

Antwerp Street Stories: ±90,000 Machine-Readable Pages of Handwritten Local Police Reports (1876–1945)

Lith Lefranc (corresponding author)

Centre for Urban History, University of Antwerp; Antwerp Centre for Digital Humanities and Literary Criticism, University of Antwerp, Antwerp, Belgium, Lith.Lefranc@UAntwerpen.be 

Mike Kestemont

Antwerp Centre for Digital Humanities and Literary Criticism, University of Antwerp, Antwerp, Belgium, Mike.Kestemont@UAntwerpen.be 

Ilja Van Damme

Centre for Urban History, University of Antwerp, Antwerp, Belgium, Ilja.VanDamme@UAntwerpen.be 

Received 20 February 2025, revised 22 May 2025, accepted 1 September 2025

Abstract

We present Antwerp Street Stories: a machine-readable dataset of approximately 90,000 pages from the incident books of Antwerp, Belgium, spanning from 1876 to 1945. In these books, local police officers meticulously documented daily incidents on the streets, capturing information on the location, time, and date of each event, along with personal details of the involved citizens. The records offer a unique window into the social dynamics of everyday urban life in a rapidly expanding metropolis. This paper outlines the digitization process, including handwritten text recognition (HTR). It also describes the dataset's key characteristics, such as its size, temporal and geographical scope, languages, and HTR accuracy. Despite some inherent limitations, such as data gaps and registration biases, the dataset holds significant potential for further research. Antwerp Street Stories supports data-driven approaches as well as in-depth explorations for (urban) histories of all kinds, among others urban crime, police-community interactions, and the lived experiences of lower social groups.

Keywords

handwritten text recognition – police reports – digital history – urban history – social history

Related Data Set

Raw data on Zenodo: <https://doi.org/10.5281/zenodo.12805005> (released under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International license (CC BY-NC-SA 4.0))

Necturus XML viewer (Shahidzadeh Asadi, 2025): <http://gebeurtenissen.flw.uantwerpen.be/>

1. Introduction

European cities, including Antwerp, experienced rapid urban growth during the latter half of the nineteenth century. Antwerp's rise as one of the world's largest harbors not only brought economic prosperity but also social challenges (Blondé et al., 2005). As the city's population swelled, so did concerns over overpopulation, poverty, and anxiety for increased crime. One of the strategies to deal with these problems was through large-scale investments in the local police (De Koster & Vrints, 2020). These police reforms created new administrative instruments documenting in great detail what transpired on the Antwerp streets. In the so-called "incident books" (Dutch: *gebeurtenisboeken*; French: *registre des événements*) local policemen wrote daily reports after patrolling their neighborhoods (see Figure 2). Based on these reports, deputy commissioners decided whether an official report (*proces-verbaal*) was in order. Consequently, the incident books document many aspects of everyday Antwerp street life: offenses, petty crimes, neighborhood quarrels, and other irregularities, such as missing children, stray dogs, or injured people lying on the street (see Figure 2 and Table 1). They offer insights into the everyday challenges faced by both the authorities and the residents in a rapidly growing urban environment.



FIGURE 1: Local Police Officer Patrolling the Vismarkt, Antwerp, Ca. 1930 (Felixarchief, GP#513).

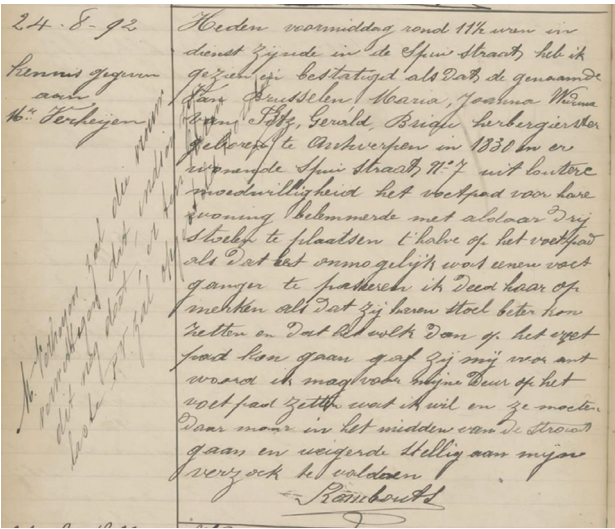


FIGURE 2: Example of an Incident (Felixarchief, 450#58, p. 507).

TABLE 1: Manual Transcription of One Incident From Figure 2 and Translation to English (FelixArchief, 450#58, p. 507)

24-8-92 Kenniss gegeven aan Mr Verheyen	Heden voormiddag rond 11 ½ uren in dienst zijnde in de Spuistraat heb ik gezien en bestatigd als dat de genaamde Van Brusselen Maria, Joanna Weduwe van Pitz, Gerald, Briaux herbergierster geboren te Antwerpen in 1830 en er wonende Spuistraat n°7 uit loutere goedwilligheid het voetpad voor hare woning belemmerde met aldaar drij stoelen te plaatsen t'halve op het voetpad als dat het onmogelijk was eenen voet ganger te paseeren ik deed haar op merken als dat zij haren stoel beter kon zetten en dat het volk dan op het voet pad kon gaan gaf zij mij voor ant woord ik mag voor mijne deur op het voetpad zetten wat ik wil en ze moeten daar maar in het midden van de straat gaan en weigerde stellig aan mijne verzoek te voldoen Rombauts
24-8-92 Notification given to Mr. Verheyen	This morning around 11:30 while on duty in Spuistraat, I saw and confirmed that the named Van Brusselen Maria, Joanna, widow of Pitz, Gerald, Briaux, innkeeper, born in Antwerp in 1830 and residing there at Spuistraat n°7, deliberately obstructed the sidewalk in front of her residence by placing three chairs there, making it impossible for a pedestrian to pass. I pointed out to her that she could position her chair better so that people could walk on the sidewalk, to which she replied that she could put whatever she wanted on the sidewalk in front of her door and that people should walk in the middle of the street. She firmly refused to comply with my request. Rombauts

The serial character of the incident books ensures a fairly consistent registration over time of the following information:

- The location, date, and time of the incident;
- Personal data of the citizen(s) involved: name, marital status, profession, place of birth, date of birth, and place of residence;
- Description of the incident;
- Personal data of the officer(s) that observed the incident: name, number, and signature;
- Information on the follow-up of the incident: whether an official report was in order (if yes, the number of the official report), name and signature mark of the deputy officer who followed up on the incident.

The incident books thus provide a detailed view of the broader context of the event and the social backgrounds of all the people involved. This is especially valuable for urban historians exploring the everyday lives of social groups who have left almost no “paper trail” of their own (Thompson, 1963; Wolf, 1982).

Until now, qualitative and sample-based methods have significantly shaped historical research “from below”. Only recently, social historians have been discovering the potential of the application of data-driven methods toward criminal records (Saygi & Yasunaga, 2021; van den Heuvel et al., 2020). While much attention in digital history has focused on large-scale analysis, Julia Laite (2020) points out that digitization also opens new possibilities for small-scale, narrative-driven research—allowing historians to follow marginal individuals across scattered records. One of the most straightforward explanations for the delayed uptake of such approaches is the lack of machine-readable data. Grand digitization initiatives such as Old Bailey Online remain scarce because they require time, money, and labor (Hitchcock & Shoemaker, 2006). In response, platforms created by crowdsourcing and citizen science projects have been mushrooming, yet these still pose problems with consistent data quality and are not always in line with topical questions and debates forwarded by the academic community (Greefs, 2023).

In the context of a data-driven research project on nocturnal street life in modern Antwerp, we have used open-source software for handwritten text recognition (HTR) to automatically transcribe approximately 90,000 pages of the incident books. In this paper, we consecutively explain the methods we used to create this dataset, its main features (size, temporal and geographical scope, bilingualism, and HTR quality), and the limitations and possibilities for further research. The goal of this paper is twofold: first, to make our dataset more widely known to historians, and second, to ensure transparency regarding both the nature of the source and the steps taken during its digitization, which are essential for using and interpreting the data in future research (Fickers, 2012).

2. Methods

The workflow used for converting the incident books to machine-readable text consisted of:

- 1) *Selection*: The Antwerp City Archives preserve 597 incident books (1829; 1876–1970). Due to the timeframe of our research project, we have selected 271 incident books for the current dataset (written between 1876 and 1945).
- 2) *Scanning*: We outsourced the scanning of the incident books to the digitization company Picturae. They provided us with high-quality (300 dpi) images in jpg format for each page.
- 3) *Layout and handwritten text recognition*
 - 1) *Creation of ground truth*: While a ground truth training set of 15,000 words is typically sufficient for good HTR results (Muehlberger et al., 2019), the diverse handwriting styles, the irregular layout (marginal and main body text; see Figure 3), and the bilingual nature of the incident books required more data. An expert (PhD student) annotated the layout and transcription of 271 randomly chosen pages (78,581 words). Later, undergraduate history students contributed annotations for an additional 3444 pages (900,091 words), although the quality of the student subset is less consistent compared to the expert annotations. To further enhance the ground truth for the incident books, we incorporated the publicly available IJsberg dataset (1,538,478 words), which includes 4,735 pages from the seventeenth- and eighteenth-century Dutch East-India Company archive and 1,615 pages of nineteenth-century Dutch notarial deeds (Keijser, 2020).
 - 2) *Model training*: Because of the high cost of proprietary software for large datasets, we trained a series of open-source layout and HTR models with YALTAi (Clérice, 2023) and *kraken* (Kiessling et al., 2019), which we baptized under the name ARletta (Lefranc et al., 2024). For the HTR, we used the publicly available *ManuMcFrench* (Chagué et al., 2023) as a foundational model and fine-tuned it with models based on the IJsberg and the Antwerp data. ARletta achieves an absolute Character Accuracy Rate (CAR) of 92.7% and a relaxed CAR (excluding white space, punctuation, and capital letters) of 93.97% on the test set.
 - 3) *Application of LA and HTR models*: We applied the models to the images in the following order: baseline detection (*kraken*), layout segmentation (YALTAi), and handwritten text recognition (*kraken*).

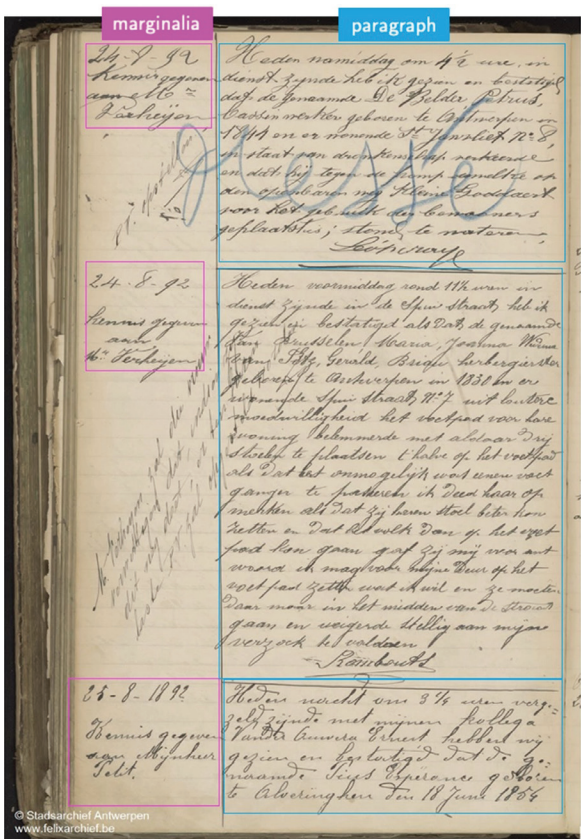


FIGURE 3: Page of an Incident Book With Marked Marginal (*Marginalia*) and Main Body Text (*Paragraph*) (Felixarchief, 450#58, p. 507).

- 4) *Post-processing*: The raw output consists of page XML files for each page (image) of the incident books with the image coordinates for region and line detection, and their corresponding transcriptions. On the region level, the model distinguishes vertically between marginal (*marginalia*) and main body text (*paragraph*), and horizontally between different incidents (see Figure 3). We post-processed the page XML files to extract transcriptions and text region tags only. To determine the reading order, we alternated text regions tagged as *marginalia* (containing dates and information on the follow-up of the incident) with text regions tagged as *paragraph* (describing the actual incident). We used the image coordinates of the text regions to pair *marginalia* and corresponding *paragraphs* (based on vertical overlap) (see “postprocess-xml.ipynb” in the repository). Because this method is not fool-proof, markers were added in the new XML files to indicate different text regions (see example in Table 2).

TABLE 2: Automatic Transcription Including XML Tags of the Page Displayed on Figure 3 (Felixarchief, 450#58, p. 507)

<page number="507"> <incident> <marg>24-8-92 Kennis gegeven aan Mr Perheyen D D</marg> <p>Heden namiddag om 4 1/2 ure, in dienst zinde heb ik gezien en bestatigd dat te genuamde De Perner Peturs, bassinwerker geboren te Antwerpen in 1844 en er wonende St Janvliet n° 8 in staat van dronkenschap verkeerde en dat hij tegen de pomp dewelke op den openbaren weg Kleine Goddaert voor het gebrutk der bemooners geplaatstis; stond te wateren Lopcray</p> </incident> <incident> <marg>24-8-92 kennis gegeven aan Mr Verheijen i</marg> <p>Heden voormiddag rond 11 1/2 uren in dienst zijnde in de Spui straat heb ik gezien en bestatigd als dat de genaamde Van Brusselen Maria, Joanna Weduwa van Pitz, Gerald, Briau herbergierster geboren te Antwerpen in 1830 un er wonende Spui straat n°7 uit loutere moedwilligheid het voetpad voor hare woning belemmerde met aldaar drij stoelen te plaatsen t'halve op het voekpad als dat het onmogelijk was eenen voet ganger te paseren ik deed haar op merken als dat zij haren stoel beter kon zetten en dat het volk dan op het voet pad kon gaan gaf zij mij voor ant woord ik mag voor mijne deur op het voetpad zetten wat ik wil en ze moeten daar maar in het midden van de straat gaan en weigerde stellig aan mijne verzoek te voldoen Rombouts</p> </incident> <incident> <marg>25-8-1892 Kenns gegeven aan Mijnheer Petit</marg> <p>Heden nacht om 3 3/4 uren verge zeld zijnde met mijnen kollega Vander Auwera Ernert hebben wij gezien en bestatigd dat de ge naamde Pius Esperance geboren te Alveringhen den 18 Juni 1854</p> </incident> </page>
--

3. Data exploration

The resulting machine-readable version of the dataset allows us to identify and calculate its main features: size, temporal and geographical distribution, languages, and HTR quality. Transparency about the gaps and emphases of the dataset is vital to structurally acknowledge prevalent source biases, offering invaluable assistance in developing future research questions and historical interpretations.

3.1. Size

The HTRed dataset of 271 incident books contains 87,593 pages (sides), consisting of 410,829 incidents, 3,804,556 lines, 21,555,587 words, and 128,408,985 characters (see average per book and page in Table 3).

TABLE 3: The Average Number of Pages (Sides), Incidents, Lines, Words, and Characters Per Book, Page, and Incident

	Book	Page	Incident
Pages	323.92	N/A	N/A
Incidents	1521.61	4.71	N/A
Lines	16,645.63	51.39	10.91
Words	80,401.07	248.21	52.74
Characters	472,610.25	1459.04	310.59

3.2. Temporal Evolution

Overall, there is an incident increase towards the final years of the dataset (see Figure 4). Although there are many inconsistencies in terms of preservation conditions, we can explain the steady rise by the mounting investments in the Antwerp police force (e.g. a fivefold increase in the number of police officers per capita (from one officer for every 1020 inhabitants in 1870 to one officer per 216 inhabitants in 1940)). During both war periods (1914–1918 and 1940–1945) policing on the Antwerp streets intensified (Saerens Lieven, 2000), so the peak toward the final years of the dataset produces a slightly distorted picture.

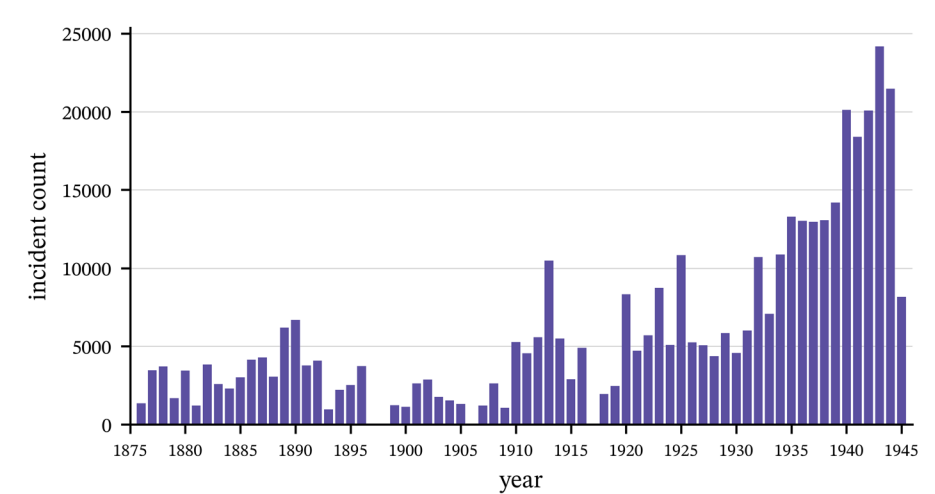


FIGURE 4: Total Incident Count for Each Year, 1876–1945.

3.3. Geographical Distribution

Due to urban expansion and (subsequent) population growth during the first decades of the twentieth century, the Antwerp police districts changed fundamentally in 1921. The incident count for each district is therefore displayed on two different maps (see Figure 5a and 5b) and tables (see Table 4).

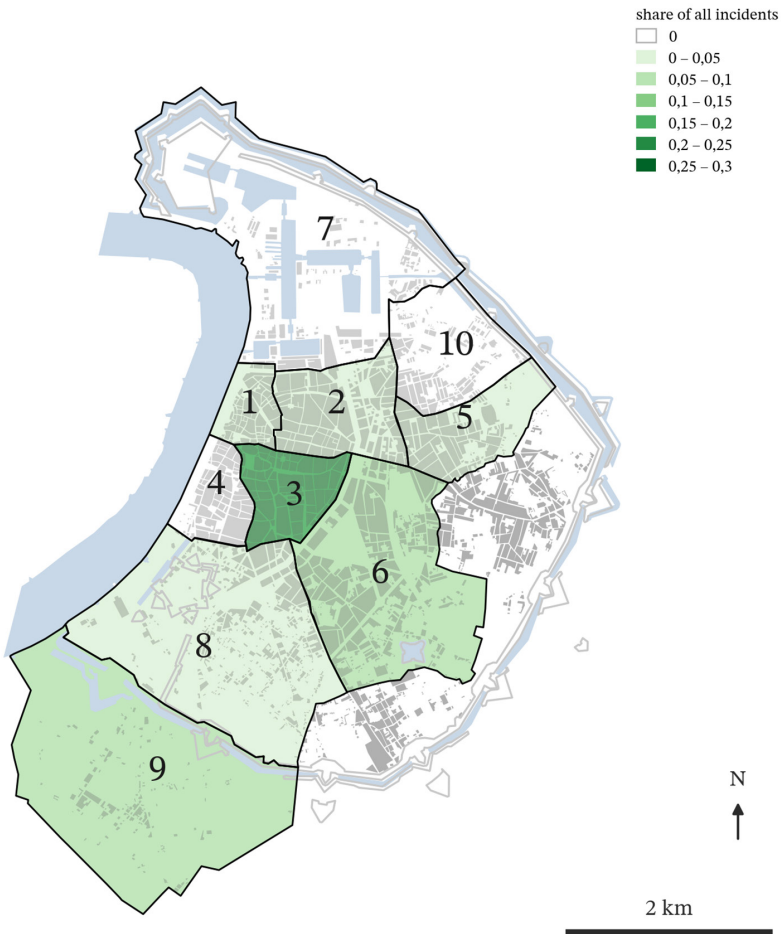


FIGURE 5a: Map of Relative Incident Count for Each District, 1876–1920.

It is challenging to determine whether the higher incident count for a particular district reflects the preservation conditions of the incident books or heightened police activity. Among all districts, the third district is the most represented in the dataset: not only does it account for the highest share of incidents (35%), but it is also the most consistently present over the years

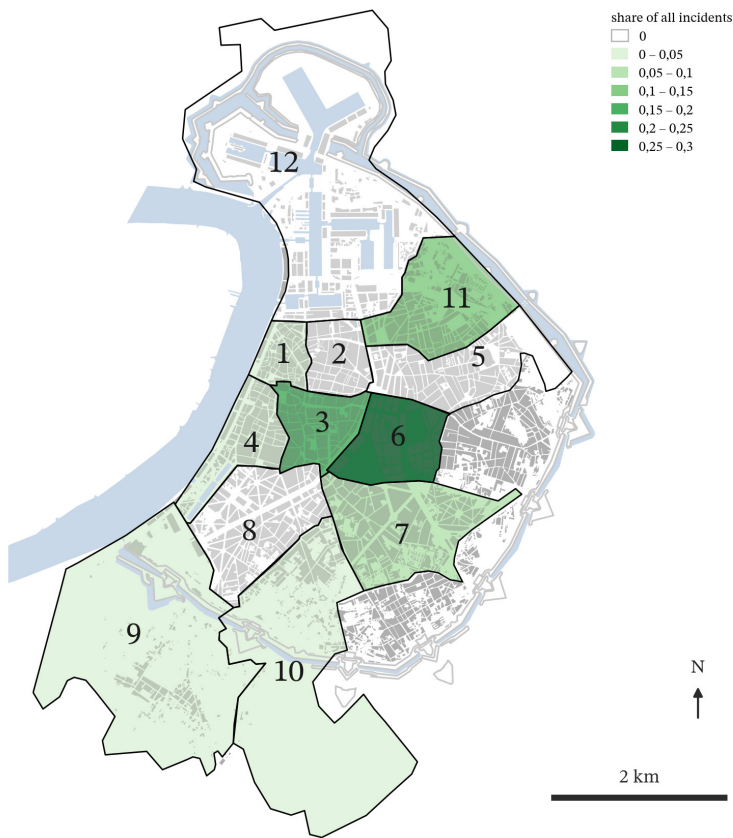


FIGURE 5b: Map of Relative Incident Count for Each District, 1921–1945.

(see Figure 6a and 6b). Its prominence could be attributed to its dual commercial and bourgeois-residential functions. The police likely prioritized this area to safeguard Antwerp's shopping image—home to the city's first department stores and the most fashionable shopping streets—and, to monitor the boundaries of bourgeois respectability (with smaller working-class ghettos situated between predominantly bourgeois areas) (De Koster & Vrints, 2020). The sixth district has the second highest share of incidents (31.5%; including the seventh district which used to be in the same area before 1921 even 39.6%), and is represented in the dataset from 1908 onwards, with a peak during the 1930s and the Second World War (see Figure 6b). This neighborhood was characterized by the presence of hotels, bars, and dancehalls, as well as the train station, the diamond industry, and a significant Jewish population. Possible explanations for the peak during the late 1930s and WWII could be the increasing number of motorized vehicles or rising antisemitism, but further research is necessary.

TABLE 4: Incident Count for Each District Before and After 1921

district	incident count 1876-1920 (absolute)	incident count 1876-1920 (relative)	incident count 1921-1944 (absolute)	incident count 1921-1944 (relative)
1	16,171	.040	2,676	.007
2	1,508	.004	N/A	N/A
3	64,445	.158	76,979	.189
4	N/A	N/A	5,717	.014
5	3,296	.008	N/A	N/A
6	24,688	.061	103,392	.254
7	N/A	N/A	33,090	.081
8	2,064	.005	N/A	N/A
9	20,903	.051	8,892	.022
10	N/A	N/A	1,442	.004
11	N/A	N/A	41,439	.102
12	N/A	N/A	N/A	N/A
total	133,075	.327	273,627	.673

3.4. Language

The incident books’ main language is Dutch, but they occasionally feature French passages because of Belgian language policies at the time. We used the Python library LangID (Lui & Baldwin, 2012) to detect the relative presence of French in the HTR output (see Figure 7). The values are not flawless, since the HTR is not perfect and language identifiers for Dutch are usually not adapted to the large number of French words in Flemish, the version of Dutch spoken in Belgium. Nevertheless, the results can be used to broadly estimate the evolution of the French-Dutch ratio in the incident books over time. In Figure 7, a clear influence from the Belgian Law of Language Equality is notable, which in 1898 elevated Dutch to the same status as French as an official language in Belgium (Witte & van Velthoven, 2010).

3.5. HTR Quality: Confidence Score

In the XML output, *kraken* provides a confidence score (CS) in the 0–1 range for each individual character, which represents the model’s level of certainty about the accuracy of its recognition results. We have calculated the average confidence score for each page and each book to measure the quality of the HTR (see “metadata” folder in repository). The majority of the CSs cluster

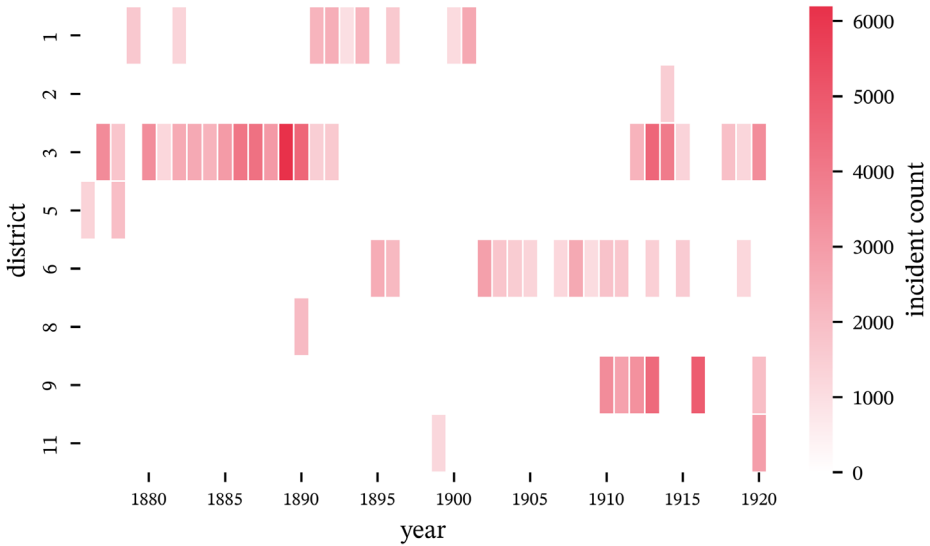


FIGURE 6a: Heatmap of Incident Count by Year and District, 1876–1920.

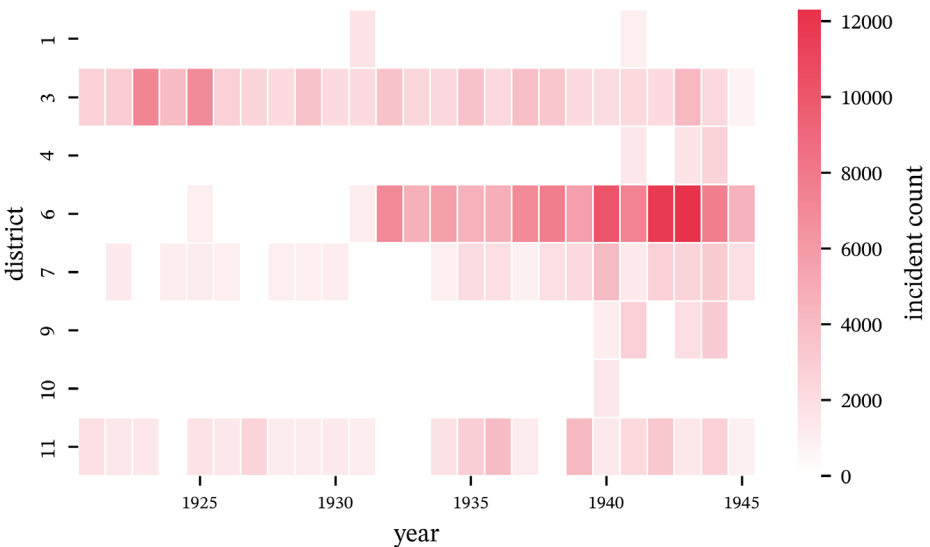


FIGURE 6b: Heatmap of Incident Count by Year and District, 1921–1944.

around an average of .97 (see Figure 8 and Table 5). Manual inspection has revealed that pages with the lowest scores (below .90) are unusable (empty, upside-down, printed instead of handwritten letters, et cetera). A number of pages with a CS between .90 and .96 (first quartile) have an unusual layout (e.g. table), content (e.g. personnel administration), or handwriting, but still

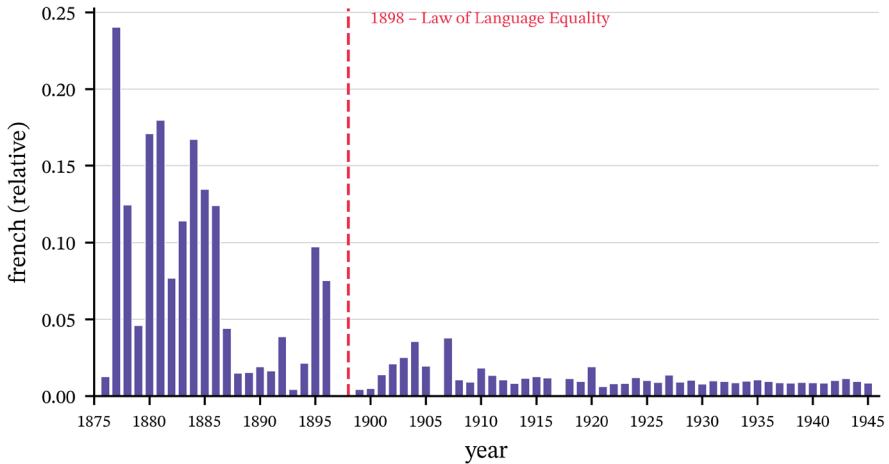


FIGURE 7: Barplot of (Relative) Proportion of French in the Incident Books Between 1876 and 1945 Estimated With LangID. The LangID Python library calculates language percentages of a text by using a Naïve Bayes classifier trained on n-grams (contiguous sequences of characters or words) from a large, multilingual corpus to predict the most probable languages.

contain passages that are useful for further analysis. If we take a look at the average CS for each book, we notice that the results are worse for the period before 1900 (see Figure 9). This could be attributed to the higher proportion of French or less uniform handwriting styles.

4. Using the Dataset

4.1. Structure of the Dataset

/scripts

- /scripts/requirements.txt: all code was applied and tested using Python (version 3.11.10)
- /scripts/postprocess-xml.ipynb: scripts used to postprocess the raw xml files after layout analysis and HTR of the manuscript images
- /scripts/enrich-metadata.ipynb: scripts used to analyze the postprocessed xml files and enrich the metadata of each incident book
- /scripts/visualize-metadata.ipynb: scripts used to make visualizations of the analyses of the metadata

/data

- /data/xml-raw: folder with the raw xml files after layout analysis and HTR of the manuscript images (organized by incident book)
- /data/xml-processed: folder with the post-processed xml files created by the scripts in /scripts/postprocess-xml.ipynb (organized by incident book)

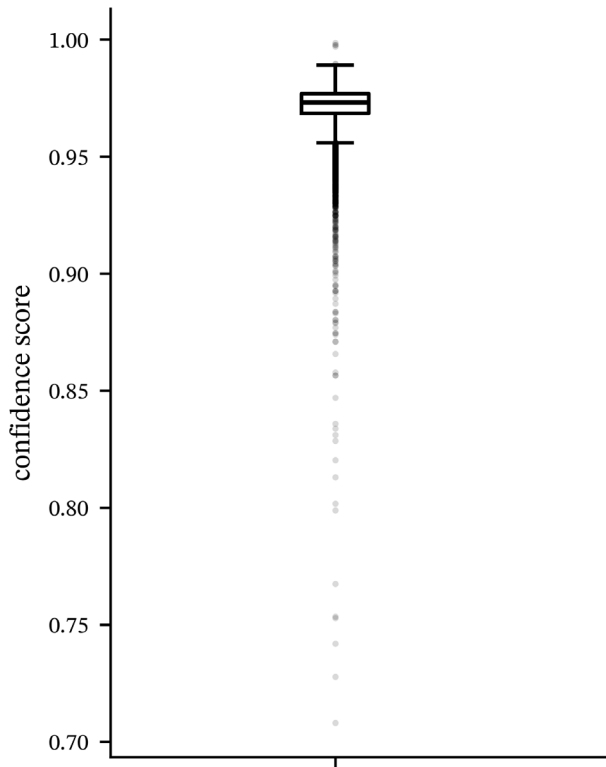


FIGURE 8: Boxplot of Confidence Scores (average per page).

TABLE 5: Summary Statistics Confidence Scores (Average Per Page)

mean	.9716
std	.0081
min	.7079
25%	.9683
50%	.9731
75%	.9767
max	.9986

- /data/IB-metadata: folder with csv files of the original and enriched meta-data of the incident books
- /data/visualizations: folder with png images created by the scripts in /scripts/visualize-metadata.ipynb

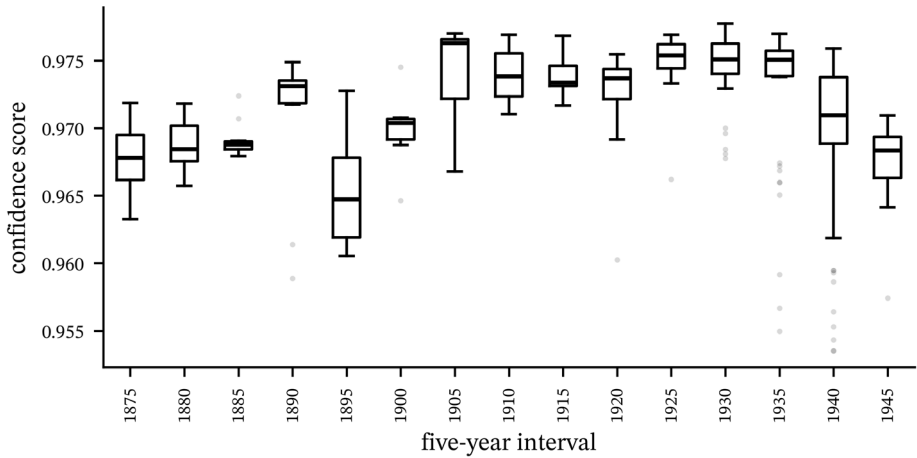


FIGURE 9: Boxplot of Average Confidence Scores for Each Book Aggregated for 5-Year Intervals.

4.2. Limitations of the Dataset

The above exploratory study of the data has revealed some structural inconsistencies that should be taken into account for its analysis. First, the LA and HTR quality is not perfect. Even small segmentation and character errors can skew research results. Second, as outlined above, the data is not evenly distributed across time and place. Some neighborhoods and periods are better represented than others. Lastly, the presence of both Dutch and French requires sophisticated natural language processing techniques to accurately interpret and analyze the text, further complicating the data analysis process.

In addition to the structural limitations of the dataset, the incident books are subject to several inherent source-critical remarks independent of the digitization process, but which must be considered when using this source for historical research. First, the incident books only reflect the point of view of the policemen. Because these were not official reports, the people involved didn't proofread or sign the description of the events. Even when the reports included citations from the lawbreakers, victims, or witnesses, these were always paraphrased by the police officer and fitted into legal-administrative jargon. Furthermore, all the information that the officer deemed irrelevant to the criminal context was omitted. Apart from the distortion caused by the officer, all the parties involved tried to present their motivations and actions in the most favorable way possible, which inevitably gave way to a certain level of "fiction in the archives" (Davis, 1997).

The events mentioned in the incident books are not representative of the experience of street life in general. Despite the high density of policemen

in Antwerp, most of the inhabitants probably lived their lives without ever interacting with the local officers of their respective neighborhoods. During their daily rounds, the police officers often remained anonymous or unnoticed. After the introduction of motorized vehicles during the interwar years, their contact with the community became even more distant (De Koster, 2011). Even though the incident books record more and smaller events than the official reports, the reported events were always related to a certain criminal or safety concern. This could create the false impression that the streets of Antwerp were a dangerous or disturbing place (Pooley, 2021).

On top of that, the incident books fall short in truthfully representing criminal street life as well. Dark numbers are a persistent methodological question in historical as well as present-day criminology (Archer, 2008; Karsdorp et al., 2024; Taylor, 1998; Williams, 2000). Not only did many illegal acts go unnoticed, the police often adopted non-official means to resolve conflicts as well. The police's dependence upon the population to maintain public order forced them to apply pragmatic solutions that were not necessarily recorded in the police archives. Additionally, their selective policing strategies (whether consciously or not) inevitably led to a biased representation of the facts. By targeting specific demographic groups, neighborhoods, or forms of behavior, the incident books inform us less about an objective, actual crime reality than about prevalent bias and stereotypes (De Koster & Vrints, 2020).

4.3. Suggestions for Future Research

When its shortcomings are taken into account, the incident books can be valuable for various fields of historical research. Police historians can find evidence of street-level police practices and police-public relations. The incident books shed light on the strategies local police officers used for law enforcement and how they had to negotiate with the community to maintain public order (e.g. Klein, 2010; Churchill, 2014). Another potential area of study is the history of (petty) crime. By examining minor offenses and everyday transgressions, historians can gain insight into the nature of urban crime and changing societal attitudes towards deviant behaviors (such as drunkenness, vagrancy, littering, indecency, noise pollution, vandalism, et cetera).

Because maintaining public order in a large city encompasses many aspects of daily life, the police records contain information relevant to historians of all kinds: urban mobility, leisure activities, street vending, animal welfare, children and adolescents, urban infrastructure, emotional and sensory experiences, and so on. On top of that, the abundance of personal details is invaluable for social historians of the city. Thanks to the fairly consistent registration of the gender, class, age, origin, and residence of the Antwerp

streetwalkers, all of the above research domains can be studied from a diversity perspective.

The aforementioned limitations of the dataset make comprehensiveness impossible. While the dataset does not offer representativeness over the entire city or period, it contains a wealth of information for specific neighborhoods and shorter contiguous periods, allowing for a sample-based data-driven approach. Additionally, police bias against the working class can be advantageous for historians who adopt a “from below” approach. Given that the illiterate classes left few records of their own, the incident books provide an exceptional window into their lives. The extensive detail and volume of the dataset offer unique opportunities for data-driven research, making it a valuable resource for uncovering the lived experiences of marginalized groups in the modernizing city.

Finally, it is important to note that this dataset was produced before the recent rise of large language models (LLMs). At the time of writing, we are experimenting with LLMs to further clean and improve the existing HTR output. Because such models can draw on contextual and linguistic cues that fall outside the scope of a purely character-based recognition engine, they hold promise for correcting recurrent segmentation and character errors, as well as for normalizing the historical and bilingual spelling variation that complicates text analysis (Boros et al., 2024; Kim et al., 2025). While these efforts are still ongoing, they suggest that the quality and usability of the dataset can be enhanced further without a costly re-transcription of the source material.

Acknowledgements

Many thanks to Nooshin Shahidzadeh Asadi for processing the incident books data for the Necturus XML viewer (Shahidzadeh Asadi, 2025), to Jef Peeraer for setting up the website, to Thibault Clérice for the layout analysis and to the Antwerp City Archives for providing the archive files.

References

- Archer, J. (2008). Mysterious and suspicious deaths: Missing homicides in North-West England (1850–1900). *Crime, History & Societies*, 1, 45–63.
- Blondé, B., Greefs, H., & Clark, P. (2005). The growth of urban industrial regions: Belgian developments in comparative perspective, 1750–1850. In J. Stobart & N. Raven (Eds.), *Towns, regions and industries. Urban and industrial change in the Midlands, c. 1700–1840* (pp. 211–227). Manchester University Press.
- Boros, E., Ehrmann, M., Romanello, M., Najem-Meyer, S., & Kaplan, F. (2024). Post-correction of historical text transcripts with large language models: An exploratory study. In *Proceedings of the 8th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL 2024)* (pp. 133–159). Association for Computational Linguistics.

- Chagué, A., Clérice, T., Norindr, J., Humeau, M., Davoury, B., Van Kote, E., Mazoue, A., Faure, M., Doat, S. (2023). Manu McFrench, from zero to hero: impact of using a generic handwriting recognition model for smaller datasets. In *Digital Humanities 2023: Collaboration as Opportunity*. Alliance of Digital Humanities Organizations, University of Graz. <https://inria.hal.science/hal-04094241>
- Churchill, D. (2014). 'I am just the man for upsetting your bloody bobbies': popular animosity towards the police in late nineteenth-century Leeds. *Social History*, 39(2), 248–266. <https://doi.org/10.1080/03071022.2014.912424>
- Clérice, T. (2023). You Actually Look Twice At it (YALTAi): Using an object detection approach instead of region segmentation within the Kraken engine. *Journal of Data Mining & Digital Humanities*. <https://doi.org/10.46298/jdmdh.9806>
- Davis, N. Z. (1997). *Fiction in the archives: Pardon tales and their tellers in sixteenth-century France*. Stanford University Press.
- De Koster, M. (2011). Politieoptredens en het dagelijks ordenen van de Stad Antwerpen, eind negentiende—Begin twintigste eeuw. In A. Winter, M. De Koster, B. De Munck, H. Greefs, & B. Willems (Eds.), *Werken aan de stad: Stedelijke actoren en structuren in de Zuidelijke Nederlanden, 1500–1900: Liber alumnorum Catharina Lis en Hugo Soly* (pp. 251–268). VUBPress.
- De Koster, M., & Vrints, A. (2020). The new police as agents of class control?: Urban policing and its socio-geographical focus in nineteenth-century Antwerp. In B. Blondé, S. Geens, H. Greefs, W. Ryckbosch, T. Soens, & P. Stabel (Eds.), *Inequality and the city in the Low Countries (1200–2020)* (pp. 345–358). Brepols.
- Fickers, A. (2012). *Towards a new digital historicism? Doing history in the age of abundance*. <https://mediarep.org/handle/doc/15001>
- Greefs, H. (2023). Burgerwetenschap in stadsgeschiedenis. Reflecties vanuit de historische praktijk. *Stadsgeschiedenis*, 18(2), 79–132.
- Hitchcock, T., & Shoemaker, R. (2006). Digitising history from below: The Old Bailey Proceedings online, 1674–1834. *History Compass*, 4(2), 193–202. <https://doi.org/10.1111/j.1478-0542.2006.00309.x>
- Karsdorp, F., Kestemont, M., & de Koster, M. (2024). Beyond the register: Demographic modeling of arrest patterns in 1879–1880 Brussels. *Proceedings of the Computational Humanities Research Conference 2024*, 265–281.
- Keijser, L. (2020). 6000 ground truth of VOC and notarial deeds 3.000.000 HTR of VOC, WIC and notarial deeds (Version 5.0) [Dataset]. Zenodo. <https://doi.org/10.5281/zenodo.4159268>
- Kiessling, B., Tissot, R., Stokes, P., & Stökl Ben Ezra, D. (2019). eScriptorium: An open source platform for historical document analysis. *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, 2, 19–19. <https://doi.org/10.1109/ICDARW.2019.10032>
- Kim, S., Baudru, J., Ryckbosch, W., Bersini, H. & Ginis, V. (2025). Early evidence of how LLMs outperform traditional systems on OCR/HTR tasks for historical records. *arXiv*. <https://doi.org/10.48550/arXiv.2501.11623>
- Klein, J. (2010). *Invisible men: the secret lives of police constables in Liverpool, Manchester, and Birmingham, 1900–1939*. Liverpool University Press.
- Laite, J. (2020). The Emmet's Inch: Small history in a digital age. *Journal of Social History*, 53(4), 963–989. <https://doi.org/10.1093/jsh/shy118>
- Lefranc, L., Van Damme, I., Clérice, T., & Kestemont, M. (2024). ARletta. Open-source handwritten text recognition models for historic Dutch. *Journal of Open Humanities Data*, 10, 43. <https://doi.org/10.5334/johd.225>

- Lui, M., & Baldwin, T. (2012). langid.py: An Off-the-shelf language identification tool. In M. Zhang (Ed.), *Proceedings of the ACL 2012 System Demonstrations* (pp. 25–30). Association for Computational Linguistics. <https://aclanthology.org/P12-3005>
- Muehlberger G. et al. (2019). Transforming scholarship in the archives through handwritten text recognition: Transkribus as a case study. *Journal of Documentation*, 75(5), 954–976.
- Pooley, C. (2021). On the street in nineteenth-century London. *Urban History*, 48(2), 211–226.
- Saerens, L. (2000). *Vreemdelingen in een wereldstad: Een geschiedenis van Antwerpen en zijn joodse bevolking (1880–1944)*. Lannoo.
- Saygi, G., & Yasunaga, M. (2021). The digital urban experience of a lost city using mixed methods to depict the historical street life of Edo/Tokyo: Using mixed methods to depict the historical street life of Edo/Tokyo. *Magazén*, 2, 193–224.
- Shahidzadeh Asadi, N. (2025). eXtant-CMG/Necturus-Viewer-Compact: v0.1.0. *Zenodo*. <https://doi.org/10.5281/zenodo.15472466>
- Taylor, H. (1998). The politics of the rising crime statistics of England and Wales, 1914–1960. *Crime, History & Societies*, 2(1), 5–28.
- Thompson, E. P. (1963). *The making of the English working class*. Victor Gollancz Ltd.
- van den Heuvel, D., Pierik, B., Amaro, B. V., & Kisjes, I. (2020). Capturing gendered mobility and street use in the historical city: A new methodological approach. *Cultural and Social History*, 17(4), 515–536.
- Williams, C. (2000). Counting crimes or counting people: Some implications of mid-nineteenth century British police returns. *Crime, History & Societies*, 4(2), 77–93.
- Witte, E., & van Velthoven, H. (2010). *Strijden om taal: De Belgische taalkwestie in historisch perspectief*. Pelckmans.
- Wolf, E. (1982). *Europe and the people without history*. University of California Press.