Uitgewerkte aanvraag

1) Main applicant: Prof.dr.J.S. (Judith) Pollmann, Leiden University Institute for History, Leiden University, PO Box 9515, 2300 RA Leiden

2) Title: *Chronicling novelty. New knowledge in the Netherlands, 1500-1850.*

3) Summary

For innovation to happen it is not enough that new ideas and technologies are being invented. Cultural factors play an essential role in their acceptance and appropriation. Recent scholarship hypothesises that Europeans after 1650 became more receptive to new technology and innovation than their ancestors, and so enabled the Enlightenment and Industrial Revolution. The spread of new knowledge and techniques among scholars and specialists between 1500-1850 is indeed well-documented. Yet since acceptance by specialists does not guarantee wider acceptance, we will study how and to what effect, new knowledge actually anchored among the wider public.

This project focuses on the circulation and evaluation of new knowledge, ideas and technologies among a non-specialist public of middle-class authors in the Netherlands, who kept handwritten chronicles to record events and phenomena that they considered important. We develop a method to use them in large numbers and comparatively, so as to track and analyse the circulation, evaluation and acceptance of old and new ideas and information over time and spatially.

We will create a large high quality annotated corpus of texts, develop computational tools to trace patterns in topics, perspectives and appreciation of novelty and to alert us to passages that require further, qualitative analysis by close reading. In this way we will assess the circulation of new ideas, their reception, and the impact on attitudes to novelty and tradition in wider society. (229 words)

4) NWO research fields:

Hoofddiscipline	Geschiedenis	
Subdiscipline	27.60.00	Culturele geschiedenis
Hoofddiscipline	Computers en geesteswetenschappen	
Subdiscipline	37.20.00	Tekst- en content analyse

5) Infrastructurele component Yes

6) Herindiening: Yes, Geschiedenis, 2016 (VC.GW16.024)

7) Institutionele omgeving: Leiden University Institute for History; VU Inter-facultary Research Institute CLUE+: research programme on 'Knowledge Formation and its History', VU Network Institute: Computational Lexicology & Terminology Lab (CLTL) and the VU-based Stevin Centre for History of Science and Humanities.

8) Subsidieperiode: 1 September 2018 – 31 December 2023

9) Research team

Name	Affiliation	Expertise	Type of involvement
Prof. dr. Judith Pollmann	Leiden University	Cultural History	Project leader, thesis supervisor
Dr. Erika Kuijpers	VU	Social and cultural history, digital humanities	Project leader infrastructure, co-supervisor PhDs
NN	Leiden University	History + training data analysis and visualization (space, time, networks)	PhD 'mediascapes'
NN	Leiden University	History + training automatic text analysis (sentiments, topics)	PhD 'reasoning'
NN	VU	computational linguistics, semantic text analysis	Postdoc 'moments of change'
NN	VU	Knowledge of palaeography, Digital Humanities	Student assistant crowdsourcing and data curation
NN	VU	Digital Humanities	Student assistant annotations/evaluations
Prof. dr. Piek Vossen	VU	Computational linguistics, semantic text analysis	Advisor, supervision Postdoc
Dr. Roser Morante	VU	Computational linguistics, semantic text analysis	Advisor, supervision Postdoc
Dr. Martin Reynaert	Tilburg University	Computational linguistics	Advisor
Prof. dr. ir. Fokko Jan Dijksterhuis	VU/Universiteit Twente	Cultural History, History of Knowledge	Advisor

Prof. dr. James Amelang	Universidad Autónoma de Madrid	Cultural History, chronicling	Advisor
Prof. dr. Susanne Rau	Universität Erfurt	Cultural History, chronicling	Advisor
Prof. dr. Anne Laure van Bruaene	Ghent University	Medieval History, chronicling	Advisor
Dr. Chiara De Caprio	University of Naples Federico II	Cultural History, chronicling, orality	Advisor

10) Structure of the proposed research

I.PhD1 (UL), **Changing mediascapes and the collection of knowledge,** main supervisor Pollmann, co-supervisor Kuijpers

II.PhD2(UL): **Reasoning, causality and the appreciation of novelty,** main supervisor Pollmann, co-supervisor Kuijpers

III.Postdoc (VU Network Institute, CLTL) **New knowledge: tracking topics, attributions and speculations,** (VU Network Institute, CLTL, co-supervision computational linguistics: Prof. dr. P. Vossen/ Dr. Roser Morante)

IV. Synthesis: **Chronicling change** (UL & VU), Pollmann & Kuijpers

11a) Description of the proposed research: (1196)

Scholarly relevance:

In 1712 Holland was struck by a devastating epidemic of cattle-disease. In the village of Huizen, local Lambert Lustigh chronicled day by day in which households cows had died. In some respects Lustigh's understanding of the epidemic was entirely traditional. He saw it as a sign of divine anger, that required a moral response. Yet from his observations, by conversations and reading about the epidemic in other parts of Holland, Lambert Lustigh simultaneously developed a theory that focused on scientific explanations. In his view, beasts got sick when they were stabled, because close proximity and higher temperatures exposed them to bad air. Unfortunately, his explanation did not persuade his fellow-villagers; these continued to stable their cows together (Peeters).

For innovation to happen it is not enough that new ideas and technologies are being invented. Research on current-day innovations suggests that the chances of a new idea catching on strongly depend on human factors, that vary from culture to culture. Sluiter (2017) argues that an analysis of innovation in historical societies can help us to map, evaluate and diversify strategies that are

used successfully to 'anchor innovation'. This project focuses on one such historical society: the Netherlands between 1500-1800.

Seeking an explanation for the acceleration of economic change in Europe between 1650-1900, economic historians have suggested an important role for cultural changes that facilitated a greater acceptance of innovation (Slack, Davids). Mokyr (2016) sees the relatively open exchange of ideas in early modern Europe as the key to understanding why notions of progress emerged, which ousted earlier suspicion of 'novelty'. This made Europeans after 1650 more receptive to new technology and innovation than their ancestors, and so enabled the Enlightenment and Industrial Revolution.

While the move to connect economic to cultural developments is very welcome, this analysis of the relationship between cultural change and innovation in this period raises **issues that require further investigation**.

First, it assumes that once people welcome one form of innovation, they will also be open to another. Considering that people are adept at operating several, and even contradictory, cultural systems at once (Swidler), it is not self-evident that enthusiasm for novelty in one domain precipitated acceptance of other forms of innovation, or that the rise of new knowledge led to the abandoning of older ways of thinking. Although attitudes towards new technologies, for instance, were positive at least as early as the sixteenth century, the negative evaluation of social, political and religious innovation persisted far longer (Pollmann 2017).

For that reason, we need to know much more about the **reception** of new knowledge if we are to establish its relationship to broader cultural change. Studies of the reception of new ideas and technologies after 1650 have focused on issues of circulation and acceptance by other scholars (e.g. Burke 2000 & 2012, Blair, Shapiro, Mokyr 2016) and/or on the spread of new skills and techniques among professional users and producers (Mokyr 2002, Epstein, Smith and Schmidt, De Munck, Dupré and Mijnhardt, Davids). Yet Roberts, among others, stresses the importance of 'distributed cognition': since acceptance by specialists does not guarantee wider appropriation and broader cultural change, we need to study when, how, and to what effect, new knowledge actually anchored among the wider public. Chroniclers like Lustigh both consumed and produced shared cognition (Roberts).

In order to analyse the anchoring of new knowledge in this period, this project focuses on the **appropriation of new knowledge, ideas and technology by a public of non-specialists** between 1500-1850. To study this we have identified a corpus of **chronicles**: high quality, non-institutional source-material that lends itself to comparison over time. From the late Middle Ages until 1850, thousands of middle-class men (and some women) across Europe kept chronological records of current events in their communities and beyond, so as to archive, frame and evaluate events and phenomena that they considered important, 'memorable' and 'useful'. Chroniclers' understanding of '**useful**

knowledge' was diverse and broader than that defined by Mokyr (2002): information about the weather, (super-)natural phenomena, prices, murders, epidemics, politics and technical innovations all had a place (Amelang, Pollmann 2016). In the process, chroniclers generated a remarkably stable, cross-European genre of narrative evidence, that we hypothesize can be used as a benchmark across five centuries.

We propose to create a **digitized corpus** of chronicles as this will allow us to combine in depth analysis with tracing and comparing changes over time and spatially. To reassess the relationship between cultural change and innovation, we focus on three key issues (Sarasin) in the history of knowledge:

- What knowledge is considered *collectible* and why? What are the authors' sources of authority? Do these change as the flow of information increases? What do chroniclers do with contradictory information?
- How do people *appropriate*, reproduce and structure knowledge? How does new knowledge acquire epistemic and moral authority? Does it replace or complement old knowledge?
- How do the information-flows and new ideas *affect* people? Does local knowledge become less important? Does increased exposure to the new make people less or more averse to novelty?

Method

Our method combines the techniques of 'close' and 'distant reading'. Close reading is a hermeneutical method for interpreting an individual text in its historical context. 'Distant reading' is a method that uses computational techniques to detect, visualize and interpret cultural trends in a larger collection of texts. To discover such trends in our data we will develop automatic semantic processing techniques.

Although our long-term aim is to do cross-European research, in this project we will study the genre in Dutch-speaking areas in the Low Countries. These areas make an interesting **pilot** since people here were exposed to many types of new knowledge. Literacy and numeracy levels were high, the many towns were centers of printed media and knowledge production, but the pace of innovation and change varied regionally (Dupré and Mijnhardt). Chronicling was a widespread practice and a good number of chronicles are already published (Bruaene, Pollmann 2016). We start in 1500, when non-institutional chronicling took off in the Low Countries. The end date 1850 allows us to include the period that is associated with the coming of modernity.

Creating a corpus. The project's **basic viability** is guaranteed by the existence of 115 Dutch chronicles that exist in older print editions (c. 75) or in digital transcript (40) (c. 25.000 pages). The Royal Library/DBNL will undertake digitizing, OCR correction and encoding of this printed material. In order to

prepare the data set for computational analysis, metadata and the annotation of named-entities (persons, institutions, locations) as well as dates and sources of information will be added manually through crowdsourcing to ensure a high quality coded corpus. In order to **broaden the corpus** and also include unprinted manuscript material we will combine automatic character recognition (via Transkribus/READ) with proven crowdsourcing techniques to transcribe and code a selection of c. 100 unpublished manuscripts. We will do so with the help of the crowdsourcing platform VeleHanden (see 11b).

Proven and improved supervised computational techniques will be used to evaluate the data, to normalize the corpus into modern Dutch and feed it into a Natural Language Processing (NLP) pipeline. A high quality curated, normalized and manually annotated corpus will allow us to use existing tools for manual and automatic text analysis developed for modern Dutch in order to track trends over time in the choice of topic, the use of sources of information and to assess and visualize knowledge horizons.

The team's expert knowledge on both historical language use and cultural meanings will allow the postdoc, a computational linguist, to develop automatic methods for detecting novelty and uncertainty with the help of manual annotation, machine learning and evaluation by the team of experts. The semantic processing will allow the historians on the team to identify moments of uncertainty, speculation and change for further qualitative analysis.

Analytical approach. While chronicles retained their basic features throughout this period, preliminary research suggests that we will be able to identify and analyse a number of significant **changes** that can help us establish the evolving appropriation and evaluation of new knowledge.

1.Chroniclers around 1800 used a **broader range of sources of information** than they had three centuries earlier. While continuing to use other eyewitnesses and local authority figures, there were changes in their reliance on written and visual sources (newspapers, statistics, and maps), in the tools they used (watches and thermometers), and the way they presented data themselves (tables, more numerical lists).

2.The **range of topics expanded**. This may be because printed information arrived from further afield, and the known world got larger, but also because exposure to new knowledge and media suggested new topics.

3.Studies of individual chronicles show that authors carefully weighed the relative merits and authority of different sources of information (Rau, Mauer). By studying how they assessed these merits, we can establish, whether **the authority** of written texts increased relative to local knowledge, or whether traditional authoritative voices (like pastors) retained their authority. During epidemics, wars and political **crises**, particularly, chroniclers indicated the importance of evaluating reliable and unreliable information (Van Nierop).

Finally, we can use chronicles to test earlier claims that the proliferation of new media led to **information overload** and gave rise to scepticism (Dooley, Shapiro, Rosenberg).

4. Preliminary work (Pollmann 2016) argues that by selecting, evaluating and combining local observations and information from other sources, authors probed **correlations and patterns**: between wet winters and infestations by mice, high prices and war, dissent and crime. Explicit expression of causation was uncommon at first, yet may have developed over time. Most of Lustigh's ideas on cattle disease belonged to the eighteenth century; a sixteenth-century author would have been more likely to consider witchcraft or astrological factors. Yet as his example also indicates, chroniclers operated new and old ideas side by side. By studying these changes, we can analyse to what extent the arrival of new knowledge encouraged people to abandon existing sources and thought systems, or merely expand their repertoire of ideas.

Innovative aspects

1 Unlike their well-studied medieval antecedents, early modern chronicles have been rather neglected. When studied at all, they have been analysed as individual texts and as a form of urban historiography (e.g. Schmidt, Fuchs, Rau 2002). We propose **to study chronicles as collections of useful and memorable knowledge** that were archived for future use (Pollmann 2016; Rau 2017), and so reflect the appropriation and circulation of information and knowledge in their community.

2 To study systematically how chroniclers acquired their information, how they judged reliability, and how they established explanatory patterns, we propose to study chronicles thematically, comparatively and in much larger numbers than has been attempted so far, so as to track **long-term and spatial variation**.

3 **Interdisciplinary collaboration** of computational linguists with historians allows for the use of expert knowledge of the historical source material to develop a method for systematic analysis of linguistic cues that signal the introduction of new ideas. Thus we develop **a model for automatic detection of novelty** in an early modern corpus.

4 Finally we expect to make a **contribution to computational linguistics** as we will work on methods for normalization of historical Dutch and semantic processing.

Coherence and collaboration:

For this project collaborative team effort is essential, first, to create the corpus, secondly, to develop appropriate metadata and annotation models, and third to carry out the manual annotation and expert evaluation of the results of computational methods. Historical knowledge and qualitative approaches are essential to the understanding of the source material and to further refine and

evaluate the semantic processing. The team members will share the corpus and technical infrastructure, and jointly consider methodological and interpretative meta-issues. There will be bi-weekly, and later monthly, meetings.

The team will also benefit from: (a) the joint expertise in social and cultural history and proven collaborative success of the two applicants; (b) close collaboration with the VU experts in semantic text mining technology; (c) the vibrant research communities in early modern cultural history at Leiden and the VU.

Subprojects

I.   **Changing mediascapes and the collection of knowledge** (PhD1)

Scholars agree that cultural changes in early modern Europe were both accompanied and precipitated by an information revolution, that consisted not only of the spread of more printed material and of new instruments for observation, but also of new models for collecting, archiving, listing, tabulating and mapping information. Recent media historians have enormously expanded our understanding of the production, circulation and marketing of news media New media, such as newspapers and journals, as well as new tools, such as thermometers, watches and maps, promised access to 'accurate' data (Pettegree, Soll).

Preliminary work by the applicants suggests that much of this filtered down into local chronicles, which, while retaining their basic chronological structure and an interest in traditional topics (climate, prices, crime, deaths) came to include new topics and used new textual tools such as lists and tables. Authors frequently copied excerpts from official documents, local announcements and by-laws, and increasingly inserted printed material, like ballads, pamphlets, and newspapers. Others kept chronicles in almanacs (Salman). By the late eighteenth century, some chronicles consisted of a running commentary on clippings from other material (Walleghem). Interestingly, the plethora of printed information did not make the chronicling redundant. Chroniclers selected the information of their choice, and structured and framed it as they saw fit, so imposing meaning on events as they occurred.

This means that chronicles offer us an opportunity to study not just the *consumption* of knowledge, but also the *engagement* with it. This is significant, because although we have some evidence for book ownership in inventories, systematic evidence of readership, let alone appropriation, of media among a non-scholarly public is rare (Salman, Van Groesen). Case studies of the reading habits in the Dutch Republic and elsewhere have established how useful chronicles and journals can be in this respect, yet so far they have relied on individual case studies (Blaak, Mauer). Important as these are, such studies have not allowed an analysis of developments over time.

The size of the corpus and the manual annotation of the chroniclers' sources and topographical information (see part 11b) will enable the PhD **to map the evolving use of media and new instruments as sources of information among non-specialists between 1500-1850 and across the Netherlands**. Through spatial visualisations of media use and topics over time this researcher will analyse

a) **information flows** and the extent to which non-specialists encountered newer types of media and information (observation, oral, manuscript, books, newspapers, images, instruments).
b) what sources they **privileged**, and how this changed. Did the importance of oral information recede as there was more printed information available? What was the relative weight of different sources?
c) how this affected the range of topics in chronicles and the chroniclers' **knowledge horizon**? Did the world get larger as more and other media became available over time, did this inspire new topics of interest, and lead people to contextualise local and trans-local events differently?
d) how the emergence of new media and the growing volume of information affected the **authority** of the sources. What information, from which source was considered (un)reliable? How did they frame new knowledge? Did chroniclers experience information overload and what impact did this have? (Dooley, Blair, Rosenberg).
e) **regional differences** in media use e.g. between the Dutch Republic and the Habsburg Netherlands, because of differences in religious and censorship regimes, and between towns and countryside, printing centers and places without printers, as well as between ports and commercial centres and the agrarian hinterlands.

Within the chronicles, we expect the closer study of **crises** to be especially rewarding. Chronicles tend to get much richer and more detailed in situations of uncertainty. During political and military crises, when there were frequent attempts to manipulate and frame news, many authors were very explicit in their reflections on the (un)reliability of information (Van Nierop, Deseure). Crises caused by epidemics, floods and freak weather, as well as economic crises, triggered a hunger for additional data, as well as speculation on correlations, causes and consequences. For that reason, the latter category will also be important for subproject II.

II. Reasoning, causality and the appreciation of novelty (PhD2)

This subproject investigates changes over time in the

- **evaluation of cause and effect**
- **the evaluation of change**

Using a combination of distant and close reading, this project will

(a) assess and analyse **correlations and patterns** that chroniclers identified.

(b) analyse the **emergence of new knowledge**, concepts and evidence in patterns and causal relationships that authors indicated.

(c) assess how this affected their **ideas on tradition and innovation.**

Many early modern chroniclers collected data on prices, climate, crime, epidemics and (super-)natural phenomena. While earlier scholars have often seen these as 'trivial' side-issues (Fuchs), their ubiquity points to the fact that these issues were core business for chroniclers. In line with our Huizen example, we hypothesise that the perceived *value* of such information was that it could be used to establish **patterns and correlations**, and so outline scenarios for the future.

For our purposes, it is precisely the consistent interest in these topics that make chronicles ideal for a comparison over time, and to chart how patterns and correlations evolved under the influence of new knowledge (e.g. cattle disease – God - sin, as opposed to cattle disease – weather conditions – infection). Chronicles discuss the **weather** so frequently that historians have used them to chart climate change (Draelants). While many left it implicit, some chroniclers articulated that they recorded weather data to establish a 'rule'. The way they did that changed over time. In 1550 merchant Jan de Pottre preceded his journal with maxims about the coincidence and correlation between the day of the week on which New Year's Day began, and weather conditions later in the year. By the eighteenth century, some chroniclers had developed their observations into the new science of phenology, that recorded the exact moment in any one year when natural phenomena (the first snowdrop) occur at a particular location. Scholars have argued that the weather gained a new cultural significance in the Enlightenment (Golinski, Jancovic), and we are interested to see whether we can see this reflected in Dutch chronicles, too.

Price-recording was also done by chroniclers of all walks of life. We frequently see authors comparing prices past and present. Sometimes they did so to express the extremity of conditions during sieges and famines, or as evidence for disorder. But seeking patterns in prices was always relevant. Early modern prices were in wild flux, not only because food prices were highest in early summer, when the last year's supplies ran out, but also because they were influenced by the volume of harvests and changes in demand, for instance because of armies passing through. Any information that could help predict changes was welcome.

Like in weather and prices, authors also looked for patterns in **death and disease**. Epidemics, especially, were often seasonal, and authors recorded numbers of deaths, for instance, to work out whether an epidemic was still growing, or receding (Amelang).

While the interest in prices, weather and disease explains why authors chose to structure their collections chronologically, other patterns were less obviously

associated with annual rhythms. Many authors saw a causal relationship between **human sin and divine retribution**, and included exemplary tales of the wages of sin. Recording good and bad deaths, the behaviour of people being executed, the decomposition of corpses, was a way of assessing the moral state of the dying.

In this way chroniclers not only made sense of the fate of individuals, but also of the political, religious and military **fate of local communities** as a whole. This was especially important during times of crisis. Looking back on the period 1740-1749, for instance, chronicler Cornelis Veen connected the death of the Emperor in Vienna, the outbreak of the war of the Spanish succession and the restoration of the stadholderate in Holland, with divine punishment, a comet, cold weather, high prices, an infestation with mice, and cattle disease in his own village of Zaandam.

We expect providential explanations for war and collective misfortune to have persisted, but changes in the way these were combined with natural and scientific factors, as well as moral and political ones. A key question is in what contexts chroniclers identified tradition or 'novelty' as a **cause**, or conversely as the **solution** for crisis phenomena. Many authors framed situations and behaviour by descriptors like 'old, traditional, customary' or 'new, unheard of, modern, novel, innovative'. More implicitly, authors indicated their **views of change** by phrases like 'these days', 'this was the beginning of' and 'this was the end of'. The context usually allows us to see whether this is meant positively or negatively, and to reconstruct in what circumstances this was so. In collaboration with the postdoc, the mining of expressions of *hesitancy*, *approval* and *disapproval* and their correlation with specific topics should also allow this researcher to detect attitudes to news and novelty.

III. New knowledge: tracking topics, attributions and speculations
(Postdoc)

This subproject will investigate computational methods for detecting the appropriation of new knowledge in chronicles. Semantic processing techniques will be applied to detect when and how the authors of chronicles explored and acquired new knowledge. The main research questions are:

- Which linguistic phenomena are indicative of the expression of novelty and appropriation of knowledge in early modern chronicles?
- How can these phenomena be identified automatically in passages of text where they occur?

The methods that will be applied (factuality profiling, attribution detection, modality and hedge processing, topic modelling) have already been tested for modern English and some of them have also been applied to Dutch.

An initial qualitative analysis and manual annotation of the data will serve as starting point to develop computational methods for detecting novelty, factuality, speculation and uncertainty using rule and lexicon-based techniques as well as machine-learning. Collaborating with the other team members, the postdoc will analyse a sample of chronicles from different periods and regions to study the following phenomena:

- The introduction of **new concepts**. New techniques and new ideas or concepts can in some cases be traced through keyword searches (e.g. *infection*, *thermometer*) and more advanced searches for co-occurrences of themes and concepts (cattle disease – god - sin, versus cattle disease – weather conditions – infection), yet they may also be hidden in other text properties like the ones listed below.
- The reference to and appreciation of **sources of information** in chronicles. Since the sources mentioned by the authors will have been manually annotated by the crowd, we can analyze the linguistic patterns in which these references occur (eg. *sayde men sterck*) in order to determine how the authors appreciated them in terms of reliability and authority.
- **Attribution** of information/knowledge to sources. Attribution is a discourse phenomenon used to establish a relation between a third party expressing an attitude and some text. Attribution relations consist of three components, a source, a cue and a content (Paretti). We expect that fragments of text that introduce new knowledge can be in part identified by analyzing the content of attribution relations.
- **Uncertainty and speculation**. Writers use specific linguistic expressions (hedges, modals) to indicate that they have doubts about the information that they or other sources are presenting (Morante and Sporleder). We expect that authors of chronicles use specific expressions (eg. *misschien, wellicht (maybe)*, *het zou kunnen zijn dat (it could be that))* when they are reporting about new knowledge.

Additionally, in order to find new topics, topic modelling techniques will be applied that allow us to extract topics, i.e. lists of words that co-occur (for example *cattle disease* and *infection*) in statistically meaningful ways without the need of annotated data. Since dates in the chronicles will be manually annotated, it will be possible to track how topics change over time and when a topic is newly introduced.

One of the challenges of this project is language diversity due to historical change (the chronicles cover a period of over three centuries), regional variety and a lack of spelling standardization. Because of this linguistic heterogeneity the data cannot be processed easily with available NLP tools. Retraining existing tools for different historical periods is not an option, due to the lack of annotated corpora. This problem has been addressed in previous work by normalizing older spelling variants of a language into its standardized modern variant (Archer et al,

Schneider et al., Tjong Kim Sang, Van Elburg and Wijckmans). In this project we will work with a new well performing workflow, PICCL (Reynaert et al. 2015), which incorporates the spelling normalization tool TiCCL (Reynaert 2005, 2010). This will be done in close collaboration with Martin Reynaert's CLARIAH/eScience project TICCLAT.

The evaluation of the semantic processing methods will be qualitative and performed in several iterations. The fragments of texts extracted with the tools will be provided to the PhD students so that they can evaluate whether the fragments are relevant. Based on their judgement, the tools will be fine-tuned. The normalization phase will be evaluated in coordination with the TICCLAT project. For this purpose, small samples of text from different periods will be selected to perform a manual evaluation.

The main outcomes of this subproject will be:

- Systematic analysis of linguistic cues that signal the introduction of new ideas.
- The adaptation of existing NLP methods in order to detect attribution, factuality, uncertainty and speculation in the normalized version of the historic variants of Dutch.
- A digitized corpus of chronicles normalized automatically into modern standardized Dutch spelling.
- A subset of the digitized chronicles including manual annotations of expressions referring to novel phenomena, sources, attributions and expressions of speculation.

The tools, automatic annotations and lexical resources developed will be made public.

IV.   Chronicling change.

By way of **synthesis** the applicants will write at least three articles,

(1)   *Chronicling novelty.* Using the results of the three subprojects the first article aims to answer the central question of this project, and reassess the relationship between the circulation of new ideas and practices in the Low Countries, their reception, and the impact on attitudes to novelty. It will do so in the context of the evolving discussion about the emergence of innovation as a positive value in European society, and the impact of this way of thinking on older ideas. Possible venues are *Continuity & Change* or the *Journal for the History of Ideas.*

(2)   *How to read chronicles*? With a view to further work, this article, probably co-authored with the postdoc, will evaluate the methodological side of the project, both in terms of its interdisciplinary

collaboration, and the gains of working with a larger corpus as opposed to working with single chronicles. We are particularly keen to evaluate how our expectations of its potential compare to the final results, and report on obstacles, as well as potential we had not anticipated. Venue may be the *Digital Humanities Quarterly.*

(3) Preliminary research points to the existence of thousands of chronicles across Europe, which offer the potential not only for a European comparison regarding the questions posed in this project, but also for future work in a range of other areas (e.g. on changing experiences of political authority, but also of identities, gender, space, emotions (Pollmann 2016, Rau 2017). We have already begun to establish an international network of experts in this field, some of whom have agreed to serve on our advisory board. For an international workshop towards the end of the program, we will invite colleagues to apply a number of the questions we have developed on knowledge circulation and innovation to one or more chronicles with which they are familiar in their own context, and, if so desired, work with some of the tools we have developed. While a selection of those papers will be published as a special issue of a journal (perhaps in the *Journal for early modern history*), we also want to write a separate article that further explores the suitability of our research method for cross-European comparisons, and the practicability of doing so, perhaps through sharing a protocol and tools.

Knowledge utilization

The corpus of chronicles has potential for many **scholarly purposes in other disciplines** than cultural history: e.g. environmental humanities, historical lexicography, the study of material culture, media studies, memory and religion. We will propose a Lorentz workshop in order to explore the interdisciplinary potential of the material and to set up new collaborations

The corpus of chronicles normalized and pre-processed with NLP tools is of great value for the digital humanities and computational linguistics communities. To begin with, corpora of this size and diversity of historic variants are very scarce, especially for Dutch. This corpus will provide a gold standard for regional diachronic variants of Dutch which will further help to fine-tune linguistic normalization by PICCL and to train new tools.

The additional layers of semantic annotation that will be provided with the corpus will allow the computational linguistics community to train new tools for the semantic processing of historical variants of Dutch and their normalized version. The corpus can be used not only for research purposes, but also for teaching purposes. Students can be taught how to process this type of corpora with hands-on assignments. Finally, the corpus can be used to organize international shared tasks on processing historic variants of languages. Within the

computational linguistics community shared tasks are organised to boost advances in a specific field.

The **general public** will both help to create, and benefit from, our corpus and the tools we build. At the end of the project about two hundred new texts will have been made available on the very popular Digitale Bibliotheek der Nederlandse Letterkunde (DBNL, 3 million visitors per annum) for all to use (see 11b).

Our transcription and manual annotation project builds on the extensive *crowd sourcing* expertise, volunteer network and infrastructure created by Picturae. Through their *VeleHanden* (Many Hands) platform, c. 12,000 volunteers are participating in all sorts of crowd sourcing initiatives, most of them in historical transcription, annotation and indexing projects by archives and cultural heritage institutions. Key to the success of a project running on *VeleHanden* is the effort made by the project leaders to create a sense of community and involvement with the volunteers by showing their own enthusiasm and commitment and frequent communication through the management of an online forum, and by sharing the goals and outcomes of the project as a whole. We will therefore organise a number of meetings for the volunteers we engage, to explain the purposes and progress of the programme.

Because **local history** is immensely popular in the Low Countries we also want to encourage further use of the transcriptions at a local level. In many local archives volunteers or **transcription groups** are currently active who publish the fruits of their labour through the websites of the archives. We will ask them to turn their attention to local chronicles, and assist in transcribing and enriching them at the platforms of Transkribus and/or Picturae's *VeleHanden*. In return, we will of course offer the enriched texts for publication on their websites, but also pay site visits to show the volunteers and others interested in local history how these can be used to work on questions in local history. We expect an especially enthusiastic uptake of the possibilities for local mapping. What were the crime hotspots? Where did people hear about news? When was food scarce? So as to encourage further use in local history associations, schools and archives, as a work placement project, we will ask students to make a video to demonstrate the potential of the collection and the tools, and invite two trainee history teachers to develop a teaching package for secondary schools.

11b. Description of the proposed infrastructural component

The infrastructural component will entail the creation of a searchable digitized corpus of about 200-220 Dutch language chronicles from the period 1500-1850. Previously edited chronicles contain an average of c. 200 printed pages source text. A corpus of 200-220 chronicles will thus amount to a minimum of 40.000 pages text.

Corpus selection

In the past ten years, the applicants created a list of chronicles with the help of many colleagues in Belgium as well as our Dutch colleagues and students. We have also made grateful use of the work of Dekker and Baggerman who published extensive bibliographies of pre-modern Dutch ego-documents in the past decades (http://www.egodocument.net/), some of which fall within our definition of chronicles. Since they included only those texts which also contain personal information about the author, there is only limited overlap between our lists.

For us, a *chronicle is a text resulting from an act of literacy by someone who decides that he is well suited to keep a record of events in his surroundings, who believes that these events are worth recording, and that the best way to structure this information is to do so chronologically*. This could be done retrospectively, yet the early modern authors who interest us here are those who tried to record the events of their own lifetime; those who wrote what Germans call *Zeitgeschichte.*

Accordingly, our selection criteria for chronicles for this project are:

1. their strictly chronological organization – distinguishing them from common place books used among the highly educated

2. they describe a period that may start in an (ancient) past but that evolves into an account of events occurring during the lifetime of the author

3. focus on public life in the local community of the author. The main distinction with ego-documents is that in chronicles historical or public events and collective experience are the main topic, not the person or autobiography of the author.

For purely practical reasons, we have chosen not to include texts written in French, Latin or Frisian. If the NLP procedures and tools we develop for analysis work well for the Dutch-language material, they can be tested on other languages in future research.

While these texts were not written for publication, some of them have been edited and published in print, mostly between 1850-1950, in editions of varying quality. Unpublished chronicles can be found in virtually every archive in the Netherlands, and carry a variety of titles and descriptions. We have so far identified 108 that seem suitable for our purpose.

The corpus will thus be drawn from:

1 Modern digital transcriptions or high quality digitized editions of chronicles (including 25 published chronicles, digitized and published by DBNL); making a total of c.40

2 Published editions of chronicles that are partly scanned by Google Books, but not available in high quality OCR; so far, we made a selection of 75 texts that met our criteria concerning the quality and completeness of the editions.

3 Unpublished manuscripts that are deposited in archives and library collections in the Netherlands and Belgium. Our list of 108 items still needs to be evaluated and completed. We will try to select for regions and years otherwise unrepresented. We estimate that the final selection may amount to 120 manuscripts or more.

The aim of the project is to create a corpus of high quality TEI XML annotated text files of all three sets of chronicles that will be published and preserved on the DBNL website by the KB and simultaneously made available for further processing, enrichment and analysis by the research team.

In the image below the workflow for each category is sketched.
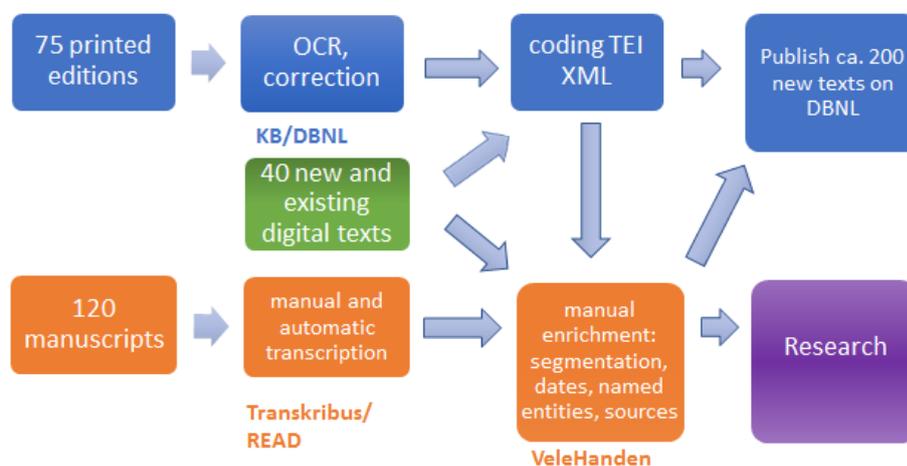
Digitization of printed editions

A selection of c.75 printed editions of chronicles will be digitized by the online library *De Bibliotheek der Nederlandse Letteren* (DBNL) that is part of the Dutch Royal Library (KB). Collaboration with the KB is secured and they provided us with an estimate of the costs of digitization as well as publication of the entire corpus after enrichment. The Royal Library already possesses scans of most of the publications we selected, a number of 29 books need to be scanned first. DBNL has a well-established pipeline for digitization that produces high quality

TEI XML documents based on manually corrected and coded OCRs of the scans. Their standard codes include all bibliographical metadata as well as the coding of the structural elements of the text (such as headings, tables, images, para-text, footnotes etc.)

A total number of 25 chronicles is already digitized and published in DBNL. A set of transcriptions published on websites by archives and groups of amateur historians will be equally coded and converted to XML by DBNL.

Manuscript transcription

Transcripts of the manuscripts will be produced with the help of automatic HTR (Handwritten Text Recognition) via the platform *Transkribus* set up by the European Horizon 2020 funded READ Project (Recognition and Enrichment of Archival Documents). Collaboration with this group is already discussed and ensured through email contact.

Per group of manuscripts with a similar handwriting, (e.g. 16th-17th century gothic style and 18th century humanistic style) a representative selection of 300 pages will be transcribed by the applicants with the help of a number of expert volunteers making use of the applications provided by Transkribus. This data set will serve as training material ("ground truth") for the HTR engine that is trained by Tanskribus at the University of Rostock and will also be used to evaluate the results of the HTR. After a number of corrections / iterations the engine may produce automatic transcriptions with character error rates below 10%. (https://www.citlab.uni-rostock.de/work/) The results for this project seem very promising as Transkribus already processed a substantial amount of early modern Dutch language text material. Moreover, Transkribus provides exact ways to measure the quality of the automated HTR processing. In 2018 they plan to implement a tool for measuring precision and recall.
Once the model produces acceptable results it can be run over the whole corpus, an estimated 20-30.000 pages. Transkribus allows for the exportation of the data in various formats including the original image files, TEI XML files and the metadata added to the document.

Correction, segmentation and annotation

The final corrections of the transcripts as well as the enrichment of the corpus generated by both DBNL and the HTR engine will be carried out through crowdsourcing. For this phase we will make use of the services of the very successful crowdsourcing platform *VeleHanden,* developed by the private company Picturae.

VeleHanden will develop a new interface (for this and future projects) in which the scans of manuscripts and TEI encoded XML documents produced by DBNL and the automatic HTR can be shown side by side to the users. The volunteers

will be asked to read and correct the transcripts. A number of complex or otherwise difficult hands will need manual transcription by specialized volunteers.

The volunteers will also be asked to identify and code text segments (e.g. distinguish main text from headings, margin texts, additional notes, lists, illustrations, inserted documents etc.) and to select and annotate the **dates** of entries, **named entities** (mainly persons and locations) and **sources of information** adding tags such as 'newspaper', 'rumour/hearsay', 'eye witness' etc.. Additionally, place names will be tagged by selecting modern geo-names from a list. Both the geo-tagging and the tagging of dates with modern dates may require the expert knowledge of the team members as the spelling of place names and the use of historical calendars.

The quality of the volunteer work in this type of projects is usually very high. Moreover, every page will be corrected and annotated twice. The team members will check and disambiguate the differences in outcomes and authorize the final versions of the transcripts and annotations.

The project will be designed by VeleHanden in close collaboration with the applicants. A student assistant and, in due course, the PhD students, will be involved in the monitoring of the project. On a project page the project will be presented and instructions provided for the volunteers. A forum allows for communication with the volunteers, questions and answers.

VeleHanden works with a system of remuneration based on credit points that can be earned for transcription or annotation work. In exchange for credit points we will invite volunteers to lectures and workshops.

VeleHanden will generate an output in TEI XML. They have agreed to collaborate with DBNL while developing the project, in order to use the desired codes for e.g. the segmentation of the text. DBNL will further complement and process the enriched data for publication on their website.

The annotated transcriptions of manuscripts will also be made available to the archives where they come from. Archives may publish them as they wish. We will provide documentation on how the documents can be made searchable as well as visualized.

Linguistic processing of the corpus

On top of the digitized and manually enriched corpus, automatic processing tools will be applied in order to link the older spelling and form variants to their modern Dutch counterparts. This normalisation process enables the use of NLP tools developed for modern Dutch. For example, automatic part-of-speech tagging and syntactic parsing tools can then be applied.

The results of these linguistic processing tools will be made available as an extra layer of annotations in combination with the digitized corpus. Manually annotated

samples of the data will also be provided in order to enable evaluating the results.

Data curation

The corpus, including the transcriptions, meta-data, manual and automatic annotations and documentation will be made publicly available for future use under the creative commons license CC 4.0. All data will be stored on GitHub. Additionally, the data will be archived by one of the existing Dutch data preservation centers such as DANS (https://dans.knaw.nl/en/about) or the TST-centrale (http://tst-centrale.org)

Summary of the infrastructural component in work packages

WP	Who	Tasks
1 Digitization Published chronicles	KB/DBNL	Collection of scans/scanning published chronicles OCR & OCR correction TEI XML coding Check and handle copyright issues Transfer coded files to Picturae for further annotation
2 Collection and selection manuscripts	Applicants, student assistant, PhDs	Extend database/inventory relevant manuscripts Complete metadata Contact archives, order scans Administration incoming scans and metadata Select scans for transcription in Transkribus
3 Transcription & training HTR	Volunteers, student assistant, Transkribus/REE	Create project, upload scans in Transkribus Segment images, link text to image Transcribe and correct texts Train HTR engine Run engine corpus, evaluate, iterate Export data
4 Correction and manual annotation of the transcripts	Picturae, applicants, PhDs, volunteers, student assistant	Develop and test crowdsourcing environment Upload scans and data Organize workshops volunteers and archives Manage forum of volunteers and section Q and A's Supervise transcription, corrections and annotation Return scans with transcriptions to archives Transfer output to DBNL
5 Evaluating the quality of the digitized material	PostDoc, PhDs, Student assistant	Evaluation of the quality of the segmentation of the text. Evaluation of the coding of named entities, dates and sources of information Evaluation of geotagging

		Corrections
6 Normalization of texts into modern Dutch	PostDoc, PhDs, Student assistant	Normalization into modern Dutch using the PICCL workflow Several iterations will be planned for manual evaluation (student assistant) of the PICCL output. Errors detected in each iteration can be fed back into PICCL. Quality will be checked by determining whether the texts can be processed with an NLP pipeline of modern Dutch Final evaluation of quality
7 Pre-processing the text with NLP tools for modern Dutch	PostDoc	Setting up an NLP pipeline using existing resources: POS tagging, lemmatization, shallow parsing Evaluating quality of output in terms of suitability for project use
8 Data management and publication	KB/DBNL & team	Royal Library publishes transcriptions and scans of c. 200 chronicles in DBNL with metadata, search tools and download options. Postdoc and PhDs deposit enriched data sets (1 in the original language, the second in normalised modern Dutch) in Folia/XML in repository DANS with documentation and tools Transcripts of the manuscript will be made available for the archives with documentation, tools and instructions

Planning infrastructural Component

	2018	2019				2020				2021
	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1
WP1 Digitization Published chronicles										
WP2 Collection and selection manuscripts										
WP3 Transcription & training HTR										
WP4 Correction and manual annotation										
WP5 Evaluation										
WP6 Translation into modern Dutch										
WP7 Pre-processing (NLP)										
WP8 Data management and publication										
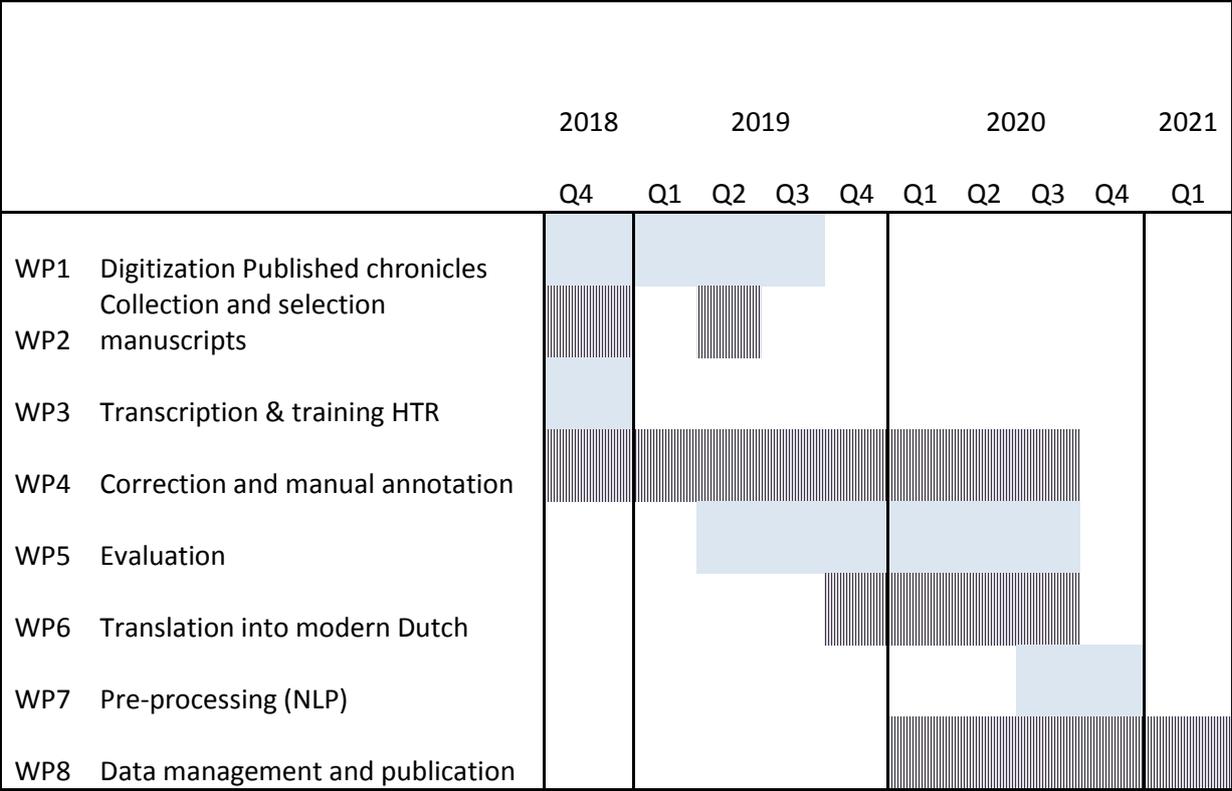
Viability

We have discussed and agreed financial costs, effort and planning with all three **external partners**: The Royal Library, Transkribus and Picturae/VeleHanden. The Royal Library and Picturae provided us with estimates of the costs, the services by Transkribus are free as the project is European funded and additionally funded by several projects running until the end of 2021. (https://read.transkribus.eu/transkribus/)

Technically, the project is innovative – both the use of automatic transcription and the new type of crowdsourcing project that will be developed by Picturae - but the risks for the researchers are limited as the Royal library will certainly be able to provide us with a base corpus of c. 100 texts within half a year after the start of the project. This enables the PhD students and postdoc to start familiarizing themselves with the data and to develop their methods, while simultaneously collaborating in the supervision of the crowdsourcing and annotating process.

The experiences with both the speed and the quality of transcriptions of historical material at the platform VeleHanden are very positive. An example of a current project that is doing extremely well is *Alle Amsterdamse Akten*

(http://alleamsterdamseakten.nl/) in which currently 535 volunteers are participating of which quite an impressive number are familiar with seventeenth century palaeography.

We will recruit volunteers from local groups of history amateurs who are active in many towns in Belgium and the Netherlands. Via local archives and with the help of a student assistant we will contact these groups and organize meetings during which we will explain our project and invite them to collaborate.

We need **student assistance** to maintain contacts with these groups, manage the forum of the project on VeleHanden and help with the administration of files and versions, corrections and evaluations.

A crucial milestone of the project is to obtain a version of the corpus normalized into modern Dutch. TICCL has been used extensively in large scale corpus projects of modern (Oostdijk et al. 2013) and historical (Brugman et al. 2016) Dutch texts. Its performance on modernizing historic texts has been demonstrated on Portuguese, but the more accurately dated Dutch historical word forms in the lexicons to be derived from the Nederlab corpus in project TICCLAT are guaranteed to further boost its performance on historical Dutch.

To make sure that this step is successful, we will work in close collaboration with Martin Reynaert, who is an expert on normalization of Dutch and author of PICCL, the tool that we will use. Members of the CLTL group have worked with Martin in the past and keep close professional relations.

We estimate that the application of NLP methods to the corpus once it has been normalized is realistic because the methods and tools have been tested in previous work, even if for English and non-historical Dutch. The work will mostly consist of adapting the tools by building new lexicons and updating rules based on syntactic knowledge.

National relevance of the corpus and knowledge utilization (see also 11a)

On DESIDERIA, the Royal Netherlands Academy of Arts and Sciences roadmap for the Humanities to 2025, chronicles are included because they meet the criteria of underrepresented original hand-written, non-canonical and non-fiction pre-modern material. They are considered of prime importance to historical linguists, and literary scholars as well as historians. This was not always so. While medieval chronicles have been very well studied as a genre, and for the Netherlands have been digitally available for many years now (http://www.narrative-sources.be/colofon_nl.php), early modern chronicles have only recently been rediscovered as an important resource. Falling between the genres of history and autobiography, chronicles of this type have traditionally been used as a source for early modern local history only. Nevertheless, German

and Swiss scholars have since the 1950s been aware that chronicles are a good lens through which to study urban culture, social struggles and political identity, and in recent decades they have used them to gauge experiences of religious division and the Thirty Years war. In the process, they have also considered the early modern history of the genre, broadening the traditional assumption that these texts should be seen as a form of historiography to thinking of them also as a form of 'memory', especially of urban communities (Rau, Fuchs). Historians of the early modern Low Countries have begun to use chronicles increasingly to study the local experience of political and military change (Pollmann 2011 & 2016, Deseure, Van Nierop). They argue that chronicles give us a very valuable insight in the everyday experiences of life in historical urban and village communities.

This is all the more important for understanding the history of the Low Countries, since much power there was devolved to a local level (notoriously so in the Dutch Republic, but also in the Habsburg Southern Netherlands), and people cherished local and regional privileges and differences. It was only after 1795 that they became more integrated politically – a process which can itself be traced and followed in chronicle texts. Historical linguists, on their part, are interested in chronicles because they give access to a historical linguistic variety that was 'filtered out' by professional printers, proofreaders and editors. For literary scholars, they offer vital access to reading and writing practices beyond the canonical authors (Blaak).

So far, these texts have only been available for close reading. By making them available in a standardized format, we also open them up for a whole range of distant reading purposes. As pointed out above, the corpus of chronicles normalized and pre-processed with NLP tools is also of great value for the digital humanities and computational linguistics communities. To begin with, corpora of this size and diversity of historic variants are very scarce, especially for Dutch. Such a corpus will allow to make progress in processing historic variants of Dutch not only because it will be used to improve linguistic normalization tools such as PICCL, but also because it will allow to train new tools. The additional layers of semantic annotation that will be provided with the corpus will allow the computational linguistics community to train new tools for the semantic processing of historical variants of Dutch and their normalized version. The corpus can be used for research purposes, as well as for teaching purposes. Students can be taught how to process this type of corpora with hands-on assignments. Finally, the corpus can be used to organize international shared tasks on processing historic variants of languages.

12. Word count

11a: General:1992
11a: Subprojects & Synthesis:674+795+784+389

11a: Knowledge Utilization: 602
11b: Infrastructure: 3107
 Total: 8343

13. Summary in key words:

knowledge, innovation, chronicles, media, reasoning

14. Work programme

Milestones

		2018	2019	2020	2021	2022	2023
Q1			Kick off meeting with volunteers, launch project	Final meeting with volunteers		meeting 2 advisory board	
Q2			Start PhD's and postdoc	meeting 1 advisory board	International conference	Manuscript Conference volume completed	2 Dissertations
Q3			Upload second batch of scans	Completion corpus			
Q4		Kick off 1 september 2018, start infrastructural project		Interdisciplinary workshop semantic analysis of historical corpora			3 synthesizing articles

Work plan Applicants
(Replacement 2023 Q3-4)

2018 Q3-4	Recruit team Data collection and selection (WP2)
2019	Supervise PhDs Organize team meetings Participate in WP4: workshops for volunteers
2020	Supervise PhDs Organize team meetings Organize meeting with Advisory board
2021	Supervise PhDs Organize team meetings Organize international workshop Write article conference volume/special issue Edit volume/special issue
2022	Supervise PhDs Organize team meetings Organize second meeting with advisory board
2023 Q3-4	Write 3 synthetic articles with co-applicant

Work plan co-Applicant
(Replacement 2023 Q3-4)

2018 Q3-4	Recruit team Data collection and selection (WP2) with student assistant Supervision transcription project volunteers and student assistant with Transkribus Project design with picturae/Velehanden
2019	Co-supervise PhDs Coordinate WP4
2020	Co-supervise PhDs WP4-WP8 collaboration/supervision Meeting with Advisory board
2021	Supervise PhDs Co-organize international workshop Write article conference volume/special issue Edit volume/special issue
2022	Co-supervise PhDs Second meeting with advisory board
2023 Q3-4	Write 3 articles with applicant

Work plan PhD project 1 **Changing mediascapes and the collection of knowledge** (1,0 fte 1 April 2019- 31 March 2023)

2019 Q2-4	Background reading, framing questions Data collection (second batch WP2) Participate in WP4 supervision crowdsourcing, workshops for volunteers Evaluation annotations, data curation (WP5) IT training (basic programming and GIS visualization) (c. 20 ects) 0.2 fte x 4 months teaching
2020	Research and experimentation data visualization Collaborate in WP 4-8 Present first paper and research plan to advisory board Education in (National) Research School Draft chapter 1 and 2 0.2 fte x 4 months teaching
2021	Research Co-organize international workshop and present paper Draft chapter 3 and 4 Write article conference volume/special issue
2022	Second meeting with advisory board Education in (National) Research School Present paper or poster at a Digital Humanities conference Write Chapter 5, Introduction and Conclusion Start composing and revising thesis
2023 Q1	Revise and submit thesis

Work plan PhD project 2 **Reasoning, causality and the appreciation of novelty** (1,0 fte 1 April 2019- 31 March 2023)

2019 Q2-4	Background reading, framing questions Data collection (second batch WP2) Participate in WP4 supervision crowdsourcing, workshops for volunteers Evaluation annotations, data curation (WP5) IT training (basic programming automatic text analysis techniques) (c. 20 ects) 0.2 fte x 4 months teaching
2020	Research and experimentation with data analysis Collaborate in WP 4-8 Present first paper and research plan to advisory board Education in (National) Research School Draft chapter 1 and 2 0.2 fte x 4 months teaching
2021	Research Co-organize international workshop and present paper Draft chapter 3 and 4 Write article conference volume/special issue
2022	Second meeting with advisory board Education in (National) Research School Present paper or poster at a Digital Humanities conference Write chapter 5, Introduction and Conclusion Start composing and revising thesis
2023 Q1	Revise and submit thesis

Work plan Postdoc: **New knowledge: tracking topics, attributions and speculations** (0,5 fte 1 April 2019- 31 Sept 2022)

2019 Q2-4	Coordinate evaluation annotations, data curation (WP5) Develop and evaluate method normalization corpus in modern Dutch (WP6) Research design Collaborate with PhDs on methodological issues
2020	Submit proposal Lorenz workshop Normalization of corpus into modern Dutch and evaluation iterations (WP6) Work on NLP normalized corpus: pre-processing (WP7) Data management and publication (WP8) Write two articles/conference papers First meeting advisory board Qualitative analysis of data for topic modelling, attribution and uncertainty (I)
2021	Qualitative analysis of data for topic modelling, attribution and uncertainty (II) Produce guidelines for manual annotation of novel phenomena, attribution and uncertainty Monitor annotation process for novel phenomena, attribution and uncertainty Work on NLP normalized corpus: topic modelling (Q1, Q2), attribution processing tool (Q3, Q4) Provide data for qualitative evaluation; evaluate NLP output

	Write two articles/conference papers Organize interdisciplinary (Lorenz) workshop
2022 Q 1-3	Work on NLP normalized corpus: processing uncertainty (Q1,Q2) Provide data for qualitative evaluation; evaluate NLP output Second meeting with advisory board Write one article/conference paper

15. Planned deliverables:

Applicants:

1 conference volume/special journal issue of e.g. *Journal for early modern history* edited by applicants 2022

1 peer reviewed article on the key conclusions of the project on the chronicling of novelty in *Continuity & Change* or the *Journal for the History of Ideas.*

1 or 2 peer reviewed articles evaluation of the method, co-authored with the Postdoc in e.g. *Digital Humanities Quarterly.*

1 peer reviewed article on the feasibility and promises of using our tools and approach for European comparison in a special issue of *Journal for early modern history* or conference volume edited by the applicants

PhD student 1

1 PhD dissertation, submitted 1 March 2023

1 peer reviewed article by PhD student in English, e.g. in the conference volume or special issue

1 (poster) presentation of the spatial analysis of media use and knowledge horizons in this project at a DH workshop or conference such as DH Benelux

PhD student 2

1 PhD dissertation, submitted 1 March 2023

1 peer reviewed article by PhD student in English, e.g. in the conference volume or special issue

1 (poster) presentation on distant reading techniques used in a DH workshop or conference such as DH Benelux

Postdoc:

5 articles/published conference papers by postdoc as main author on the methods developed in the field of Natural Language Processing (on the normalization of historical Dutch) and on knowledge attribution to the chroniclers' sources as well as on the analysis of uncertainty and speculation about sources and new knowledge. These articles could appear in *Digital Humanities Quarterly* or *DSH*

Team:

> In collaboration with the Royal Library: Publication of c. 200 chronicles in DBNL with search tools and download options.

> Enriched data sets (first in the original language, the second in normalised modern Dutch) in Folia/XML in repository DANS for future use and possible integration in Nederlab/CLARIAH infrastructure.

> Transcripts of the manuscript will be made available to the originating archives.

> Tools, instruction-video and teaching package on how to search and visualize chronicles. These will be developed in collaboration with students in the context of an internship as part of the educational Master in History.

16. Short curriculum vitae applicants

Judith Pollmann (1964) is Professor of Early Modern Dutch History at Leiden University and Academic Director of its Institute for History. She has published widely on the history of the Dutch Revolt, on news and propaganda, and on the experience and impact of change in early modern Europe. Much of her work makes use of diaries, memoirs and chronicles. When working on her most recent book, *Memory in early modern Europe, 1500-1800* (Oxford 2017) she developed an interest in modernity, the role of custom and attitudes towards novelty and innovation in early modern Europe.

Judith Pollmann studied history at the University of Amsterdam and the Warburg Institute in London. Her UvA doctoral dissertation (1998) was awarded the Keetje Hodshon Prize of the Holland Academy of Science. Before coming to Leiden in 2005, she taught early modern history at Oxford University, where she was a CUF lecturer and Fellow of Somerville College from 1998-2005. A VICI laureate, from 2008-2013 she directed the NWO project *Tales of the Revolt. Memory, oblivion and identity in the Low Countries, 1580-1700*. Since 2015 she has been directing, with Henk te Velde, the Vrije Competitie NWO project *The persistence of civic identities in the Netherlands, 1747-1848.* She has successfully supervised five PhD students, and expects to see another three graduate in 2018.

Pollmann likes bringing early modern Dutch history to a larger audience in the Netherlands, most recently by advising the Dordts Museum and the Rijksmuseum on new exhibitions, and the NTR on a 2018 television series on the Dutch Revolt. Nationally, she is active as the chair of the board of the Huizinga Institute for Cultural History, as chair of the *Werkgroep Zeventiende eeuw* and as a member of the Royal Library's advisory board for the Metamorfoze program to digitise archives. Internationally, she is a member of the editorial board of *Past &*

Present and a Fellow of the Royal Historical Society, as well as a member of the academic board of the Leibniz Institut für Europäische Geschichte in Mainz.

From 2008-13, Pollmann directed a VICI project on early modern war memories, in which she collaborated and published extensively with co-applicant Erika Kuijpers.

Erika Kuijpers (1967) studied social and economic history at the University of Amsterdam. Her doctoral dissertation (University of Utrecht 2005) on Migrants and social relations in seventeenth century Amsterdam was awarded with the Keetje Hodshon Prize of the Holland Academy of Science (2007). Her most recent work is on early modern war memory, on the practice of chronicling and on the history of emotions. She is one of the organizers of the *Amsterdam Centre for Cross-Disciplinary Emotion and Sensory Studies.* In that capacity she co-ordinated a pioneering digital humanities project entitled 'Embodied Emotions', that aimed to map the bodily expression of emotions in a corpus of 300 seventeenth- and eighteenth-century Dutch plays. She teaches medieval and early modern cultural history at the Vrije Universiteit in Amsterdam.

17. Key publications applicants:

- Judith Pollmann, *Memory in early modern Europe, 1500-1800* (Oxford University Press 2017)
- Judith Pollmann, 'Archiving the present and chronicling for the future in early modern Europe', in; *The Social History of the Archive. Record Keeping in Early Modern Europe*, ed. Liesbeth Corens, Kate Peters and Alexandra Walsham, Past & Present Supplement (2016), 231-252
- Judith Pollmann, 'Iconoclasts Anonymous, or why did it take historians so long to identify the image-breakers of 1566?', *BMGN* 131/1 (2016), 155-176
- Judith Pollmann and Erika Kuijpers, 'Introduction. On the early modernity of modern memory' in Erika Kuijpers et al. (eds.), *Memory before modernity. Practices of memory in early modern Europe* (Brill: Leiden and Boston 2013), 1-23.
- Judith Pollmann, 'Of living legends and authentic tales. How to get remembered in early modern Europe' *Transactions of the Royal Historical Society*, sixth series 23 (2013), 103-125.

- Erika Kuijpers, 'Expressions of fear, counting the loss. Managing emotions in war chronicles in the Netherlands (1568-1648)' Jennifer Spinks and Charles Zika, eds, *Disaster, death and the emotions in the shadow of the Apocalypse, 1400-1700* (Basingstoke: Palgrave 2016) 93-111
- Erika Kuijpers and Cornelis van der Haven eds., *Battlefield emotions 1500-1850. Experience, practices, imagination* (Basingstoke: Palgrave 2016)
- Janneke Van Der Zwaan, Inger Leemans, Erika Kuijpers and Isa Maks,

'HEEM, a Complex Model for Mining Emotions in Historical Text' (IEEE eScience, Munich 2015) DOI: 10.1109/eScience.2015.18

- Erika Kuijpers, 'O, Lord, save us from shame'. Narratives of emotions in convent chronicles by female authors during the Dutch Revolt, 1566–1635' in: Susan Broomhall (ed.) *Destroying order, structuring disorder. Gender and emotions in Medieval and Early Modern Europe* (Farnham: Ashgate/Gower 2015), pp. 127-146
- Erika Kuijpers, 'Between storytelling and patriotic scripture: the memory brokers of the Dutch Revolt' in: Erika Kuijpers et al. (eds.) *Memory Before Modernity. Practices of Memory in Early Modern Europe* (Leiden/Boston Brill 2013)

18. Data management

We will generate a high quality annotated corpus of historical Dutch texts from the period 1500-1850 as well as a normalized version using appropriate international standards and conventions for text encoding (TEI XML/Folia) ([https://github.com/proycon/folia)](https://github.com/proycon/folia). We will keep track of provenance and versions of the data throughout the various curation and processing cycles and provide documentation on how the corpus was selected, transcribed and processed in compliance with disciplinary and ethical conventions.

The corpus, including the transcriptions, meta-data, manual and automatic annotations and documentation will be made publicly available for future use under the creative commons license CC 4.0. All data will be stored on GitHub. Additionally, the data will be archived by one of the existing Dutch data preservation centers such as DANS (https://dans.knaw.nl/en/about) or the TST-centrale (http://tst-centrale.org)

During the project, data will be stored, annotated and managed on a number of servers. The Royal Library will store the scans of published chronicles. Scans of manuscript chronicles will be collected and uploaded in an existing VRE (Virtual Research Environment) at Leiden University. During the crowdsourcing phase, the images will also be uploaded at the Transkribus platform as well as at the VeleHanden project managed by Picturae. In the budget by Picturae the costs of storage are included.

Publication for a wider audience of the scans and transcripts and management of the metadata will be handled by the Royal Library/DBNL to allow access to the wider public. DBNL offers several download options and search tools and aims to further develop this infrastructure in the future. The Royal Library will also take care of the copyrights issues that may arise with a small number of published chronicles. The transcripts of manuscripts will be free of any copyrights and shall also be shared with local archives and interested parties (amateur historians) for their own use.

19. Public summary

Nieuwe kennis in de Lage Landen, 1500-1850.
Voor innovatie heb je meer nodig dan knappe koppen. Nieuwe kennis moet ook worden geaccepteerd. Tussen 1650 en 1850 ontstond er heel veel nieuwe kennis. Wij onderzoeken hoe die kennis bij gewone mensen in de Lage Landen terechtkwam, en of die daardoor positiever over innovatie gingen denken dan hun voorouders.

(50 words)

20. Summary for non-specialists

Een centrale vraag in de kennisgeschiedenis is onder welke omstandigheden nieuwe ideeën, kennis en technologie aanslaan en leiden tot brede acceptatie en succesvolle innovaties. Waarom is de ene samenleving innovatiever dan de andere? Inmiddels weten we dat culturele factoren sterk mee bepalen of nieuwe kennis of technologie aanslaat. Maar wat zijn dan precies de culturele voorwaarden daarvoor? En hoe komen die tot stand? Historisch onderzoek stelt ons in staat te analyseren hoe culturele factoren op de langere termijn innovatie bevorderen of tegenhouden.

Recente studies veronderstellen dat door culturele ontwikkelingen, zoals de opkomst van nieuwe media, Europeanen na 1650 meer ontvankelijk werden voor nieuwe technologie en innovatie dan hun voorouders en zo de Verlichting en de Industriële Revolutie mogelijk hebben gemaakt. Het onderzoek naar deze ontwikkelingen heeft zich echter vooral gericht op de producenten en professionele gebruikers van kennis en nieuwe technologie. Waar we veel minder van weten is hoe en wanneer gewone mensen in Europa zich die ideeën en kennis toe-eigenden, zodat die daadwerkelijk konden leiden tot brede culturele verandering.

Onderzoek naar de veranderende belevingswereld van een meer algemeen publiek in de vroegmoderne periode is heel lastig omdat de bronnen die we daarover hebben onderling vaak slecht vergelijkbaar zijn. In dit project willen we daarom een omvangrijk corpus ontsluiten van een type teksten dat wel in grote aantallen en over een heel lange termijn is geproduceerd in heel Europa: particuliere kronieken. Zo'n kroniek is een tekst waarin de auteur memorabele zaken en gebeurtenissen uit de wereld om hem heen chronologisch geordend optekent. In onze periode zijn de auteurs meestal mannen uit de stedelijke middenklasse.

Voor dit project willen we ca. 200 van dergelijke teksten uit het Nederlands taalgebied uit de periode 1500-1850 – ongeveer 40.000 pagina's tekst – digitaliseren om ze te kunnen doorzoeken, vergelijken en analyseren. Op die manier kunnen we de receptie, toe-eigening en waardering van nieuwe kennis op

een meer systematische, vergelijkbare en controleerbare manier onderzoeken. Een digitaal corpus stelt ons bovendien in staat om naast de grondige bestudering van specifieke teksten en auteurs in hun lokale context, via digitale methoden ook trends te traceren in het hele corpus, dus over de langere termijn.

Het project zal worden uitgevoerd in vier deelprojecten:

I.   Een AIO onderzoekt hoe kroniekschrijvers kennis verzamelden, selecteerden, ordenden en waardeerden in een veranderend medialandschap. Zowel variatie in tijd als ruimte wordt hierbij (letterlijk) in kaart gebracht. Daarbij staan de volgende vragen centraal: Welke bronnen van informatie (mondeling, manuscript, boeken, kranten) gebruikten zij? Werd hun wereld groter naarmate er meer en andere media beschikbaar kwamen? Welk gezag werd aan de verschillende bronnen van informatie en kennis toegekend? Veranderde de waardering voor traditionele bronnen door de opkomst van nieuwe media? Kregen de kroniekschrijvers ook last van informatie *overload*?

II.   De tweede AIO doet onderzoek naar de impact van nieuwe kennis op de manier van denken over oorzaak en gevolg van natuurlijke en maatschappelijke verschijnselen en naar de waardering van nieuwe ideeën en technologie. Vragen die hierbij centraal staan zijn: Hoe veranderde nieuwe kennis de manier van denken en waarnemen van de kroniekschrijvers? Gingen ze andere causale relaties leggen tussen hun observaties? Hoe waardeerden zij oude en nieuwe kennis? Verminderde de weerstand tegen nieuwe ideeën? Dit onderzoek zal ook gebruik maken van computationele technieken om uitingen van positieve en negatieve waardering voor nieuwe kennis en innovatie te traceren.

III.   De postdoc, computationele linguïstiek, onderzoekt methoden om *markers* van verandering en nieuwe kennis te vinden in het corpus. Sommige van die markers voor innovatie kunnen worden gevonden door te zoeken naar nieuwe concepten (bijv. infectie, thermometer) en meer geavanceerde zoekopdrachten waarbij veranderende correlaties rond bepaalde thema's in kaart worden gebracht (bv. epidemieën - zonde, versus epidemieën - weersomstandigheden - infectie), maar ze kunnen ook verborgen zijn in andere teksteigenschappen zoals uitdrukkingen van onzekerheid en speculatie.

IV.   De synthese door de aanvragers zal zijn beslag krijgen in een drietal publicaties. Het eerste zal de patronen analyseren van bottom-up toe-eigening van nieuwe kennis in de vroegmoderne Nederlanden en analyseren welke rol oudere denkkaders hadden bij de evaluatie, afwijzing of acceptatie van nieuwe kennis. Het tweede artikel zal verslag doen van de instrumenten en methodes die we hebben ontwikkeld voor dit onderzoek. En het derde zal het potentieel onderzoeken om deze methode te gebruiken voor een Europees vergelijkend onderzoek.

Kronieken zijn erg aantrekkelijke bronnen, die door hun lokale en verhalende karakter kunnen rekenen op de belangstelling van een groot publiek. Zowel in Nederland als in België is lokale geschiedenis zeer populair. De ervaringen met *crowdsourcing* leren dat deze liefhebbers graag meewerken aan de ontsluiting van lokale bronnen. In ruil daarvoor zullen wij workshops en lezingen geven waarin wij laten zien hoe die nieuwe kennis ook lokaal kan worden benut. Ook wetenschappers in andere disciplines kunnen met ons corpus aan de slag; vooral de historische taalkundigen zullen er veel aan hebben.

(798 words)

21. Budget

See excel file submitted with application

Bibliography:

- Amelang, J. 1998. *The flight of Icarus. Artisan autobiography in early modern Europe*, New Haven
- Archer, Dawn, Merja Kyto, Alistair Baron, and Paul Rayson. 2015. 'Guidelines for normalising Early Modern English corpora: Decisions and justifications', ICAME Journal. DOI: 10.1515/icame-2015-0001
- Blaak, Jeroen, 2009, *Literacy in everyday life. Reading and writing in early modern Dutch diaries*, Leiden
- Blair, Ann M. 2010. *Too much to know. Managing scholarly information before the Modern Age*, New Haven
- Brezina, Vaclav, Tony McEnery, and Stephen Wattam. 2015. 'Collocations in Context: A New Perspective on Collocation Networks'. *International Journal of Corpus Linguistics* 20 (2): 139–73
- Bruaene, Anne-Laure van. 1998. *De Gentse memorieboeken als spiegel van stedelijk historisch bewustzijn (14de tot 16de eeuw)*, Ghent
- Hennie Brugman, Martin Reynaert, Nicoline van der Sijs, René van Stipriaan, Erik Tjong Kim Sang, and Aantal van den Bosch. 2016 'Nederlab: Towards a Single Portal and Research Environment for Diachronic Dutch Text Corpora' In: Nicoletta Calzolari et al., eds., *Proceedings of the Tenth International Conference on Language Resources and Evaluation* (LREC-2016), pages 1277–1281, Portorôz, Slovenia
- Burke, Peter. 2000. *A social history of knowledge, from Gutenberg to Diderot*, Cambridge
- Burke, Peter. 2012. *A social history of knowledge from the Encyclopédie to Wikipedia*, Cambridge
- Davids, Karel. 2008. *The rise and decline of Dutch technological leadership. Technology, economy and culture in the Netherlands*, 1350-1800, 2 vols. Leiden
- Deseure, Brecht. 2010, 'Den ouden luijster is verdwenen. Geschiedenis, herinnering en verlies bij Jan Baptist van der Straelen (1792-1817), *Belgisch Tijdschrift voor Nieuwste Geschiedenis* 40.4
- Draelants, Isabelle. 1998. 'Le temps dans les textes historiographiques du Moyen Age' in Claude Thomasset en Joelle Ducos (eds.), *Le temps qu'il fait au Moyen Âge. Phénomènes atmosphériques dans la littérature, la pensée scientifique et religieuse*, Paris, pp. 91-138
- Dupré, Sven, and Wijnand Mijnhardt 2015. *Creating a knowledge society in a globalized world (1450-1800)* http://www.globalknowledgesociety.com/
- Epstein, S. R. 2013. 'Transferring Technical Knowledge and Innovating in Europe, c.1200 – c.1800'. In *Technology, Skills and the Pre-Modern Economy in the East and the West*, edited by Maarten Prak and Jan Luiten van Zanden, 25–67. Leiden; Boston: Brill
- Dooley, Brendan, 1999 *The Social History of Skepticism: Experience and Doubt in Early Modern Culture,* Baltimore
- Fuchs, Thomas. 2006. *Geschichtsbewusstsein und Geschichtsschreibung zwischen Reformation und Aufklärung. Städtechroniken, Kirchenbücher und historische Befragungen in Hessen, 1500-1800*, Marburg
- Golinski, Jan. 2010, *British Weather and the Climate of Enlightenment,* Chicago
- Groesen, Michiel van. 2016, 'Reading Newspapers in the Dutch Golden Age', *Media History* 22
- Jankovic, Vladimir. 2001. *Reading the Skies: A Cultural History of English Weather, 1650-1820*, Chicago
- Lindeman R., Y. Scherf and R.M. Dekker, 1993, *Egodocumenten van Noord-Nederlanders van de zestiende tot begin negentiende eeuw*, The Hague
- Mauer, Benedikt. 2001 *'Gemain Geschrey' und 'teglich Reden': Georg Kölderer. Ein Augsburger Chronist des konfessionellen Zeitalters*, Augsburg
- Mokyr, Joel. 2002. *The gifts of Athena. Historical origins of the knowledge economy,* Princeton UP
- Mokyr, Joel. 2016, *A culture of growth. The origins of the modern economy,* Princeton UP
- Morante, Roser, and Caroline Sporleder. 2012 'Modality and negation: An introduction to the special issue.' *Computational linguistics* 38.2 (2012): 223-260
- Munck, Bert de. 2015. 'The Fabrication of Useful Knowledge: An artisanal approach to literacy and theory', paper at *Knowledgeable Youngsters: Youth, Media and Early Modern Knowledge Societies,* 26-27 June 2015, Amsterdam (University Museum)
- Nierop, Henk van, 2009 'And ye shall hear of wars and rumours of wars' : rumour and the revolt of the Netherlands' in J. Pollmann and A. Spicer (eds), *Public opinion and changing identities in the early modern Netherlands*, Leiden
- Oostdijk, Nelleke, Martin Reynaert, Véronique Hoste, and Ineke Schuurman. The construction of a 500-million-word reference corpus of contemporary written Dutch. In: *Essential Speech*

and Language Technology for Dutch: Results by the STEVIN-programme, chapter 13, pages 219–247. Springer Verlag, 2013. URL http://rdcu.be/vOMV

- Pareti, Silvia. 2015. *Attribution: A Computational Approach*. PhD Thesis. University of Edinburgh
- Peeters, Thérèse. 2014. '"Sweet Milk-Cows" in Huizen and "Memorable Incidents" in Oost Zaandam. Identity and Responsibility in two eighteenth-century Rural Chronicles', *Volkskunde*, 2, 163-79
- Pettegree, Andrew. 2014. *The invention of news. How the world came to know about itself* London
- Pollmann, Judith. 2011. *Catholic identity and the Revolt of the Netherlands, c. 1520-1635* Oxford UP
- Pollmann, Judith. 2016. 'Archiving the Present and Chronicling for the Future in Early Modern Europe.' *Past & Present* 230 (suppl_11): 231–52. doi:10.1093/pastj/gtw029
- Pollmann, Judith. 2017. *Memory in early modern Europe*, Oxford UP
- Rau, Susanne. 2002. *Geschichte und Konfession. Städtische Geschichtsschreibung und Erinnerungskultur im Zeitalter von Reformation und Konfessionalisierung in Bremen, Breslau, Hamburg und Köln*, Hamburg and Munich
- Rau, Susanne, 2017. 'Die Städte und ihre Chroniken. Überlegungen zu einer Geschichte der Verflechtungen auf regionaler und globaler Ebene', in: S*tudien zur Stadtchronistik (1400-1850). Bremen und Hamburg, Oberlausitz und Niederlausitz, Brandenburg und Böhmen, Sachsen und Schlesien*, hg. von Lars-Arne Dannenberg und Mario Müller, Hildesheim
- Reynaert, Martin. 2005. *Text-Induced Spelling Correction*, PhD Thesis, Tilburg University
- Reynaert, Martin. 2010. 'Character confusion versus focus word-based correction of spelling and OCR variants in corpora', *International Journal on Document Analysis and Recognition*, 14:173–187
- Reynaert, Martin, Marten Gompel, Ko van der Sloot and Antal van den Bosch. 2015. 'PICCL: Philosophical Integrator of Computational and Corpus Libraries', In Koenraad De Smedt (Ed.), *Proceedings of CLARIN Annual Conference 2015: Book of Abstracts* (pp. 75-79). Wrocław, Poland: CLARIN ERIC
- Rosenberg, Daniel. 2003, 'Early Modern Information Overload'. *Journal of the History of Ideas* 64.1, 1-9
- Roberts, Lissa. 'Situating Science in Global History: Local Exchanges and Networks of Circulation'. *Itinerario* 33, no. 01 (2009): 9–30
- Salman, Jeroen 1999, Populair drukwerk in de Gouden eeuw. De almanak als lectuur en handelswaar, Zutphen
- Sarasin, Philipp. 2011. 'Was ist wissensgeschichte?' *Internationales Archiv für Sozialgeschichte der deutschen Literatur (*IASL). 36, pp. 159–172
- Schmidt, Heinrich. 1958. *Die deutschen Städtechroniken als Spiegel des bürgerlichen Selbstverständnisses im Spätmittelalter*, Göttingen
- Schneider, Gerold, Hans Martin Lehmann, and Peter Schneider. 2015. 'Parsing Early and Late Modern English corpora', *Literary and Linguistic Computing*, Volume 30, Issue 3, 1 Pages 423–439, https://doi.org/10.1093/llc/fqu001
- Shapiro, Barbara J. 2000. *A Culture of Fact. England, 1550-1720* Ithaca NY
- Slack, Paul. 2015 *The invention of improvement. Information and material progress in seventeenth-century England* Oxford
- Sluiter, I. 2017. 'Anchoring Innovation: a Classical Research Agenda', *European Review* 25.1, p. 1-19
- Smith, Pamela H., and Benjamin Schmidt. 2007. *Making Knowledge in Early Modern Europe. Practices, Objects, and Texts, 1400-1800*, Chicago, IL
- Soll, Jacob. 2010. 'From Note-Taking to Data Banks: Personal and Institutional Information Management in Early Modern Europe', *Intellectual history review* 20, 355-375
- Swidler, Ann. 2001. Talk of love. How culture matters; Chicago
- Schneider, Gerold, Hans Martin Lehmann, and Peter Schneider. 2015. 'Parsing Early and Late Modern English corpora', *Literary and Linguistic Computing*, Volume 30, Issue 3, 1 September 2015, Pages 423–439, https://doi.org/10.1093/llc/fqu001
- Walleghem, Jozef van 1985, *Merckenweerdigste voorvallen en daegelijksche gevallen Brugge 1790,* eds. Yvan vanden Berghe et al. Brugge