

# Opening the Gates to the Dutch Republic: A Comparison between Analogue and Digital Editions of the Resolutions of the States General

Ronald Sluijter<sup>1</sup>, Joris Oddens<sup>1</sup>, Rik Hoekstra<sup>1</sup>, Marijn Koolen<sup>1</sup>, Rutger van Koert<sup>2</sup>, Menzo Windhouwer<sup>2</sup>, Hennie Brugman<sup>2</sup> and Femke Gordijn<sup>1</sup>

<sup>1</sup>*Huygens Institute for the History of the Netherlands*

<sup>2</sup>*KNAW Humanities Cluster - Department of Digital Infrastructure, Netherlands*

## Abstract

REPUBLIC is a five-year project to compile a digital online edition of the resolutions of the Dutch States General issued between 1576 and 1796. In this article we compare the expected outcome of the project with the book and online editions that preceded it, and we assess the losses and gains of the new approach.

## Keywords

States General of the Dutch Republic, editions, Digital Humanities

## 1. Introduction

REPUBLIC is a five-year project aimed at creating a digital online edition of the resolutions issued by the States General of the Dutch Republic between 1576 and 1796.<sup>1</sup> The user interface of the online edition will include scans of the original resolutions in addition to a transcription. The search facets will enable various types of users to quickly find answers to realistic research questions, thanks to the structure we bring to the transcriptions using advanced methods tailored to the resolutions. By linking named entities that appear in the resolutions to other datasets that contain information about these entities, we will give more context to the data and increase research opportunities. In addition, we will provide an API to help technically savvy researchers perform more complex analyses on the data.

---

*Digital Parliamentary Data in Action (DiPaDA 2022) workshop, Uppsala, Sweden, March 15, 2022.*

✉ ronald.sluijter@huygens.knaw.nl (R. Sluijter); joris.oddens.huc@knaw.nl (J. Oddens); rik.hoekstra@di.huc.knaw.nl (R. Hoekstra); marijn.koolen@di.huc.knaw.nl (M. Koolen); rutger.van.koert@di.huc.knaw.nl (R. v. Koert); menzo.windhouwer@di.huc.knaw.nl (M. Windhouwer); hennie.brugman@di.huc.knaw.nl (H. Brugman); femke.gordijn@huygens.knaw.nl (F. Gordijn)

🆔 0000-0002-2896-9986 (R. Sluijter); 0000-0002-0518-2136 (J. Oddens); 0000-0002-6951-8014 (R. Hoekstra); 0000-0002-0301-2029 (M. Koolen); 0000-0001-6535-2849 (R. v. Koert); 0000-0002-2204-4018 (M. Windhouwer)

© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

<sup>1</sup>REPUBLIC is an acronym for REsolutions PUBLISHED In a Computational Environment (2019-2023), a project funded by the Dutch Research Council (NWO). The project is a partnership between The Huygens Institute for the History of the Netherlands and the National Archives of the Netherlands. For the project website see <https://republic.huygens.knaw.nl/>.

REPUBLIC is the successor of a book edition and a subsequent born-digital edition, which together took a century to produce.[1, 2, 3].<sup>2</sup> Yet these three earlier projects only covered the resolutions from 1576 to 1630. When it became clear that completing the edition in this form would take many more decades, if not centuries, and given the availability of new and rapidly developing technical solutions, it was decided in 2018 to submit a proposal for a new approach with better chances of completion. The proposal was honoured by the Netherlands Research Council (NWO) and the project started in 2019.

The book editions compiled by our predecessors are known among specialists for their thoroughness and quality, as well as for the accessibility and usability of the material they provide by means of critical and explanatory footnotes, lists and indices. In the REPUBLIC project, we are obviously also striving for an edition of high quality that will be accessible and usable by a broad public, but primarily by researchers. It is clear, however, that due to the enormous amount of material, there is a certain tension between meeting the high standards of the book editions and striving to publish the entire series of resolutions. On the other hand, a digital edition naturally has a number of advantages over a book edition, and as we shall see, the first series of book editions in particular can also be criticised in some respects. Therefore, we do not only want to present the new project, but also to examine what we lose and what we gain by switching to the new approach. In this way, we hope to initiate further discussion on the implications of the transition to digital editions in general.

## **2. The significance of the resolutions of the Dutch States-General**

The Republic of the Seven United Provinces, or the Dutch Republic for short, came into being at the end of the sixteenth century as a result of a revolt against the overlord, King Philip II of Spain. The rebellious provinces formed a new confederate state, and sent delegates to the States General, which served as an arena for negotiating their common interests, such as foreign policy, defence and related finances, and religious matters. The decisions ('resolutions') made during the daily meetings were drafted by the greffier (chief clerk) of the council and then copied into the fair register.

The total number of pages of the series of resolutions from 1576 to 1796 amounts to about half a million. The contents are of great importance as a historical source. The resolutions are an indispensable source for Dutch history, because they concern both high and low politics and virtually all aspects of society. But their significance is not limited to Dutch history. In the seventeenth and eighteenth centuries, the Dutch Republic held a powerful position in Europe and in various parts of the world, based on its central position in a global maritime trade network and a relatively tolerant religious and scientific climate. The country was therefore often involved in important developments in international politics, trade, arts and sciences, which is also reflected in the resolutions. Moreover, the resolutions are not only of great value for historical research into specific events. The availability of two and a half centuries of uninterrupted decision making offers an important insight into the functioning of a political body as well as the development of (pre)parliamentary culture in the Early Modern Age.

---

<sup>2</sup>Although the first volume appeared in 1915, work on it started several years earlier. There's even an older edition of the resolutions, covering only 1576 and the first half of 1577 ([4])

Furthermore, the resolutions function as a gateway to the underlying documents (petitions, reports, letters, accounts et cetera) in the archive of the States General, as these have been archived according to the date on which a resolution was taken on them.

### **3. The editions of the resolutions: changing attitudes**

Because of the importance of this source, the publication of the resolutions has long been on the agenda of the Huygens Institute of History of the Netherlands and its predecessors. The first book series [1] offered a mixture of transcriptions and summaries of the resolutions in contemporary Dutch. In this series, the resolutions are arranged according to a classification system with a limited number of classes (e.g. War, Foreign Affairs, Finance, Religion), sometimes in turn subdivided into sub-classes. Within these (sub)classes, the resolutions were arranged chronologically. The volumes also provided footnotes that were either explanatory or referred to other relevant documents in the States General archives; contained lists of those present at the meetings; lists of appointments of members of other councils by the States General; chronological lists of the resolutions and of incoming and outgoing documents; indices of names and subjects; and a list of references to printed books and archive sources. It is important to note that not all resolutions were included in this first series. This practice was similar to other editions of large archive series produced in the same period: which resolutions were included reflected the historical interest of the time, and also depended greatly on the personal opinion of the editor. [5, p. 71]. As interest shifted over time, historians of later generations, who were often interested in subjects previously considered less important, found the editions less useful for their research.

In the 1960s, awareness of this problem led to a different approach. In the new series [2] the compilers included all resolutions - those that were considered less important in footnotes - and arranged them chronologically. Initially, the old classification was added to the summaries of the resolutions, but this stopped from volume 3 onwards. The chronological lists of resolutions that had been included in the first series were now superfluous and are therefore not included in the second series. The lists of incoming and outgoing documents, and of appointed members in other councils, were also dropped. Another change was that from now on the resolutions were numbered and the indices referred to these numbers instead of page numbers.

Although the new approach better suited the changing interests of historical research, it did not speed up the work. In the last decade of the twentieth century, plans were made for a new type of edition using rapidly developing digital techniques. This resulted in the complete abandonment of book publishing and the adoption of an XML-based approach. [6]. The new online edition was, in retrospect, an intermediate form between the book edition and the edition to be produced by the REPUBLIC project. The resolutions were summarised in modern Dutch, as in the books, while personal names, institutions, places, books and pamphlets were manually marked in the text. These marked elements were indexed in searchable databases. The person database also contained offices held. This made it possible to search for persons holding the same office, which was not the case in the book series. It was now also possible to search the text of the summaries of the resolutions, which further increased accessibility. However, this advantage over the book editions was cancelled out when the books were digitised in 2007 and

2008, resulting in a machine-readable, searchable version of the text.

## 4. The REPUBLIC project

Although fewer references to other documents in the States General archives were included in the born digital edition than in the book edition, this did not speed things up either. This fact led to a long pause in the publication of the resolutions. During this period, however, the idea of a continuation in some form was never entirely abandoned. There was an awareness of the rapid technical developments in various fields that could benefit a future project, such as Handwritten Text Recognition. In 2018, a grant proposal for the REPUBLIC project was awarded by the Netherlands Research Council (NWO), and a year later the project started.

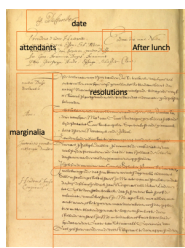


Figure 1: Structure of the handwritten resolutions

Today many digitised printed text corpora are online, also from before the nineteenth century. Apart from the fact that the OCR is often of poor quality, they often lack or provide only raw metadata to structure the text. As a result, finding answers to research questions is often only possible by reading through large amounts of text. The compilers of the relatively small number of handwritten corpora that are online or still in the project phase - the development of HTR is, after all, still relatively new - limit themselves, it seems, mainly to indexing named entities and topics extracted from the text, without applying in-depth structure to the text.<sup>3</sup>

The ambition of REPUBLIC reaches beyond merely presenting a machine and human readable online representation of the resolutions alongside images of the originals, and a simple text-search function. Instead, we strive for enabling users to answer realistic research questions like: 'Do the resolutions over time reflect an increasing possibility for citizens to put forward their concerns?', 'Who were involved in decisions around specific topics?', 'When was economic policy discussed?'. We do this by offering searches through multiple, connected layers of metadata that are based on the structure of the resolutions themselves (see Figure 1 and 2).

Before we can go into this in more detail, it must be clear that having a good quality machine-readable text is a prerequisite for structuring it. We use contemporary printed volumes of the ordinary resolutions for the years between 1703 and 1796. The ordinary resolutions between 1576-1702, as well as the secret resolutions between 1576 and 1796, are only available in handwritten form. To convert the printed resolutions to machine readable text, we used OCR, while for the handwritten resolutions we used Handwritten Text Recognition (HTR) on the Transkribus platform.<sup>4</sup>

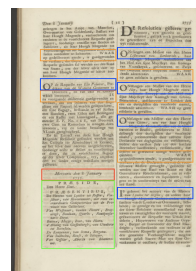


Figure 2: Elements in the printed resolutions.

Both yielded very good results. The OCR on the printed resolutions was trained with Tesseract, which was necessary because the fonts in eighteenth-century print differ from those in modern print. Currently, we have achieved an average of 0.28% Character Error Rate (CER) on the printed resolutions.

<sup>3</sup>Some examples on <https://readcoop.eu/success-stories/>

<sup>4</sup><https://readcoop.eu/transkribus>

There are still some problems with correctly recognising headings, capital letters and non-standard formats such as tables, which are sometimes included in the resolutions. We will continue to improve the OCR for these edge cases during the remaining duration of the project. For handwritten resolutions, we have trained several HTR models and so far have achieved a CER of only 2.99% on a validation set based on a ground truth of about 1,000 pages. For further improvement of the HTR model, a crowdsourcing project has been set up, in which volunteers correct the results of the HTR. This will provide more ground truth to improve the model, and will also provide high quality transcriptions for at least 10-12% of the total number of handwritten pages at the end of the project. A caveat to the HTR results is that they also depend on automatic text region and baseline detection. The 2.99% CER was achieved on pages for which this automatic layout analysis was manually corrected. Depending on the clerk who wrote the resolutions, the automatic layout analysis performs differently. Obviously, the CER on pages with uncorrected layout is higher, but we will continue to improve the automatic layout analysis until the end of the project in 2024.

The total size of the handwritten and printed resolutions is obviously too large for the errors in the OCR and HTR to be corrected manually. However, the automated methods we use to extract information from the text and structure it allow a margin of error. Three years after the start of the project, we are very confident that we will stay within these margins for the vast majority of the text.

The methods we use for information extraction differ from the tools commonly applied to modern texts, such as Named Entity Recognition (NER), part-of-speech tagging and lemmatisation, as these have been found to yield disappointing results. To investigate the value of generic NER techniques, we annotated named entities in 200 pages of manually transcribed resolutions, retrained the Spacy NER tagger with 90% of the pages, and tested on the remaining 10%. This resulted in a precision of 0.49 and a recall of 0.19, which is typical for entity recognition in historical documents.[7, 8, 9]. Since Dutch orthography has changed considerably since the early modern period, NLP tools based on modern Dutch are less effective on older texts, and the results are too generic to be of much use for research. Moreover, these techniques are not suitable for extracting other types of information that we need to provide meaningful answers to research questions.

In order to achieve this, we use repetitive structural elements in the layout of the resolutions and the repetitiveness of certain phrases that have remained remarkably stable throughout the period. Using an iterative process in which we include knowledge about the general structure of the resolutions—e.g. that they are chronologically ordered, that a meeting starts with an attendance list—, specific domain knowledge, and fuzzy string searching and matching, we can now automatically identify with high precision for the OCR output of printed resolutions the start of a meeting and the date of a meeting; identify attendance lists; distinguish one resolution from another; and identify their opening 'proposition'<sup>5</sup>, the 'conclusion'<sup>6</sup>, and the 'decision formula' [10, p. 476].

An initial pilot of reusing this pipeline on the HTR output was successful. Focused on the

---

<sup>5</sup>Typically signalled by phrases like: 'Received the petition of ...' (Ontfangen de requeste van ...)

<sup>6</sup>'Upon which, after deliberation, it was decided to ...' ('Waerop gedelibereert sijnde, is goetgevonden ende verstaen ...')

resolutions of the Dutch ‘year of disaster’ 1672, the pipeline was able to detect over 80% of the sessions and resolutions, with only minimal updates to the list of textual formulas to identify resolution boundaries. There will be more challenges for even older resolutions, because there the language variation is greater and more different from that of the 18th century, and because the consistency of formulaic language is lower.

Several parts of the pipeline can be applied more generally. The fuzzy search module has already been reused in projects to identify place name attestations in medieval charters, and to identify common phrases in 16th-18th century mandate and policy books from the Bern Staatsarchiv. The code for structuring physical and logical text elements was reused in the Golden Agents project for capturing individual deeds in the notarial archives of Amsterdam.<sup>7</sup>

We are currently developing ground truth data to extract names from the lists of attendees. These person entities will be linked to a jump-off page where users can access entries on these persons in other datasets published by the Huygens Institute.<sup>8</sup> As far as possible, the technical solution for this will line up with the ever expanding Linked Data cloud.<sup>9</sup>

In order to identify person entities in the text of the resolutions themselves, we make use of the additional ‘role’ and ‘quality’ information that is usually found around person names, such as ‘*Ambassador Van Beuningen*’, ‘*Jan Roelofsz., silversmith*’, ‘*Maertje Jansz., widow of Pieter Claesz.*’. This promises to produce much better results than conventional NER.

Further steps during the remaining time of the project are identifying index entries and their page references and linking them to the appropriate resolution; exploring ways to index subjects; and connecting the layers of metadata and extracted entities with each other and across the volumes of the resolutions.

All data and metadata will be made available online as linked open data according to common standards. Scan images will be published according to IIIF image and presentation formats and protocols. Texts will be published online in such a way that a particular text segment can be linked and retrieved, for example the text of a specific resolution. All extracted logical elements (e.g. resolutions, session days, attendance lists) will be provided with bundles of links to the corresponding image parts, text segments and metadata. These bundled links will be represented and published as W3C Web Annotations. All text, metadata and web annotations will be searchable and accessible via web based APIs.

A user interface to search and browse the resolutions is under development. We have made a prototype that is already available.<sup>10</sup> It currently provides images of the printed resolutions alongside the transcription, as well as search facets based on the metadata and the extracted person entities, and a free text search. At the end of the project, users will also be able to use an API to perform more complex analyses of the data.<sup>11</sup>

---

<sup>7</sup><https://www.goldenagents.org>

<sup>8</sup>E.g. the Repertorium van ambtsdragers en ambtenaren 1428-1861, <https://resources.huygens.knaw.nl/repertoriumambtsdragersambtenaren1428-1861>

<sup>9</sup><https://lod-cloud.net>

<sup>10</sup>See <https://demo.docere.diginfra.net/projects/republic>

<sup>11</sup>For a more detailed explanation of the connection between research questions and information extraction in the project, see [11]



## 5. From Analogue to digital: what do we gain and what do we lose?

If we compare the previous editions with the new edition, three clear advantages of the latter stand out. The first is that, for the first time, we will be publishing the entire series of resolutions from 1576 to 1796. It is the rapid technical development of OCR and especially of HTR that makes possible what would have been practically unfeasible in the analogue way. The second advantage is that we take the transcription as the basis of the edition, whereas the previous editions limited themselves to a summary of the resolutions in modern Dutch. Although one might say that these summaries offer better usability for readers who are not proficient in the older Dutch, users generally prefer to have access to the original text. Some would even argue that the earlier editions are not editions at all, as they do not offer a representation of the text in the form of an image or a transcription. [12, p. 23]. The fact that we will include images of the resolutions alongside their transcriptions is the third advantage of the new edition over the earlier editions.

As for the possible losses due to the new approach, we may wonder whether the desire to publish the entire series of resolutions leads to unjustifiable compromises in quality. After all, although our training of the OCR and HTR on the resolution texts produces impressive results, it is clear that the majority of handwritten transcriptions will still contain 2-3% errors at best. Can we claim to be producing a digital scientific edition if the transcriptions are not "the best realisation possible"? [12, p. 35]?

A simple justification would be that having a transcription at one's disposal is always better than having no transcription at all. However, this argument is not always justified, for example, in the case that OCR and HTR result in a mutilated, unintelligible text, or that the text lacks a critical apparatus [13, p. 6]. However, neither of these applies to REPUBLIC. As for the transcriptions, the results we have achieved so far on a representative part of the total material, and our expectation to further improve the automatic layout analysis, indicate that the error rate in the uncorrected OCR and HTR of most of the text to be produced by the end of the project will be small enough to make the vast majority of the text understandable to human readers, and also for automatic applications to extract information from the text. Of course, the application of these algorithms does not guarantee that the search results will be complete. But, as we have experienced in this project and in other editions, human-made transcriptions and indexes are also far from being free of errors and omissions. Moreover, a digital edition offers the possibility of making manual corrections and additions after the initial online publication. It is also very likely that OCR and especially HTR techniques and models will continue to improve in the future. This makes improving the transcriptions of the resolutions at a later stage, by applying a new OCR and HTR cycle to the images, at least a theoretical possibility.

In order to be a scholarly edition, the new edition also has to be critical, a term to be understood here as 'all processes that engage in a critical or reflective way—that is, on the basis of a scholarly agenda—with the material in question and help in "opening it up" to the intended outcome' [12, p. 24]. It's already clear from the above that the new edition will be critical, but how does REPUBLIC compare to the previous editions in this respect?

As we saw, the previous editions offered several elements that significantly increased their

usefulness for research. In fact, the manual creation of the footnotes, lists, indexes, etc. was a major reason why compiling these editions was such a time-consuming task. In the new edition, we want to make the resolutions as accessible to the users as in the previous editions. However, due to the volume of the material, we have to rely primarily on automated means, as in the case of the transcriptions. Of course, much manual work is still being done, but this is mainly limited to developing and training algorithms and checking and correcting their output in order to improve their performance. In the development and application of these algorithms, we measure the precision and recall in relation to the ground truths. The results will be included in the accompanying texts of the publication so that users can assess the value of their searches. We expect that our use of fuzzy string searching and matching will significantly reduce the negative effects of transcription errors on searchability and other methods of extracting information from the text.

It goes without saying that we have to accept that the results will not be perfect, and that access points to the text similar to those in the previous editions - indexes of names and subjects - will not be of the same quality. Also, the level of additional, explanatory information provided in the book editions in the footnotes will be lower. Nevertheless, the linking of person entities to other datasets, which is part of our current effort, will at least partially eliminate this drawback. In addition, we will make more manual efforts to identify and disambiguate the attendees at the meetings of the States General.

Moreover, apart from the already mentioned advantages that the new edition will have compared to the book edition, there is another important advantage. The fact that we are creating a truly digital edition means that it follows a 'digital paradigm', allowing for multiple layers of content and functionality that book editions usually do not, and in some cases cannot, provide due to technical and economic limitations and cultural practices of typography and printing [12, p. 31]. The aforementioned structuring of the text, combined with extracted entities, offers possibilities for quick search, selection and ranking of resolutions across the whole corpus, which in book editions would take an enormous amount of time to carry out manually. The same applies to various types of analysis of the data, which will be possible through the use of the API. Linking entities to external sources, which might even provide updates in the future, is not possible in the book paradigm. By linking person entities via a central repository that the Huygens Institute is building, the contextual information about person entities that appear in the resolutions will also increase after the project is finished, in case more external sources are linked to this central repository.

## 6. Conclusion

In conclusion, we can say that the advantages of the new edition of the resolutions outweigh the disadvantages with respect to the earlier editions. It is true that the manual labour put into the latter editions has guaranteed more reliable and complete indices and explanatory data. But apart from the fact that even handwork does not guarantee perfection, the final result of REPUBLIC will, for the first time within an acceptable time frame, offer a complete and consistent edition of the resolutions from 1576 to 1796, including transcriptions and images of the originals. Although the quality of the transcriptions, especially of the handwritten



resolutions, will for the most part be slightly lower than that of manual transcriptions, users will be able to compare them with the original text. Our automated methods of information extraction are able to reduce the negative impact on searchability and accessibility to a large extent. Moreover, the fact that the result will be a truly digital edition, following a digital paradigm, leads to possibilities that would have been unthinkable in the book paradigm.

## Acknowledgments

This research is funded by Dutch Research Council (NWO).

## References

- [1] N. Japikse (Ed.), *Resolutiën der Staten Generaal van 1576 tot 1609*, volume 1-14, Martinus Nijhoff, 's-Gravenhage, 1915-1970.
- [2] A. van Deursen et al. (Ed.), *Resolutiën der Staten Generaal*, volume 1-7 of *Nieuwe reeks, 1610-1670*, Martinus Nijhoff, 's-Gravenhage, 1971-1994.
- [3] I.Nijenhuis, et al., *Besluiten Staten Generaal 1626-1630*, 2006. URL: <http://resources.huygens.knaw.nl/besluitenstatengeneraal1576-1630/BesluitenStatenGeneraal1626-1651>.
- [4] J. de Jonge, *Besluiten van de Staten Generaal der Nederlanden = Résolutions des États Généraux des Pays-Bas*, volume 1-2, Algemeene Lands drukkerij, 1828-1831.
- [5] J. Roelevink, *Navigating new waters. The project "Resolutions of the States-General 1576-1670*, Institute of Netherlands History, 2007, pp. 67–83.
- [6] I. Nijenhuis, *Besluiten ontsloten. Resolutiën Staten Generaal digitaal (1626-1630)*, *De Zeventiende Eeuw* (2006) 272–282.
- [7] L. Borin, D. Kokkinakis, L.-J. Olsson, *Naming the past: Named entity and animacy recognition in 19th century swedish literature*, in: C. Sporleder, A. van den Bosch, C. Grover (Eds.), *Proceedings of the Workshop on Language Technology for Cultural Heritage Data, LaTeCH 2007*, 2007, pp. 1–8.
- [8] K. Rodriguez, M. Bryant, T. Blanke, M. Luszczynska, *Comparison of named entity recognition tools for raw ocr text*, *Konvens* (2012) 410–414.
- [9] M. Ehrmann, A. Hamdi, E. Pontes, M. Romanello, A. Doucet, *Named entity recognition and classification on historical documents: A survey*, arXiv preprint arXiv:2109.11406 (2021). URL: <https://arxiv.org/abs/2109.11406>.
- [10] T. Thomassen, *Instrumenten van de macht. De Staten Generaal en hun archieven 1576-1796*, Huygens Instituut voor Nederlandse Geschiedenis, 2015.
- [11] M. Koolen, et al., *Modelling resolutions of the dutch states general for digital historical research*, in: *Proceedings of the International Conference Collect and Connect 2020*, CEUR Workshop Proceedings vol. 2810, 2020.
- [12] P. Sahle, *What is a scholarly digital edition?*, in: M. Driscoll, E. Pierazzo (Eds.), *Digital Scholarly Editing. Theories and Practices*, Open Book Publishers, Cambridge, 2016, pp. 19–39.
- [13] M. Driscoll, E. Pierazzo, *Introduction: Old wine in new bottles?*, in: M. Driscoll, E. Pierazzo (Eds.), *Digital Scholarly Editing.*, 2016, pp. 1–15.