *Visual Past. A journal for the study of past visual cultures*. View date 29.3.2017 http://www.visualpast.de/archive/pdf/vp2016_0001.pdf.
- Bracker, Jörgen (2005): "Störtebeker. Nur einer von ihnen. Die Geschichte der Vitalienbrüder". In: Ehbrecht, Wilfried (Hg.): *Störtebeker 600 Jahre nach seinem Tod*. Hansische Schriften 15, p. 37-56. Trier: Pota Alba.
- Cameron, Christina; Rössler, Mechthild: *The Shift towards Conservation. Early History of the 1972 World Heritage Convention and Global Heritage Conservation*. In: *Understanding Heritage. Perspectives in Heritage Studies*. Berlin: De Gruyter.
- Dennerlein, Katrin; Hühn, Peter (2009): *Narratologie des Raumes*. Berlin [etc.]: Walter de Gruyter.
- Dormaels, Mathieu (2013): "The Concept behind the Word. Translation Issues in Definition of Heritage". In: *Understanding Heritage. Perspectives in Heritage Studies*. Berlin: De Gruyter.
- Dunn, Stuart/Schumacher, Mareike (2016): "DHO: Digital Humanities Quarterly: Explaining Events to Computers: Critical Quantification, Multiplicity and Narratives in Cultural Heritage". *Digitalhumanities.org*. Abgerufen am 18. 11. 2016 von <http://www.digitalhumanities.org/dhq/vol/10/3/000262/000262.html>.
- Eagleton, Terry (2001): *Was ist Kultur?* München: C.H. Beck
- European Union (1995-2017): "Inspire Guideline". Abgerufen am 26.9.2017 von <https://inspire.ec.europa.eu/data-specifications/2892>
- Fisher, Walter R (1984). "Narration as Human Communication Paradigm: The Case of Public Moral Argument". *Communication Monographs*. 51: 1–22. doi:10.1080/03637758409390180
- Freie und Hansestadt Hamburg: "eCulture Agenda 2020". Abgerufen am 26.9.2017 von <http://www.hamburg.de/content-blob/4359742/51291cc39d56f6ed227d05bdf6e5a4e/data/eculture-agenda-2020.pdf>.
- Geertz, Clifford (1973): *The interpretation of cultures*. New York: Basic Books.

- Georgakopoulou, Alexandra (2007): *Small stories, interaction and identities*. Amsterdam: John Benjamins Pub.
- Koppmann, Karl (1877): "Der Seeräuber Klaus Störtebeker in Geschichte und Sage". In: *Hansische Geschichtsblätter*. 7, p. 37-60, Abgerufen am 19. 07. 2017 von http://www.hansischergeschichtsverein.de/download/hgbll1877_volltext.pdf.
- Luhmann, Niklas (1995): *Die Kunst der Gesellschaft*. Frankfurt am Main: Suhrkamp.
- Luhmann, Niklas (1997): *Die Gesellschaft der Gesellschaft*. Frankfurt am Main: Suhrkamp.
- Müller-Funk, Wolfgang (2006): *Kulturtheorie. Einführung in die Schlüsseltexte der Kulturwissenschaft*. Tübingen: UTB.
- Müller-Funk, Wolfgang (2008): *Die Kultur und ihre Narrative*. Wien: Springer.
- Nünning, Ansgar (2014): "Wie Erzählungen Kulturen erzeugen: Prämissen, Konzepte und Perspektiven für eine kulturwissenschaftliche Narratologie". In: *Kultur. Wissen. Narration*. Berlin: Transcript.
- Puhle, Matthias (2012): *Die Vitalienbrüder*. Klaus Störtebeker und die Seeräuber der Hansezeit. Frankfurt: Campus.
- Ricoeur, Paul (1991): "Narrative Identity". In: *Philosophy Today*. 35 (1), p. 73-81, DOI: 10.5840/philtoday199135136.
- Rohmann, Gregor (2007): "Der Kaperfahrer Johann Störtebeker aus Danzig. Beobachtungen zur Geschichte der 'Vitalienbrüder'". In: *Hansische Geschichtsblätter*. 125, p. 77-119.
- Rohmann, Gregor (Hrsg.) (2011): "Klaus Störtebeker und die Vitalienbrüder". In: *Die Welt des Mittelalters. Erinnerungsorte eines Jahrtausends*. 1. Aufl. München: Johannes Fried; Olaf B. Rader, p. 246-260.
- Schwalm, Helga (2014): "Autobiography | the living handbook of narratology". *Lhn.uni-hamburg.de*. Abgerufen am 17. 11. 2016 von <http://www.lhn.uni-hamburg.de/article/autobiography>.
- Seitter, Walter (2004): "Michel Foucault (1926-1984). Struktur, Entscheidung, Ordnung, Stil". In: Hofmann, Ludwig Martin; Korta Tobias F.; Niekisch, Sibylle (Eds.): *Culture Club*. Frankfurt am Main: Suhrkamp.
- Silbermann, Neil A.: *Heritage Interpretation as Public Discourse. Towards a New Paradigm*. In: *Understanding Heritage. Perspectives in Heritage Studies*. Berlin: De Gruyter.
- Vecco, Marilena (2010): "A definition of cultural heritage: From the tangible to the intangible". In: *Journal of Cultural Heritage*. 11, p. 321-324

Corpus 1

- Bechstein, Ludwig (1930): *Deutsches Sagenbuch*. *Textgrid Repository*. [online] Available at: https://textgridrep.org/browse/-/browse/kh3k_0, [Accessed 18 Nov. 2016].

- Engel, Georg (1920): Claus Störtebeker. *Project Gutenberg*. [online] Available at: <https://www.gutenberg.org/ebooks/41382>, [Accessed 18 Nov. 2016].
- Fock, G. (1913): Seefahrt ist Not! *Textgrid Repository*. [online] Available at: https://textgridrep.org/browse/-/browse/n0x8_0, [Accessed 19 Nov. 2016].
- Grässe, Johann Georg Theodor (Ed.) (1854): “Klaus Störtebeker und Goedeke Michels”. In: *Sagenbuch des preußischen Staates*. 1. Aufl. Hannover. *Textgrid Repository*. [online] Available at: https://textgridrep.org/browse/-/browse/10v3n_0, [Accessed 18 Nov. 2016].
- Heinz Rölleke (Ed.) (2001): “Klaus Störtebeker und Goedeke Michels”. In: *Das große deutsche Sagenbuch*. 1. Aufl. *Project Gutenberg*. [online] Available at: <http://gutenberg.spiegel.de/buch/sagen-aus-der-hanse-33/12>, [Accessed 18 Nov. 2016].
- Herzog, Rudolf (1928): Hanseaten. *Project Gutenberg*. [online] Available at: <http://gutenberg.spiegel.de/buch/hanseaten-7657/1>, [Accessed 18 Nov. 2016].
- Jensen, Wilhelm (1924): Vor der Elbmündung. *Project Gutenberg*. [online] Available at: <http://gutenberg.spiegel.de/buch/-3807/8>, [Accessed 18 Nov. 2016].
- Klabund, (1926): *Störtebeker*. *Project Gutenberg*. [online] Available at: <http://gutenberg.spiegel.de/buch/stortebecker-2534/1>, [Accessed 18 Nov. 2016].
- Müllenhoff, Karl (1845): Klaas Störtebeker und Göde Micheel. *Textgrid Repository*. [online] Available at: https://textgridrep.org/browse/-/browse/snhc_0, [Accessed 18 Nov. 2016].
- Storm, T. (1876): Aquis Submersus. *Textgrid Repository*. [online] Available at: https://textgridrep.org/browse/-/browse/vt27_0, [Accessed 19 Nov. 2016].

Corpus 2

- Focus, (2012): Dieb erbeutet Störtebeker-Schädel. *Focus Online*. [online] Available at: http://www.focus.de/wissen/mensch/archaeologie/piraten-dieb-erbeutet-stoertebecker-schaedel_aid_472303.html [Accessed 19 Nov. 2016].
- Focus, (2012): Drei Jahre nach Raub in Hamburger Museum Gestohlener “Störtebeker-Schädel” diente als Partygag. *Focus Online*. [online] Available at: http://www.focus.de/panorama/welt/drei-jahre-nach-raub-in-hamburger-museum-milde-strafen-fuer-diebe-des-stoertebecker-schaedels_aid_877282.html [Accessed 19 Nov. 2016].
- Focus, (2012): Störtebeker-Schädel geklaut: Drei Männer angeklagt. *Focus Online*. [online] Available at: http://www.focus.de/panorama/welt/kriminalitaet-stoertebecker-schaedel-geklaut-drei-maenner-angeklagt_aid_713788.html [Accessed 19 Nov. 2016].
- Frev, C. (2014). Klaus Störtebeker war ein Danziger Kaufmann. *Die Welt Online*. [online] Available at: <https://www.welt.de/geschichte/article134534085/Klaus-Stoertebecker-war-ein-Danziger-Kaufmann.html> [Accessed 19 Nov. 2016].

Herder, Daniel (2012): Störtebekers Schädel drohte die Schrottpresse. *Hamburger Abendblatt*. 2012.

Leipold, F. (2011). Vitalienbrüder: Taliban der Ostsee. *Focus Online*. [online] Available at: http://www.focus.de/wissen/mensch/geschichte/tid-23953/vitalienbrueder-taliban-der-ostsee_aid_676450.html [Accessed 19 Nov. 2016].

Mittelacher, Bettina (2015): Der geheimnisumwitterte Schädel vom Grasbrook. *Hamburger Abendblatt*. 2015.

Schellen, Petra (2010): Archäologe Wiechmann zu Störtebeker: "Enthauptung war eine ehrenhafte Strafe". *taz*. 2010.

SHZ.DE, (2015). Klaus Störtebeker: Nur eine Erfindung von Lübecker Mönchen. [online] Available at: <http://www.shz.de/regionales/schleswig-holstein/kultur/klaus-stoertebeker-nur-eine-erfindung-von-luebecker-moennen-id9905331.html> [Accessed 19 Nov. 2016].

SVZ.DE, (2016). Das Ende eines großen Seeräubers. [online] Available at: <http://www.svz.de/deutschland-welt/panorama/das-ende-eines-grossen-seeraeubers-id13748371.html> [Accessed 19 Nov. 2016].

Welle, Florian (2016): Dem Geheimnis auf der Spur. Robin Hood der Meere. *Süddeutsche*. 2016.

TLZ.DE, (2015). Störtebeker: Ein berühmter und gefürchteter Seeräuber. [online] Available at: <http://www.tlz.de/kinder/detail/-/specific/Stoertebeker-Ein-beruehmter-und-gefuechteter-Seeraeuber-701419806>, [Accessed 19 Nov. 2016].

Corpus 3

Fock, G. & Bussmann, A. (1941): Nach dem Sturm. Projekt Gutenberg, <http://gutenberg.spiegel.de/buch/nach-dem-sturm-2600/1>, [Accessed 19 Nov. 2016].

Fontane, T. & Gross, E. (1959): Meine Kinderjahre. Textgrid Repository, https://textgridrep.org/browse/-/browse/n10p_0, [Accessed 19 Nov. 2016].

Ringelnatz, J. (1931): Mein Leben bis zum Kriege. Textgrid Repository, https://textgridrep.org/browse/-/browse/td99_0, [Accessed 19 Nov. 2016].

Tucholsky, K. & Raddatz, F.J. (1956): Schloss Gripsholm. Projekt Gutenberg, <http://gutenberg.spiegel.de/buch/schloss-gripsholm-1186/1>, [Accessed 19 Nov. 2016].