



MASTERARBEIT | MASTER'S THESIS

Titel | Title

Translate and reprint - איבערזעצן און איבערדריקן
A translation and digital edition of a Yiddish newspaper

verfasst von | submitted by
Robin Luger BA

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Master of Arts (MA)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree
programme code as it appears on the
student record sheet:

UA 066 647

Studienrichtung lt. Studienblatt | Degree
programme as it appears on the student
record sheet:

Masterstudium Digital Humanities

Betreut von | Supervisor:

Univ.-Prof. Dr. Gabriel Viehhauser

Acknowledgements

This project would not have been possible without the help and support I have received over the past two years.

First, I want to thank my supervisor Univ.-Prof. Dr. Gabriel Viehhauser for his continuous feedback throughout this process. My sincere gratitude also goes out to Assoz. Prof. Mag. Mag. Dr. Börries Kuzmany for teaching me to work with Yiddish sources in the first place, for providing resources as well as further literature, and for his detailed feedback. I also want to thank Dimitra Grigoriou, BA MA MSc and Carl Friedrich Haak, B.A. MA who helped me set up the website during my internship at the Center for Digital Humanities.

Finally, I want to thank my friends and especially my partner who kept me motivated to stay on track to finish my master's.

Zusammenfassung

Diese Masterarbeit ist eine born-digital Übersetzung einer jiddischen Zeitung ins Englische, die annotiert und als wissenschaftliche digitale Edition kontextualisiert wurde. Der Fokus liegt auf der Erstellung offener, auffindbarer, zugänglicher, interoperabler und wiederverwendbarer Daten unter Einhaltung von Standards und Richtlinien. *יד דער אמת'ר יוד* – *Der Wahre Jude* erschien 1904 in Lemberg, Ukraine, ein damaliges Zentrum für jüdisches Leben und der jiddischen Sprache. Jiddische Zeitungen und Literatur waren maßgeblich daran beteiligt, die Sprache als legitime jüdische Sprache zu etablieren und waren damit ein Pfeiler jüdischer Kultur. Die Edition ist auf einer statischen, Open-Source Website im semantischen Web frei einsehbar. Die Übersetzung ins Englische bietet einen Einblick in die jüdischen kulturellen, religiösen und sozialen Verhältnisse und Traditionen in Osteuropa zu Beginn des 20. Jahrhunderts. Dadurch kann jüdisches Kulturerbe gesichert werden. Diese Masterarbeit trägt dazu bei, Quellen in Minderheitensprachen für die Forschungslandschaft und die breitere Öffentlichkeit zugänglich zu machen.

Abstract

This thesis is a born-digital translation of a Yiddish newspaper into English, annotated and contextualised as a scholarly digital edition. Special focus is put on creating open, findable, accessible, interoperable, and reusable data by following standards and guidelines. *ידע אמת'ר יוד* – *The Genuine Jew* was published in 1904 in Lviv, Ukraine, which was a major centre of Jewish life and the Yiddish language. Yiddish press and literature was substantial for establishing the language as a legitimate Jewish language and were a pillar of Jewish culture. The edition is displayed on a static, open-source website, situated in the semantic web. The English translation provides insight into Jewish cultural, religious, and social circumstances and traditions in Eastern Europe at the beginning of the 20th century. This helps to preserve Jewish cultural heritage. This master project is an effort to bring minority language primary sources to the wider research and non-research sphere alike.

Table of Contents

1 Introduction.....	1
2 Source Material.....	3
2.1 Context of the Newspaper.....	5
2.2 Content of the Newspaper.....	7
3 Yiddish Press.....	15
4 Yiddish.....	19
4.1 Yiddish in translation.....	22
5 Translation Process.....	24
5.1 Transliteration.....	24
5.2 Translation.....	25
5.3 Translating Named Entities.....	27
5.3.3 People.....	27
5.3.4 Places.....	28
6 Challenges of Automatisation.....	31
6.1 Transcription.....	31
7 Scholarly Digital Edition.....	35
8 Digital Curation.....	37
8.1 Linked Data.....	37
8.1.1 Metadata.....	39
8.1.2 Linked Open Data.....	40
8.2 FAIR Data Principles.....	40
8.2.1 Findable.....	41
8.2.2 Accessible.....	42
8.2.3 Interoperable.....	43
8.2.4 Reusable.....	44
9 Scholarly Annotation Methodology.....	47
9.1 Encoding Rationale.....	47
9.2 Encoding Scheme.....	48
9.3 Annotation Methodology.....	51
9.3.1 Layout.....	52
9.3.2 Linguistic Characteristics.....	53
9.3.3 References.....	53
9.3.4 Named Entities.....	55
9.3.5 Temporal Information.....	58
9.4 Index Structure.....	59

9.5 Challenges in Annotation.....	64
10 Transformation Pipeline.....	67
10.1 Cookiecutter.....	68
10.2 Transformation Script.....	68
10.2.1 Formatting and Layout.....	69
10.2.2 Named Entity Reference.....	73
10.2.3 Metadata Implementation.....	75
10.3 Directory Structure.....	78
10.4 Command Line Interface.....	84
10.5 Frontend Implementation.....	85
11 Digital Infrastructure.....	86
11.1 Navigation.....	86
11.2 Search Function.....	89
11.3 Infrastructure and Hosting.....	91
12 Conclusion.....	93
Bibliography.....	95
Table of Figures.....	104
Index of Tables.....	105

1 Introduction

Yiddish Literature, poems, Memorbuchs, playscripts, and other creative writing has often been subject of translations¹, but factual texts like newspapers mostly remain available only to those who understand the language and can read the script.

However, Yiddish periodicals are vital to understand the culture and politics of Eastern European Jews before World War I. This master's project therefore aims to translate and transform a Yiddish source so far only available only in printed form into an English open-source, interoperable and linked dataset, accessible to a wider audience for further research. This approach aims to then make this dataset accessible to a wider audience for further research

During this process, I set out to explore possibilities and obstacles as well as feasible solutions of adapting a minority language text written in non-Latin script into a scholarly digital edition hosted on a website. The methodology used in this project was originally developed for more established languages, making this thesis a sort of prototype in this regard. Furthermore, tools developed for processing Yiddish sources were explored and their current capabilities tested. Since the majority of the translation and digitisation of the source material was carried out in 2024 and early 2025, only tools available up to this point could be taken into consideration. In addition to this theoretical exploration, the project deliverables are currently saved in a remote repository² and displayed on an interactive, static website³.

-
- 1 'Latest Yiddish Translations, 2024', In Geveb, accessed 8 March 2026, <https://ingeveb.org/blog/latest-yiddish-translations-2024>; 'Yiddish Women in Translation', The New York Public Library, accessed 8 March 2026, <https://www.nypl.org/blog/2020/08/17/other-words-yiddish-women-translation-month>; 'Yiddish in Translation | Yiddish Book Center', accessed 8 March 2026, <https://www.yiddishbookcenter.org/discover/yiddish-translation>.
 - 2 lug-robin, *Lug-Robin/Tgj-Data*, XSLT, 24 February 2024, released 11 February 2026, <https://github.com/lug-robin/tgj-data>.
 - 3 Robin Luger, 'The Genuine Jew', The Genuine Jew, accessed 14 February 2026, <https://lug-robin.github.io/tgj-data/>.

The thesis is structured in two parts: the first half contains a historical overview including the context of the source material, Yiddish press, as well as information about the development of the Yiddish language. First, the source material is introduced, exploring context and content of the publication before reviewing the development of Yiddish press, the Yiddish language and its history as well as translation practices. In this part I will describe the technical requirements and transformation workflow. This is followed by possible challenges an automatisisation process can entail leading into the digital methods relevant for this project. Here, I give an overview of scholarly digital editions and digital curation in the open web incorporating guidelines for open access digital archiving and publishing. Finally, I will go over the digital annotation methodology and code structures of the annotation documents, transformation files and the website structures incorporating the web hosting technicalities as well before concluding the thesis.

2 Source Material

דער אמת'ר יוד – *Der Wahre Jude*⁴ was a Yiddish newspaper⁵ that was published in L'viv⁶, today's Ukraine – which was part of the Austro-Hungarian Empire in Galicia until 1918. The first issue was published on February 5th, 1904, and at least 44 numbers were printed and published. The Austrian National Library holds 31 of these papers, all of which are scanned and uploaded to ANNO (**A**ustria**N** **N**ewspaper **O**nline). *Der Emes'r Yud* is one of four newspapers written in Yiddish available on the ANNO database and thus one of the four possible subjects of this project. The other three were all published in Kraków: *Der Jude* (1899-1902)⁷, *Der Jud* (1905-1906), and *Jüdische Mittagszeitung* (1907). I decided to work with *Der Emes'r Yud* because of the quantity and languages of the other papers. The *Jüdische Mittagszeitung* only covers about two months in total and mainly features local articles. On the other side, *Der Jude* covers four years with over 2,500 pages in total. Between *Der Jud* and *Der Emes'r Yud* is not a lot of difference in regard to themes and topics, quantity and time frame – 200 pages of *Der Jud* is available on ANNO. The reason why I chose *Der Emes'r Yud* in the end is because the non-Yiddish text in the paper is German – which I understand – as opposed to Polish – which I do not understand – in *Der Jud*. In addition to the scans, the library also provides an automated OCR transcription in a separate text document. However, the transcription is highly flawed and can only be used – if at all – to cross-check manual transcriptions. Up until the 20th number,

4 Osias H. Wilf, *Der Emes'r Yud, Organ für alle jüdischen Angelegenheiten* (L'viv, 1904; repr., Lemberg: Salat, 1904), <http://anno.onb.ac.at/cgi-content/anno?aid=emj>.

5 The paper will be referred to in the romanised spelling of its Yiddish title: *Der Emes'r Yud*.

6 The city has different names depending on the language and timeframe: Lviv (English), L'viv (Ukrainian), Lwów (Polish), Lemberg (German), and L'vov (Russian). In this thesis, it will be called L'viv following the Ukrainian spelling.

7 Roni Masel has written about *Der Jude*/*Der Yid* in detail, see Roni Masel, 'Who Is a Yid? Reading the Journal *Der Yid* beyond the Hebraist – Yiddishist Binary', *Journal of Modern Jewish Studies* 20, no. 3 (2021): 361–83, <https://doi.org/10.1080/14725886.2020.1809127>.

the library's uploads are complete, but numbers 21-22, 25-26, 35-40, and 42-43 are missing. These 31 available newspapers each contain four to twelve pages of text, in sum amounting to 218 pages of the scanned source. The 44th issue, published October 21st 1904 does not include a statement announcing the end of the newspaper or any explicit mention of a low budget. Therefore, it is possible that there are more papers of *Der Emes'r Yud* published and printed than those available in the Austrian National Library. Unfortunately, the paper yields no search results, nor is it indexed in any other library or archive catalogues. The National Library of Israel⁸ has catalogued over 2000 individual Jewish historic newspapers – half of those available to read – but *Der Emes'r Yud* is not indexed in this catalogue as well. Figure 1 visualises the available papers and a temporal estimation when those not available were printed and sold. It was not unusual for Hebrew newspapers in the Habsburg monarchy to suddenly end after one or two years. The first Yiddish newspaper published in Galicia in 1848/49 was discontinued due to lacking monetary funds as the publisher expected several thousand subscribers, only to attain close to 200 in two years⁹. It is therefore unclear whether more papers were produced after October 1904.

8 'The National Library Newspaper Collection', accessed 3 February 2026, <https://www.nli.org.il/en/discover/newspapers>.

9 Rachel Manekin, 'Die hebräische und jiddische Presse in Galizien', in *Politische Öffentlichkeit und Zivilgesellschaft, Teilbd. 2, Die Presse als der Faktor der politischen Mobilisierung*, with Kommission für die Geschichte der österreichisch-ungarischen Monarchie 1848-1918, Die Habsburgermonarchie 1848-1918 (Wien: Verlag der Österreichischen Akademie der Wissenschaften, 2006), 8:2361.

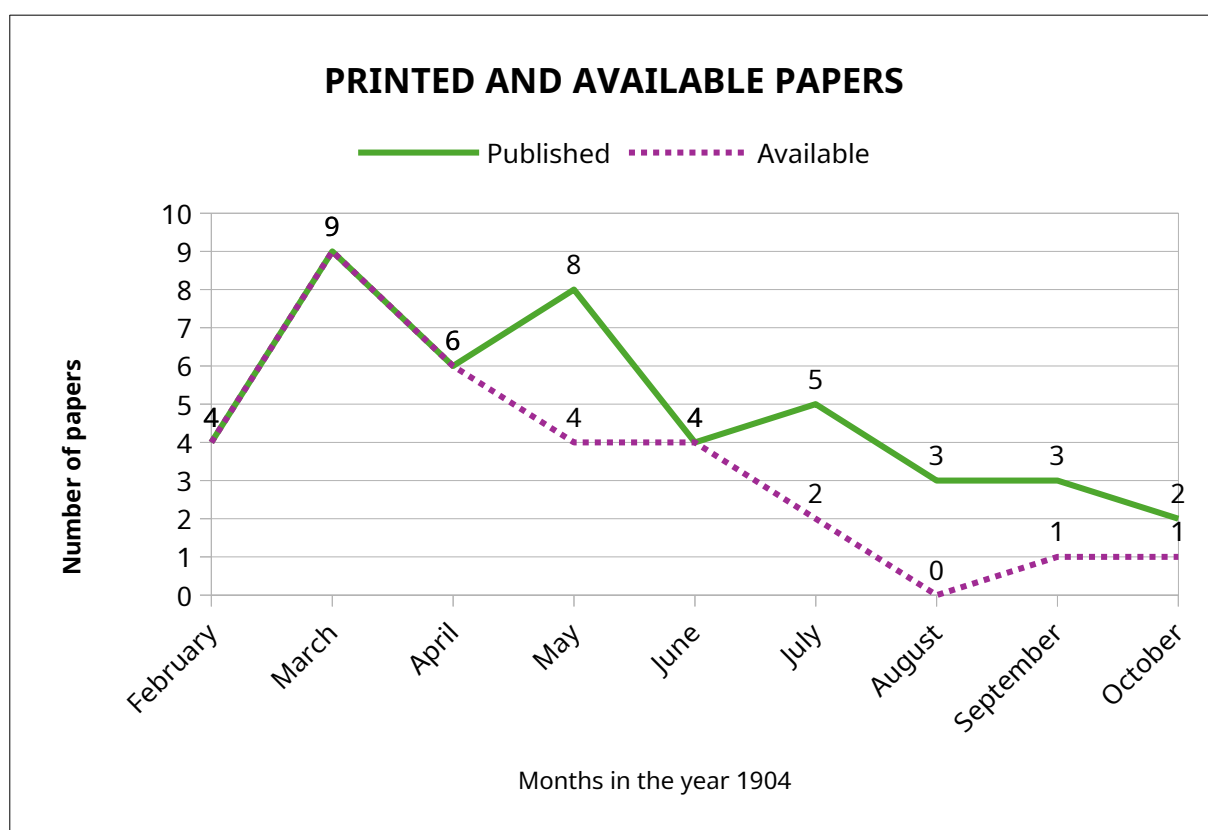


Figure 1: Printed and available papers of "Der Emes'r Yud"

For this thesis, I have chosen to focus on the first two months, February and March. During this period, thirteen papers were printed, amounting to 88 pages of text.

2.1 Context of the Newspaper

Der Emes'r Yud was published in L'viv, an important center of Jewish life in the Habsburg Empire. It was the capital of the crown land Galicia and Lodomeria which also had the biggest Jewish population in Austria-Hungary¹⁰. With a 30 % Jewish population, Yiddish at that time was the second most widely spoken language of the city, although it was not officially recognized as such¹¹. L'viv was also the fourth-

10 Miroslaw Łapot, 'Development of the Education of Galician Jews at the Turn of the 19th and 20th Centuries as Exemplified by the City of Lviv', *Scripta Judaica Cracoviensia* 18 (2021): 3, <https://doi.org/10.4467/20843925SJ.20.001.13869>.

largest city of Cisleithania, albeit one of the poorest of the region¹², not only in the Austro-Hungarian context but also in comparison to adjacent Russian Poland¹³. However, the living conditions in L'viv compared to other towns in Eastern Galicia were still better and with more educational opportunities¹⁴. L'viv was therefore fertile soil for Yiddish newspapers, as long as they were affordable and comprehensible. Jewish periodicals were commonly published throughout Europe but publishing them in Yiddish was limited to Galicia and Bucovina. In the Habsburg Empire however, Jewish periodicals were commonly written and published in Hebrew, German, Russian or Polish, since Yiddish newspapers did not survive long outside of Galicia¹⁵. Among the younger Jewish generation growing up in Galicia by the end of the 19th century, it was more common for them to be familiar with the Polish language and culture. At the same time, a part of that generation developed a Jewish National Identity for which Yiddish grew in importance not only as a Germanic dialect, but as a legitimate language. Around this time, associations to establish and strengthen Jewish autonomy and national identity as well as to promote a sense of self-esteem were founded¹⁶. One of which, the Zionist association *Machsike Hadas*,

-
- 11 Ezra Mendelsohn, 'Jewish Assimilation in L'viv: The Case of Wilhelm Feldman', in *Nationbuilding and the Politics of Nationalism: Essays on Austrian Galicia*, 3rd ed., Monograph Series / Harvard Ukrainian Research Institute (Cambridge: Harvard Univ. Press, 1995), 94–95.
 - 12 Catherine Horel, 'Midsize Cities in Austria-Hungary', in *Multicultural Cities of the Habsburg Empire, 1880-1914*, Imagined Communities and Conflictual Encounters (Central European University Press, 2023), 27, <https://doi.org/10.7829/j.ctv2kzv0mg.8>.
 - 13 Leila P. Everett, 'The Rise of Jewish National Politics in Galicia, 1905 - 1907', in *Nationbuilding and the Politics of Nationalism: Essays on Austrian Galicia*, 3rd ed., Monograph Series / Harvard Ukrainian Research Institute (Cambridge: Harvard Univ. Press, 1995), 149.
 - 14 Ela Bauer, 'Commemoration and Jewish National Identity: Lviv in the Last Decades of the Nineteenth Century', *Journal of Jewish Identities* 14, no. 1 (2021): 2–3.
 - 15 Horel, 'Midsize Cities in Austria-Hungary', 316; Manekin, 'Die hebräische und jiddische Presse', 2342.
 - 16 Bauer, 'Commemoration and Jewish National Identity', 3–5; Everett, 'The Rise of Jewish National Politics', 158.

was connected to *Der Emes'r Yud*¹⁷. By publishing in Yiddish, the language of the people, the publishers ensured that the people would understand their message although the educated strata viewed Hebrew as the real Jewish language¹⁸. While the first Jewish newspapers were written in Hebrew, it was often ideologically motivated and did not meet the contemporary journalistic standards. Additionally, the biblical metaphorical references demanded the reader to be familiar with traditional Jewish sources and thereby created further obstacles to understand the material¹⁹. Making sure their audience could understand them helped the periodicals' writers to reach the goals they set to meet by informing the public.

2.2 Content of the Newspaper

Der Emes'r Yud follows earlier Jewish periodicals in motivation and topics. The paper's goal, also according to the editor, was to inform the Jewish people who were barely or not able to understand German or Polish newspaper about current topics, news, general knowledge, and Jewish history focusing on an orthodox, Yiddish-speaking readership:

"מיר ווילען אבער זארגען ניט נאר פאר דעם פראסטען קלאס לעזער. מיר ווילען
אונזער בלאטט גאנץ בעזונדערס ווידמען אונזערע פרימע ברידער, וואס האבען
נאך נישט פערלאזען דאס אמת'ע יודישקייט, וואס זענען מחזיק בדתנו ותורתנו
הקדושה און זענען אויך זייערע קינדער מדריך בדרך הישר. אונזערע פרומע,
עהרליכע ברידער האבען ביז היינט קיין איינציגעס יודישעס בלאט געהאט, וואס
זאלל ווערין געשריבען אין זייער גייסט, לויט וייער אויפפאססונג און נאך וייערע
געדאנקען, וואס זאלל איבער אללע ענינים, וועלכע זענען נוגע לכלל ישראל,

17 Horel, 'Midsize Cities in Austria-Hungary', 318.

18 Jacob Toury, *Die Jüdische Presse im Österreichischen Kaiserreich: Ein Beitrag zur Problematik der Akkulturation 1802-1918*, Schriftenreihe wissenschaftlicher Abhandlungen des Leo Baeck Instituts 41 (Tübingen: Mohr Siebeck, 1983), 128–29; Masel, 'Who Is a Yid?', 366; Nathan Cohen, 'Yiddish Publishing Strategies in Eastern Europe (1880–1914)', *East European Jewish Affairs*, 26 February 2025, 5, <https://doi.org/10.1080/13501674.2024.2400340>.

19 Manekin, 'Die hebräische und jiddische Presse', 2341.

שרייבען און בעריכטען אין אזא שטייגער, או עט זאלל נישט פערלעטצען אונזערע
 ברידער היראים והחרדים. אוא בלאטט אין פראסט ווידישער שפראך, וואס יעדער
 זאלל עס פערשטיין, איז ביזהער נאך נישט געוועזען אונד דאס זאלל זיין אונזער
 'דער אמת' יוד".²⁰

Between Jewish, political and general news, telegrams as well as letters informing the reader about past and current conflicts, persecution, pogroms as well as trials in other countries, educational articles were strewn in throughout the publishing timeline of the paper. The dataset showed seven articles written about religious education, and in ten newspapers the history of important Rabbis through the centuries are told. Throughout the publication, further articles about health, medicine, nature, and foreign policy were printed to support general education of the readership.

The rise of Yiddish newspapers coincided with a shift in education politics. From the 1870s onwards, public schools became mandatory for Jewish children as *kheders*²¹ did not suffice for the compulsory education imposed on the entire population of the Austrian Empire, including Jewish people. Orthodox Jews partially obeyed with this new law reluctantly by sending their daughters to these public schools, while their sons remained in *kheders* and *yeshivas*²². During the 1870s and 1880s Jewish schools were established and the number of registered *kheders* shrank from 71 in 1874 to

20 However, we want to care that not only the public class reads it. We want for our paper to be particularly special for our pious brothers, which have not left behind the true Jewry, which are the religion and our Jewish teaching and also see to teach their children the truth. Our pious, honest brothers have not had a single Jewish paper until today, what should be written according to his opinion, his view and his thoughts, about all the matters, that are concerning Jewish people as a whole, writing and reporting in a way, which does not harm our ultra-orthodox brothers. This paper is written in ordinary Jewish language, which everyone should understand, which hasn't existed yet and this should be our newspaper "The Genuine Jew", Wilf, Der Emes'r Yud, Organ für alle jüdischen Angelegenheiten, 1.

21 A widespread elementary education institution, see Mordechai Zalkin, 'Heder', in The YIVO Encyclopedia of Jews in Eastern Europe (YIVO Institute for Jewish Research, 12 February 2010), <https://encyclopedia.yivo.org/article/31>.

22 in 1913²³. Orthodox Jews lamented about the public education and substituted it with educational articles in periodicals like the biographies of important rabbis. They also complained about the state of the education in articles as evident in *Der Emes'r Yud* as well.

Another objective in line with Yiddish periodicals of this time of *Der Emes'r Yud* and its editor was to form a Jewish national identity by writing about the history and current situation of the Jewish people worldwide in order to, in turn, politicise Galician Jews and create awareness of a Jewish nation outside their local interests²⁴. Especially the Russo-Japanese war, which started only three days after the first paper was published, is covered extensively. Of the available dataset there are 24 articles titled "*The war between Russia and Japan*", "*The war between Japan and Russia*", or "*The Russo-Japanese war*". In total Russia is mentioned 45 times in the articles' titles and Japan is mentioned 36 times. In addition to informing about the frontlines, the newspaper also taught its readers about the history of Japan and its connection to Judaism in a positive matter, but also about the mistreatment of Jews in Russia and the tsar's shortcomings in the war, financially and politically. The global interest is also apparent in letters sent from all over the world (featured in 14 papers) and telegrams (featured in 21 papers).

The editor of *Der Emes'r Yud* was Osias H. Wilf (Life data unknown). *Der Emes'r Yud* was not the only newspaper Wilf authored and published. He published other newspapers in Hebrew, for example one archived by the University of Potsdam²⁵, he

22 An educational institution for young men to study the Torah see Elchanan Reiner, 'The Yeshiva before 1800', in The YIVO Encyclopedia of Jews in Eastern Europe (YIVO Institute for Jewish Research, 12 February 2010), <https://encyclopedia.yivo.org/article/47>.

23 Łapot, 'Development of the Education', 9–12.

24 Joshua Shanes, *Diaspora Nationalism and Jewish Identity in Habsburg Galicia*, 1st ed. (Cambridge University Press, 2012), 191, <https://doi.org/10.1017/CBO9781139013642>.

25 Osias H. Wilf, *ha- Boneh: ' iton yotse be-arba' ah peraḳim ba-shanah le-' inyene ha-dat, ha-yahadut, ha-ḥinukh ve-ha-ḥayim* (Lemberg: A. Goldman, 1904), <http://digital.ub.uni-potsdam.de/503435>.

co-authored a Hebrew-German-Polish dictionary²⁶ and worked as a publisher for other authors. Before *Der Emes'r Yud*, Wilf was the editor of *Machsike Hadas* (מחזיקי הדת) and *Kol Machsike Hadas* (קול מחזיקי הדת), both organs of the Orthodox party *Machsike Hadas*²⁷ (Makhzikey ha-Das, Strengthening of Faith). This political party was founded in 1870 during the rise of Jewish societies and parties, expressions of the modernisation efforts in Jewish communities²⁸. After launching *Der Emes'r Yud*, the leader of the party *Machsike Hadas* fired Wilf from his position as editor of the paper *Machsike Hadas* for political differences in spring of 1904²⁹, mere weeks after *Der Emes'r Yuds* debut. From the fifth number on, a footnote on the first page of each paper read:

"אז וי"א זאלען פון היינט אן אלע בריעף און געלדזענדונגען שיקען נאר אויף דער
אדרעססע: OSIAS H. WILF, LEMBERG Sonnengasse 27".³⁰

which indicates a clean break between Wilf and his previous occupation. The termination might be one of the reasons why his new periodical was then published twice a week, instead of once, since he had more time to focus on his new enterprise.

Der Emes'r Yud was not exclusively written by Wilf. In total, only 24 articles are signed by a named author. Out of these signatures, ten have written their full name, three only provided their initials, and the other eleven remained rather anonymous

26 Abraham Robinsohn and Osias H. Wilf, *Neues hebräisches Wörterbuch: verfasst von hervorragenden und besten Schriftstellern = Millon ibrihadas me-ibrit le-germanit Nosafot we-hasamtot* (Stanislau, Lemberg: E. Salat, 1912), <https://repozytorium.biblos.pk.edu.pl/resources/40492>.

27 Shanes, *Diaspora Nationalism*, 189.

28 Łapot, 'Development of the Education', 8.

29 Shanes, *Diaspora Nationalism*, 191.

30 *We ask the subscribers of "Machsike Hadas" and "Kol Machsike Hadas" to send every letters and cash remittance from now to to OSIAS H. WILF, LEMBERG Sonnengasse 27", Wilf, Der Emes'r Yud, Organ für alle jüdischen Angelegenheiten, 1 March 1904, 1.*

by writing under a synonym, like “the observer”, “a Lvivian”, “the truth”, or “one of the Haredim”. It stands out that all articles about local politics – the cultus election in L’viv – and religion were submitted anonymously, perhaps to avoid public scrutiny for their opinion or wanting to distance themselves from a newspaper written in Yiddish. For some educated Jewish the use of Yiddish could still hold shame. The other topics seem to be less sensitive as the part of named writers surpasses anonymous writers in every other instance. The anti-Russian sentiment of the newspaper and society is mirrored in the non-existing need for the authors to remain anonymous as shown in Figure 2. These articles are more opinion-based than objective news reports.

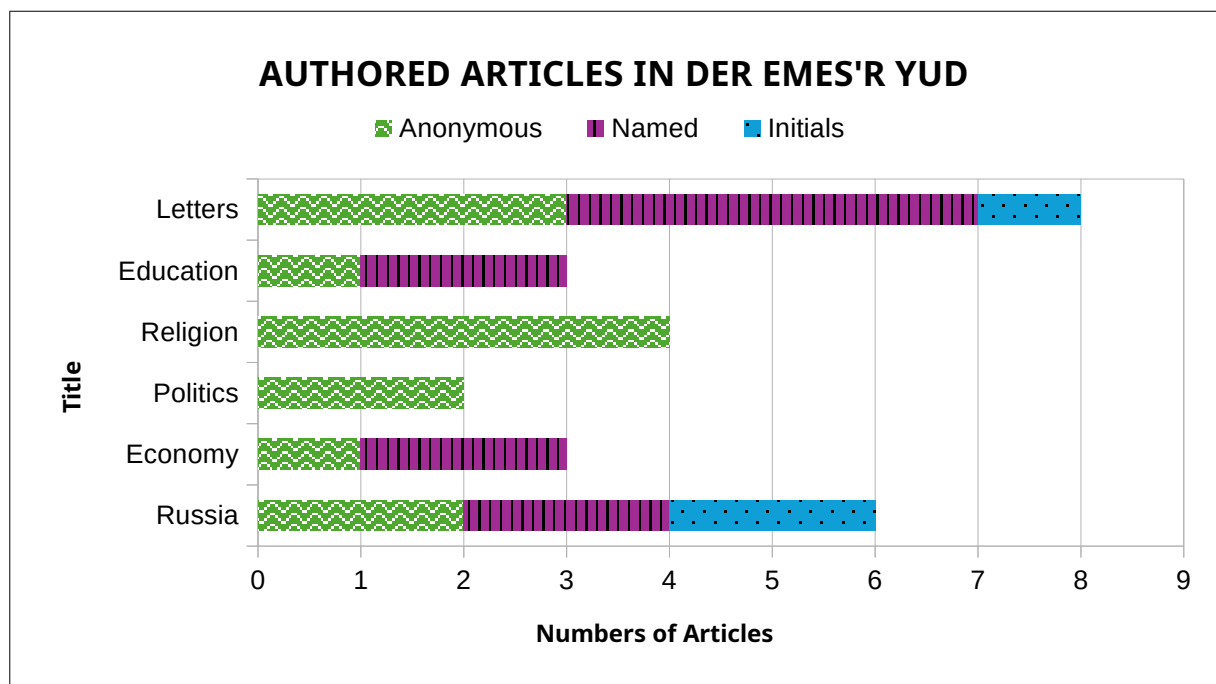


Figure 2: Articles with authors in "Der Emes'r Yud" sorted by topic

Apart from these instances of supporting authorship, the making of the paper was collaborative as the number of letters and telegrams sent from abroad shows Wilf could have had stable relationships and numerous connections with people around the world who would write articles and send telegrams as well as letters for the paper – presumably for free. It is also likely that he translated news and articles,

telegrams, and foreign writing similar to the editor of the first Yiddish newspaper in Galicia. He read 14 newspapers each day to translate and print them for a local readership³¹. Considering the price of the paper, the revenue would probably not have sufficed to pay royalties to a roster of authors. Receiving money for journalistic writing was after all uncommon at the end of the 19th century.³² The cost of subscribing to *Der Emes'r Yud* ranged from 3.50 crowns for a quarter year in the Habsburg monarchy to 20 (mostly likely French) francs for a full year for subscribers abroad. The prices increased slightly with the fifth publication, coinciding with the announcement of publishing a second weekly paper and Wilf's loss of additional income due to his parting with *Machsike Hadas*. The prices are visualised in Table 1, the lower starting price in brackets. Presumably most of the money went towards printing, the editor, and distribution, whereas the authors wrote their articles and letters for the sake of the education of their fellow Jews instead of monetary compensation.

TIME FRAME	AUSTRIA-HUNGARY	ABROAD
QUARTERLY	3.50 (2.50) Crowns	5.50 (3.50) Francs
BI-ANNUALLY	6 (4) Crowns	10 (6) Francs
ANNUALLY	12 (8) Crowns	20 (12) Francs

Table 1: Subscription prices of *Der Emes'r Yud*

However, the newspaper also printed advertisements on the last page of each publication longer than four pages. This must have additionally contributed to the newspaper's income. The advertisements are also a testament of the paper's international reach and faith in the possibility of widespread distribution thanks to a readership with disposable income. In total, there are 81 ads from 17 different advertisers in the 31 available papers. Figure 3 illustrates where the advertisers were

31 Manekin, 'Die hebräische und jiddische Presse', 2361.

32 Cohen, 'Yiddish Publishing Strategies', 3.

located. Six of those advertisers were based in L'viv and except for one from New York and one from Hamburg, the remaining ones were located in the Habsburg Monarchy, ranging from Kraków to Trieste. Four different coffee and tea traders advertised shipping their goods to L'viv, two from Galicia and two from the Adriatic coast as well as one of those explicitly offering kosher coffee. Two further newspapers were advertised. One explicitly Jewish newspaper from New York – *New York Jewish Newspaper* – and one non-Jewish one from Vienna – *Österreichische Volkszeitung* put advertisements in *Der Emes'r Yud*. Other trades advertised foods, farming equipment, books, fabrics, banks, or offered employment. The local L'viv businesses included a book shop, a bank, an editor selling reading material, a printing agency, and two fabric stores. The advertisements demonstrate the geographical reach *Der Emes'r Yud* supposedly had and the daily lives of the presumed readership. The frequency in how often the farming equipment and fabric stores were featured in the advertisement section confirms the common folk, farmers, and families as the

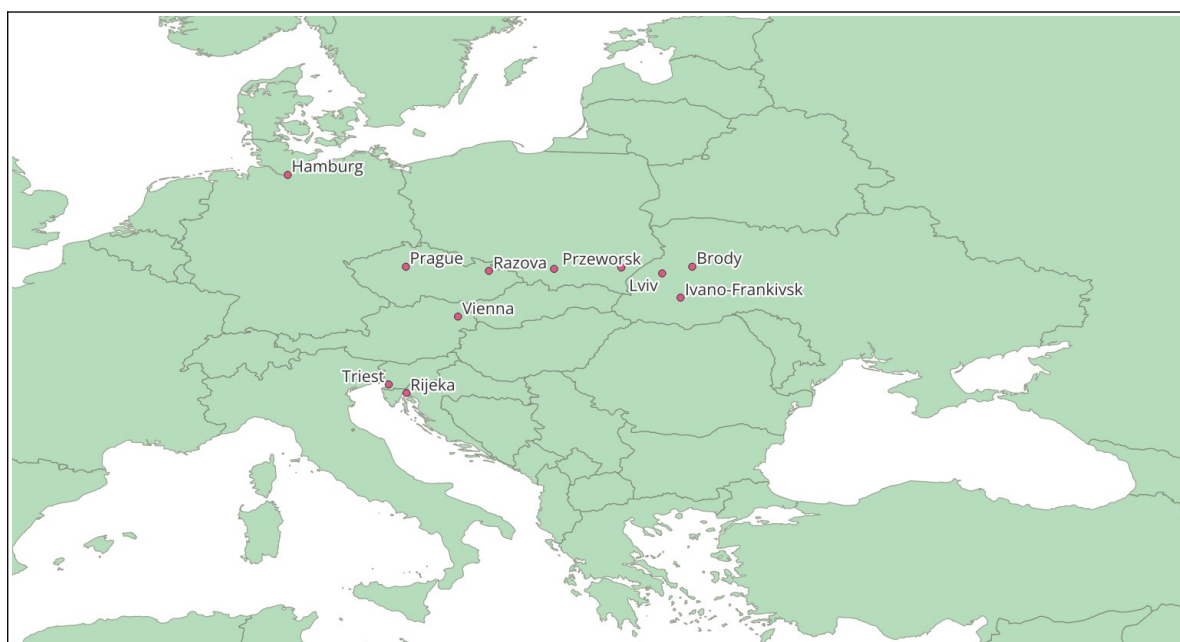


Figure 3: Cities in Eastern and Central Europe from where advertisements were sent to "Der Emes'r Yud"

targeted audience which was in line with the targeted audience of Yiddish periodicals around the turn of the century.

3 Yiddish Press

Yiddish press is a valuable historical source and one of the largest chronicles of the daily Jewish life³³ and is part of the over 10,000 Jewish newspapers and periodicals³⁴ that have been published thus far. Other than Yiddish, these publications were also written in other Jewish languages such as Hebrew and Ladino, as well as in local languages like English, German, and Polish. The oldest known Jewish journal was produced and printed in Amsterdam – the centre of Yiddish press before the 19th century³⁵: the first paper was written in Judeo-Spanish and Latin script from 1672 to 1699. The first Yiddish newspaper was published from 1686 to 1687³⁶. Two centuries later, in 1860, the first Hebrew newspaper *HaMelitz* (המליץ, *The Advocate*) was tried in the Russian Empire to meagre success since the majority of Jewish people did not master the liturgical language to the extent needed. Only in the 1880s did the modern Jewish press hold its footing in Eastern Europe³⁷. Two years after *HaMelitz* was launched, a Yiddish supplement was added to the newspaper called *Kol Mewaser* (קול מבשר – *The Herald*) which was later published as an independent print from 1869 to 1872. This first newspaper written in modern Eastern Yiddish was met with widespread enthusiasm and ran for several years with over 250 subscribers³⁸.

33 Edward Portnoy, *Bad Rabbi: And Other Strange but True Stories from the Yiddish Press*, Stanford Studies in Jewish History and Culture (Stanford, California: Stanford university press, 2018), 2.

34 The National Library of Israel has archived over 2,000 Jewish periodicals half of which can be freely accessed. The first entry was published from 1783 in Poland, see <https://www.nli.org.il/en/newspapers/titles>.

35 Marion Aptroot and Roland Gruschka, *Jiddisch: Geschichte und Kultur einer Weltsprache*, Originalausgabe, Beck'sche Reihe 1621 (München: Beck, 2010), 86.

36 Joshua Shanes, 'Jewish Press and Periodicals', in *The Routledge Handbook of Judaism in the 21st Century*, Routledge Handbooks in Religion, Book 18 (London; New York: Routledge, Taylor and Francis Group, 2023), 451.

37 Simon Perego, *Yiddish: Language, Culture and Memory from the Late 19th Century to the Present*, 15 April 2022, <https://doi.org/10.35008/TRACS-0123>.

38 Aptroot and Gruschka, *Jiddisch*, 108–9; Boris Kotlerman, 'Kol Mevasser', in *The YIVO Encyclopedia of Jews in Eastern Europe* (New York, NY: YIVO Institute for Jewish Research, 12 February 2010), <https://encyclopedia.yivo.org/article/1761>; Portnoy, *Bad Rabbi*, 6–7.

What started as a Yiddish supplement in a Hebrew newspaper was then widely accepted. It took several more decades for the Yiddish newspaper to be established as the new mass media in Eastern Europe.

Still, Hebrew was the preferred Jewish language of the educated Jews who saw Yiddish as a German dialect and a shameful jargon. But Hebrew failed at uniting Jews as it is a liturgical and theological language used for religious practices and law. It could not provide a nationalised concept of ethnicity, which Yiddish as a vernacular and therefore shared culture could achieve³⁹. On the other hand, Yiddish prevented its speakers to become proper citizens of the country they resided in, hence why the *Maskilim* – intellectuals and activists emerging from the *Haskalah* (Jewish Enlightenment) – pushed the usage of Hebrew or the local language⁴⁰. In the Austrian Empire, language was a marker and criterium for defining a nationality, but Yiddish was not granted the status of being an official language. The Galician Jews could therefore choose between being registered as Polish or German and were denied national rights⁴¹ hence why establishing Yiddish as an official language was more than aesthetic concerns.

In the last two decades of the 19th century, Yiddish media became more accepted and widely used by publishers and writers. Besides establishing Yiddish as a legitimate cultural language⁴², the choice to use this language instead of Hebrew or the locally prevalent languages was also politically and economically motivated. Editors and publishers, driven by economic gain, sought to distribute their prints farther by avoiding regional terms and phrases. Writers also aimed to establish Yiddish as a modern literary language to be accessible to as many people as

39 Masel, 'Who Is a Yid?', 376.

40 Portnoy, *Bad Rabbi*, 6–7.

41 Shanes, *Diaspora Nationalism*, 114–15.

42 Cohen, 'Yiddish Publishing Strategies', 8.

possible, which in turn also further standardised written Yiddish. By choosing the people's vernacular language, publishers and writers offered media with a Jewish perspective that could be understood by a majority of Jewish people living in Eastern Europe and Galicia as almost 5 million Ashkenazim Jews were living in Eastern Europe by 1880⁴³. This counteracted assimilation tendencies and in turn fostered a Jewish national(ist) sentiment⁴⁴ which were not necessarily Zionist in nature. Similarly, the editor of *Der Emes'r Yud* was fostering a Jewish identity, but not a Zionist one. Educating the reading masses in a language they could understand challenged the political, social, and cultural norms of the time⁴⁵. This should in turn lead to social improvement to contribute to the establishment of a Jewish national consciousness⁴⁶. Combining the educational efforts of the intellectual elite and the traditional, religious values of the Orthodox emerged the nationalist Yiddish press. Although the majority of Jewish periodicals published in Eastern Galicia had a Jewish-nationalist motivation, several other, especially newspapers written in Yiddish were socialist in nature and targeted the worker's class as their preferred readership instead of perusing national unification⁴⁷. Regardless of the political direction of a newspaper, writers saw that Yiddish was necessary if they wanted to reach the masses. Publishers further supported this development to sell higher volumes of their texts. At the same time, the nationalist newspapers constantly advocated for a Jewish national body in the vernacular of the people they were hoping to reach. Together, these contributions to the Yiddish publishing market shifted the

43 Aptroot and Gruschka, *Jiddisch*, 105–15.

44 Augusta Costiuc Radosav, "'Tsi Vos Toygt a Yudish Blat?' Of What Use Is a Yiddish Newspaper? Yiddish as a Language of the Press in Nineteenth-Century Romania', *Studia Judaica* 22 (2017): 49, <https://doi.org/10.24193/sj.2017.v22.2>.

45 Cohen, 'Yiddish Publishing Strategies', 6–8; Nathan Cohen, 'The Yiddish Press and Yiddish Literature: A Fertile but Complex Relationship', *Modern Judaism - A Journal of Jewish Ideas and Experience* 28, no. 2 (April 2008): 167, <https://doi.org/10.1093/mj/kjn004>.

46 Costiuc Radosav, "'Tsi Vos Toygt a Yudish Blat?', 44.

47 Toury, *Die Jüdische Presse*, 124–31.

perspective on the use of Yiddish and managed to turn the language into more than a tool to meet their objectives.

In the last decades of the 19th century, Yiddish turned into a commonly used language instead of a shameful vernacular only the uneducated used. From the 1880s on, The societal focus shifted from motivating a transition to learn other languages, especially Hebrew, to finally use Yiddish for the value of the language itself.⁴⁸ It was used to express love towards the Jewish people as well as to fortify pride for their language⁴⁹. By the mid-1890s, publishing in Yiddish was not seen as shameful anymore and the educated advocated for its wider usage⁵⁰. Finally, *Der Yid*, a Polish-Yiddish newspaper launched in 1899, kicked off the mass production of the Yiddish press in Galicia and five years later, *Der Emes'r Yid* was published to the reading masses of L'viv, Galicia, and the Austro-Hungarian Empire. In the subsequent years and decades, the Yiddish publishing activity grew steadily, reaching extensive numbers in the interwar years⁵¹. This leaves numerous Yiddish sources to be discovered and researched.

48 Cohen, 'The Yiddish Press and Yiddish Literature', 151.

49 Masel, 'Who Is a Yid?', 362.

50 Bauer, 'Commemoration and Jewish National Identity', 12; Cohen, 'Yiddish Publishing Strategies', 4; Masel, 'Who Is a Yid?', 362–63; Shanes, *Diaspora Nationalism*, 111.

51 Cohen, 'Yiddish Publishing Strategies', 8.

4 Yiddish

Yiddish as a Jewish language has existed for around a millennium and has since evolved through the influence of the surrounding languages spoken wherever Yiddish speakers had settled. The exact origins of the language are contested. The first written sources are dated back to the Rhineland in the 11th century where Yiddish was used for religious commentaries in supplemented glossaries to Hebrew writing. Researchers believe that Jewish migrants, before migrating to Germany before the 11th century, were already multilingual, speaking a Jewish variation of the dominant spoken language in their countries of origin. Germanic language aspects were slowly incorporated into the Yiddish vernacular and its development further fostered by manifold contacts between Jewish communities established in the different cities whereas the Middle High German dialects varied from city to city⁵². This variation of Yiddish known as Western Yiddish continued until the early 20th century, until it had come close to the brink of extinction. On the other hand, Eastern Yiddish – the language in which *Der Emes'r Yud* is written – contains a higher amount of Slavic components. While there are theories that Western Yiddish and Eastern Yiddish developed separately from Judeo-Romance in the West and Judeo-Slavic in the East respectively, it is more likely that the two linguistic entities evolved from the same root⁵³. Through emigrations from the Rhineland farer east in the 13th and 14th century, Early Yiddish came in contact with Slavic languages and subsequently adopted Slavic components. While Western Yiddish and its usage in Western Europe declined in the 19th century, Yiddish was reborn and transformed in Eastern Europe, introducing spoken dialect in its written form to be wider understood by the local Jews. Less based on its Germanic roots, Slavic characteristics

52 Aptroot and Gruschka, *Jiddisch*, 31–39.

53 Jürg Fleischer, 'The (Original) Unity of Western and Eastern Yiddish: An Assessment Based on Morphosyntactic Phenomena', in *Yiddish Language Structures*, Empirical Approaches to Language Typology 52 (Berlin Boston, Mass: De Gruyter Mouton, 2014), 107–10.

were woven into the language, for example the double negative and using one past tense instead of two, now focusing on the perfect tense only⁵⁴. Since its first proven instances, Yiddish and its speakers have evolved and adapted to the established languages of the region where they resided.

From the 19th century onwards, Yiddish speakers migrated to West Asia and North America and Yiddish evolved further incorporating English, Hebrew, and Dutch language patterns⁵⁵. The migrating Jews acquired the language of their new home and did not consistently teach their descendants the language of their ancestors, in turn decreasing the usage of Yiddish in their community. Although this had led to a discourse of Yiddish becoming an endangered language, it remained a stable Jewish language through all its changes⁵⁶. The language has been a staple for the Jewish diaspora as millions of mainly Yiddish-speaking Jews disseminated across Europe and the Americas in the nineteenth and twentieth century.

The number of Yiddish-speakers until the 20th century stayed consistent in relation to the number of Jews. Different numbers are provided by various researchers: At the beginning of the 20th century, the number of active speakers is estimated between 8 million Eastern Yiddish and 25,000 Western Yiddish speakers⁵⁷, 11 million estimated by the YIVO Institute for Jewish research⁵⁸, and 12 million as calculated in

54 Shlomo Berger, 'Translation, Interpretation and Transformation: Yiddish between Languages and Cultures', in *Trans-lation, Trans-nation, Trans-formation: Übersetzen und jüdische Kulturen*, Schriften des Centrums für Jüdische Studien, Band 21 (Innsbruck: StudienVerlag, 2012), 158–59.

55 Lea Schäfer, 'Yiddish', in *Oxford Research Encyclopedia of Linguistics*, by Lea Schäfer (Oxford University Press, 2023), 1–2, <https://doi.org/10.1093/acrefore/9780199384655.013.946>.

56 Sarah Bunin Benor, 'Hebrew and Jewish Diaspora Languages', in *The Routledge Handbook of Judaism in the 21st Century*, Routledge Handbooks in Religion, Book 18 (London, New York: Routledge, Taylor and Francis Group, 2023), 98.

57 Aptroot and Gruschka, *Jiddisch*, 125; Sarah Bunin Benor, 'How Many Speakers Are There? Statistics on Jewish Language Vitality.', *Jewish Language Website*, 2024, <https://www.jewishlanguages.org/stats>.

58 'Learning Yiddish', YIVO Institute for Jewish Research, accessed 2 September 2025, <https://yivo.org/Yiddish>.

1918⁵⁹. The systematic destruction of Jewish people and the Jewry during the Holocaust and the Stalinist efforts to liquidise Soviet Yiddish culture, which was followed by assimilation, acculturation, and repression in the second half of the twentieth century as well as the establishment of Hebrew as the official language of the State of Israel radically diminished the number of active Yiddish speakers⁶⁰. Of the 6 million Jews murdered, the majority were native Yiddish speakers, which left around 4 million speakers in the postwar period. Currently, estimates of active speakers range from half a million worldwide⁶¹ to around roughly 1.5 million speakers of which a third uses the language in their daily life⁶². Another study finds about 950,000 active speakers with great postvernacular activity⁶³. Postvernacular Yiddish is marked by a relationship with the language without fluency or proficiency in speaking or understanding, but an engagement with Yiddish culture⁶⁴. Learning the language and actively engaging with it steadily gains in popularity, rising the number of Yiddish users and speakers⁶⁵. For people who did not grow up learning and speaking the language, the postvernacular engagement is a means to connect with Jewish culture. In Ultra-Orthodox Haredi Jewish communities, Yiddish, besides Hebrew, is used to communicate with each other as vernacular.

59 Schäfer, 'Yiddish', 3.

60 Jeffrey Shandler, *Adventures in Yiddishland: Postvernacular Language & Culture* (Berkeley: University of California Press, 2006), 1.

61 Netta Abugov and Dorit Ravid, 'Noun Plurals in Israeli Hasidic Yiddish: A Psycholinguistic Perspective', in *Yiddish Language Structures*, Empirical Approaches to Language Typology 52 (Berlin Boston, Mass: De Gruyter Mouton, 2014), 11.

62 Schäfer, 'Yiddish', 3.

63 Benor, 'How Many Speakers Are There?'

64 Shandler, *Adventures in Yiddishland*, 4.

65 Benor, 'How Many Speakers Are There?', 98–99.

4.1 Yiddish in translation

The translation of Jewish texts has long been vital to Jewish studies, albeit translating Hebrew texts into vernacular languages has been focused on more than translating Yiddish⁶⁶. A large number of Yiddish literature has not been translated yet and the few that have been are so important for the Jewish literary canon to account for multiple translations⁶⁷. Literature on Yiddish translation focuses on poetry and highly regarded literature, as these are the cultural cornerstones to establish Yiddish as a language for the Jewish body⁶⁸. Media for the masses on the other hand, *schund*⁶⁹ or easy to understand periodicals, are often dismissed in Yiddish studies and translation, although they were central for the Jewish culture in Eastern Europe⁷⁰. Yiddish periodicals have been used as a vital source when researching local culture, socialist movements, or the rise of Jewish nationalism, but they remain only accessible to those who do understand the language. As a spoken language it is only natively used in Heradi or Hasidic Jewish communities. That makes translation an urgent task⁷¹ and one of the motivators for carrying out this project as part of my thesis.

66 Abigail Gillman, 'The Task of Jewish Translation Revisited', in *The Routledge Handbook of Translation History*, 1st ed., by Christopher Rundle (London: Routledge, 2021), 250–51, <https://doi.org/10.4324/9781315640129-19>.

67 Anita Norich, *Writing in Tongues: Translating Yiddish in the Twentieth Century*, 1st ed., Samuel and Althea Stroum Lectures in Jewish Studies (Seattle: University of Washington Press, 2013), 5.

68 For a 2023 bibliography of Yiddish linguistics, literature, and translation see Heidi Stern, 'Yiddish Studies: Language, Linguistics, and Literature', *Year's Work in Modern Language Studies* 83, no. 1 (2023): 655–67.

69 Schund was a category for badly-written, but popular literature in the 19th century made for the uneducated.

70 Saul Zaritt, 'A Taytsh Manifesto: Yiddish, Translation, and the Making of Modern Jewish Culture', *Jewish Social Studies* 26, no. 3 (2021): 206, <https://doi.org/10.2979/jewisocistud.26.3.07>.

71 Norich, *Writing in Tongues*, 8.

Besides being used as a tool to create accessible means for prayer to the lesser educated Jews who did not know Hebrew, by means of translating the Hebrew Torah to a vernacular language, Yiddish was also used as 'buffer language' between the surrounding cultures and their spoken language and the Jewish people living among them in the diaspora. As a successful buffer, it created a safer zone to distinguish Jews from non-Jews and form a group identity⁷². In German speaking regions Yiddish also provided another means of communication: a Yiddish speaking Jew was understood by a German speaker and could also understand the spoken German of the region to a certain degree. However, the Hebrew alphabet with which Yiddish was written still served as a functional barrier between these two languages⁷³. To translate Yiddish into another language, cultural and local knowledge is necessary to distinguish regional variations of Yiddish and understand the cultural characteristics.

72 Gillman, 'The Task of Jewish Translation', 254.

73 Berger, 'Trans-lation, Trans-nation, Trans-formation', 158.

5 Translation Process

The Digital Edition/Digital Translation of *Der Emes'r Yud* is predominantly based on manual work regarding the process of translation and annotation to engage with the language meaningfully and respectfully as opposed to seeing it as an obstacle to overcome via automation tools. While I explore the progress of Yiddish OCR and HTR in Chapter 6, not using the available automation tools was a deliberate choice to appreciate the language, preserving the cultural notions to my best abilities.

Critically engaging with my source is also supported by concerns about losing the *tam* (flavour, taste), the history and culture of the language as well as the difficulty of translating Jewish concepts and lore into a non-Jewish language⁷⁴ when translating manually⁷⁵. As a non-Jewish person, I am aware of potential shortcomings and lack of knowledge which, I tried to compensate by learning about the cultural setting of my source and engaging with Jewish and Yiddish media while and before working on this thesis.

5.1 Transliteration

The first step in the transformation process was not a direct translation from Yiddish to English, but an intermediate step of transliteration. Using the YIVO Transliteration Chart, I first transliterated the Yiddish sources into a Latin script in a separate document. The transliteration is of auxiliary nature and is therefore not available on the website or linked on GitHub. As a native German speaker and a native Latin script reader, the combination of Yiddish – a language with Germanic roots – in Latin writing accelerated the transformation process in the end despite leading to a near doubled workload. When transliterating, a Yiddish letter is replaced by the

74 One example for this is *Parshas Noyekh*, translating to "the week in which the Torah portion (beginning with Genesis 6:8) concerning Noah is read".

75 Norich, *Writing in Tongues*, 9.

approximate Latin counterpart⁷⁶. Still, it was necessary to consult the original source while translating, either because of a typo, unknown words and phrases, whole passages in Hebrew, or similar letters obfuscating the actual meaning of a word. While some Yiddish letters only seem familiar to the untrained eye – for example *beyz* (ב) and *khof* (כ) – others are only distinguishable by diacritics which are not always consistently used. The letter *alef* (א) is used to signify the Latin letter A in combination with a *pasekh* (ֿ) – as well as O when a *kometz* (ֿ) is added and is used at the beginning of a word to mark the next letter as a vowel instead of a consonant. *Alef* can then represent either A, O, or indicate a vowel dependent on the context. Similarly confusable are *fey* (פֿ) and *pey* (פּ) which are transliterated to F and P respectively. However, with time and repetition, the transliteration as well as the subsequent translation becomes easier enabling a higher accuracy rate once the words and writing style of a person becomes familiar.

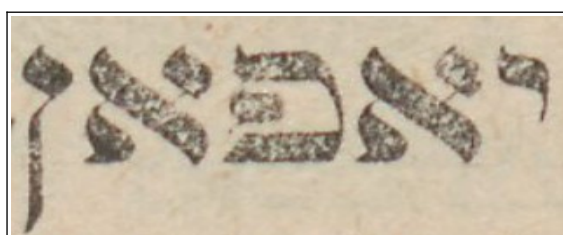


Figure 4: "Japan" written without diacritics;
Der Emes'r Yud (26.02.1904), 7.

5.2 Translation

In Yiddish, there are three words equivalent to the English word translation, all of which derive from German. There is the most commonly used *iberzetsn* – איבערזעצן – which comes closest to a literary translation which does not only convey the

76 Yiddish was at times also written with Latin letters, see Oren Cohen Roman, 'When Yiddish Was Written in Latin Letters', *Journal of Jewish Languages* 12, no. 1 (April 2024): 27–60, <https://doi.org/10.1163/22134638-bja10043>.

meaning of the text but consider the form of a text as well. Etymologically, it carries with it a notion of aggression and displacement. The second word *dolmetsh* – דאָלמעטשן – focuses more on simultaneous translation and is more bound to the literal – not literary – notions of the original to be translated. The third form *fartaytshn* – פֿאַרטײַטשן – has its etymological origins from German or *daytsh* (the name of the language itself), implying “meaning” and “interpretation” and is especially used to transfer Hebrew liturgical text into a Yiddish vernacular. Although the meaning and usage of these three words has become overlapping, notions of the original meanings are still relevant⁷⁷. The translation of *Der Emes’r Yud* for this project is then a mixture of *iberzetsn* and *dolmetsh*, since the goal was not to create an aesthetic counterpart to the original, but rather prioritise the texts meaning while following English grammar and syntax rules and transforming the Slavic characteristics of Yiddish, like the double negative and the usage of the perfect tense only when talking about the past⁷⁸, to orient the translation along the linguistic norms of English. By translating sentences and paragraphs, I could focus on the level of *parole* instead of the level of *langue* and convey what the text means, not what it says, and therefore find the closest natural equivalent of Yiddish in the English language⁷⁹. The translation is of functional nature and sensitive to possible mistranslation regarding meaning, since neither English nor Yiddish is my first language. However, for the intended purpose of this project to broaden the availability of the content to a wider audience, showcasing the significance of Yiddish periodicals, the translation is appropriate.

Besides Yiddish, Hebrew and German writing is interspersed in the text. The German writing in Latin script is constrained to footnotes as well as editorial information like

77 Norich, *Writing in Tongues*, 8–9.

78 Berger, ‘Trans-lation, Trans-nation, Trans-formation’, 158–59.

79 Werner Koller and Kjetil Berg Henjum, *Einführung in die Übersetzungswissenschaft*, 9. überarb. u. erw. Aufl (Stuttgart: utb GmbH, 2020), 105–6.

date and place of printing – all of which is also provided in Yiddish – and individual advertisements. Hebrew is regularly used in articles for liturgical quotes and Torah verses. I have translated both languages into English except for the editorial information on the first page where the Yiddish equivalent is also given. In cases in which the Hebrew phrases are not liturgical quotes, I used different dictionaries to translate and cross-check the translation since my knowledge in this language is limited. Some words, phrases, and concepts are also specifically Jewish and cannot be directly translated to English without explanation. For example, חדר is transliterated into English written *kheder* or can be translated with the word *school* or *room*, but neither of those terms possibility sufficiently conveys the meaning of the word: a religious primary school, traditionally for boys to learn Hebrew and study the Torah. Those elements are either explained in the translated text on the website and/or link to an external database for explanation.

5.3 Translating Named Entities

Additionally to the translation, the indices containing the people and places mentioned in the text was a vital part for the thesis to increase its usability. The register's functionality and technical components are further explained in Chapter 11 while this section deals with the methodology of named entity translation and allocation as well as the resources consulted in the process.

5.3.3 People

270 distinct people could be identified in the source material of which 166 could be categorically linked to their Wikidata entry. The remaining 104 people could either not be verified due to their common (e.g. H. Sholem Grosskopf) or incomplete name (e.g. Ivkov), or missing data entries (e.g. Rabbi Eliyahu Leib Shapira). Some people, especially rabbis or other important Jewish people in history do not have a Wikidata

entry but could be found in other encyclopaedias. In order to increase the chances of finding the correct person, I also consulted Jewish databases including *The YIVO Encyclopedia of Jews in Eastern Europe* (‘יוואָ-ענציקלאָפּעדיע פֿון די יידן אין מיזרח-⁸⁰ *אײראָפּע*⁸⁰ and the *Jewish Encyclopedia*⁸¹ by the Kopelman Foundation as well as tools like *Hebrew Namer*⁸² to cross-check the spelling and correct transliteration / translation of Jewish names into Latin script and English spelling. Several articles about the Austro-Hungarian parliamentary debates and its conflicts also amounted to a significant number of parliament members being mentioned by name in the paper. For this the database of parliament members *Parlamentarierinnen und Parlamentarier ab 1848*⁸³ was of immense help. Thanks to online tools and enriched datasets, 61 % of the people mentioned in the first two months of *Der Emes'r Yud* were traceable and the majority of these 61 % could be linked back to Wikidata.

5.3.4 Places

Throughout the translated text, 313 places could be indexed, including cities, countries, historical regions, oceans, continents, and street names. Out of these, 308 places are linked to Wikidata, and 302 places have an equivalent entry in GeoNames, a geographical database⁸⁴. Although every place mentioned could be deciphered, not all have a corresponding entry in an external database. The five places without a Wikidata entry are streets⁸⁵ in L'viv that are not deemed important enough to have a Wikidata entry. The additional six places not linked to GeoNames but with a

80 'The YIVO Encyclopedia of Jews in Eastern Europe', accessed 26 January 2026, <https://encyclopedia.yivo.org/>.

81 'JewishEncyclopedia.Com', Jewish Encyclopedia, accessed 2 December 2025, <https://jewishencyclopedia.com/>.

82 'HebrewNamer', HebrewNamer, accessed 26 January 2026, <https://hebrewnamer.com/>.

83 'Parlamentarierinnen und Parlamentarier ab 1848 | Parlament Österreich', accessed 26 January 2026, <https://www.parlament.gv.at/recherchieren/personen/parlamentarierinnen-ab-1848>.

84 'GeoNames', accessed 26 January 2026, <https://www.geonames.org/>.

85 Panteleimona Kulisha St, Hrodototska St, Petra Doroshenka St, Krakivska Vulytsia, Yana Zhyzhtky.

corresponding Wikidata article are historical regions (and one street in Vienna)⁸⁶ that are not represented in GeoNames. The coordinates have been acquired from GeoNames first, from Wikidata in case of a missing GeoNames entry, and from the "Lviv street" database⁸⁷ that provides the different street names throughout the 19th and 20th century and has been indispensable for matching the streets mentioned to a geographical counterpart on a contemporary map pictured in Figure 5 and Figure 6.

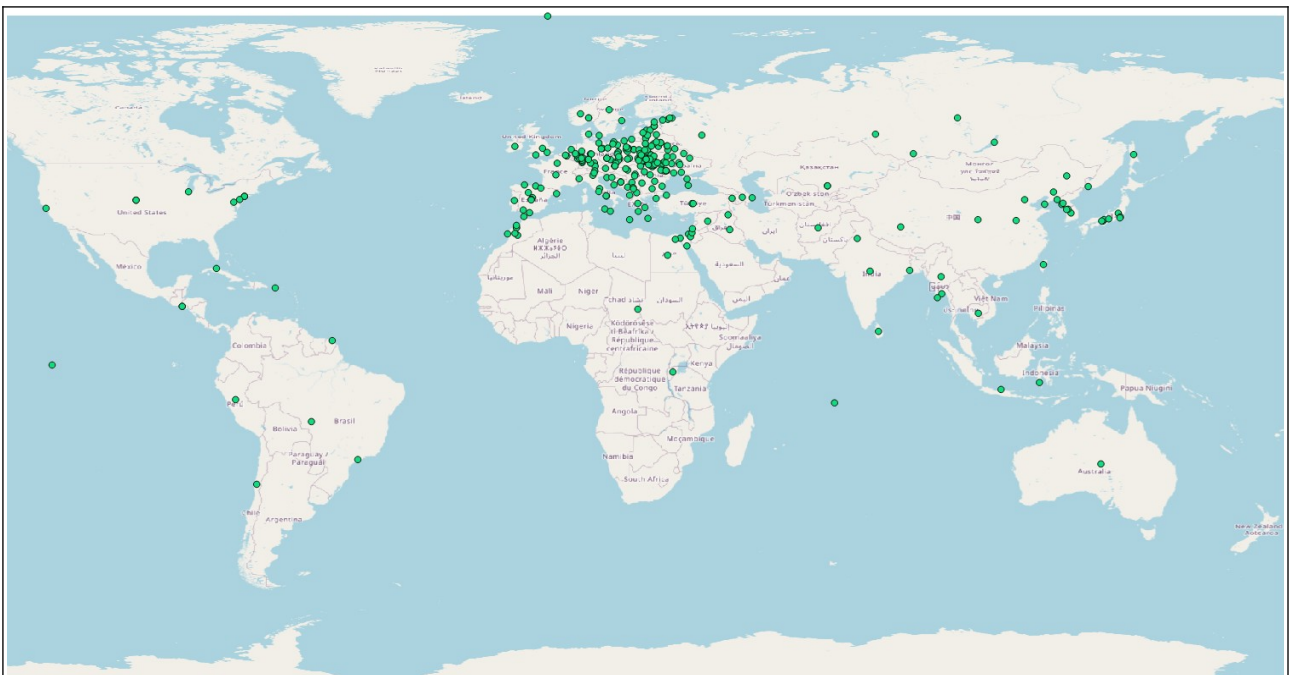


Figure 5: Places mentioned in "Der Emes'r Yid"

Since the source was written and published, the geographical and geopolitical circumstances and subsequently the names for different regions have changed manifold. In order to do justice to the historical circumstances, the places in the text are translated to their topical names in 1904. However, in the register, the current names in romanised spelling have also been provided with a note supplementing

⁸⁶ Austria-Hungary, Congress Poland, British Raj, Prussia Province, Schulerstraße, Canaan.

⁸⁷ 'The Streets of Lviv - List | Project "Lviv Streets"', accessed 26 January 2026, <https://streets.lvivcenter.org/en/streets/>.

further information about previous names and geographical affiliation (i.e. if a place

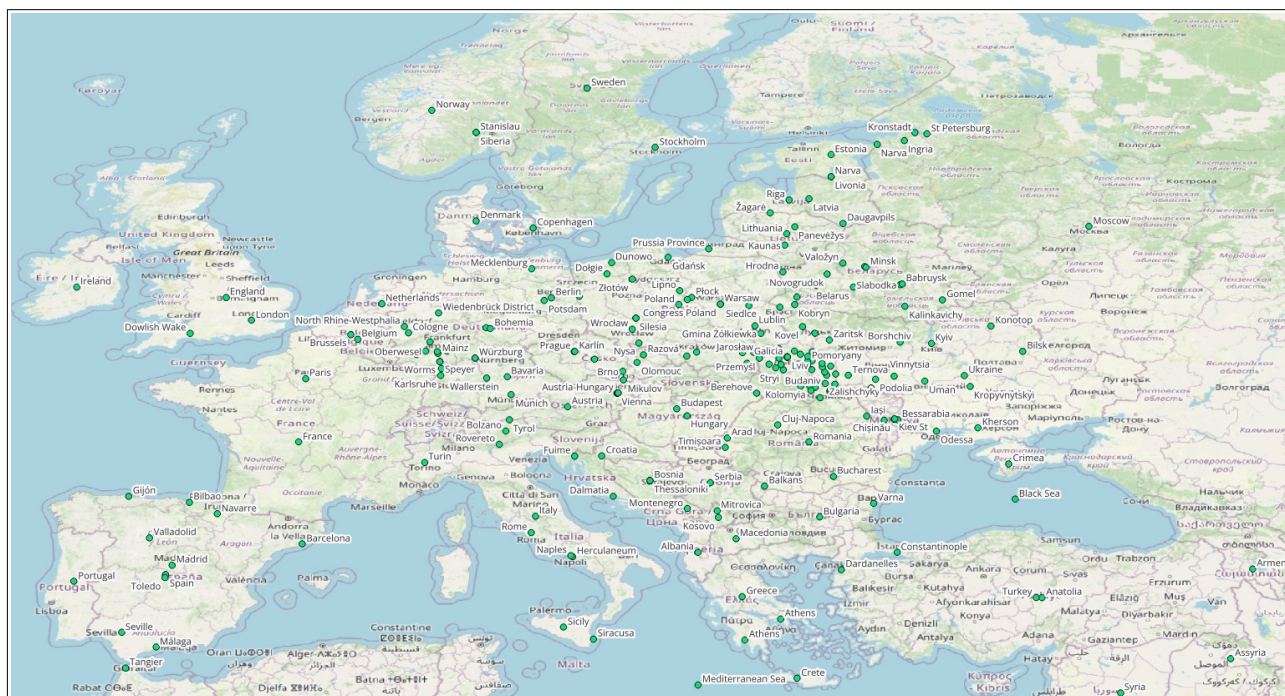


Figure 6: Map of Europe with the places mentioned in "Der Emes'r Yid"

belonged to the Russian Empire in 1904 but is now part of Moldova). Because of the non-standardised spelling, historical and geographical name changes as well as language barriers, some places were difficult to decipher. When translating from Yiddish, the names of towns and villages are often deformed, leaving the reader unable to figure out where a story took place⁸⁸. Solving this issue contributed to my efforts to work out the actual name and link to some other databases in case a town is not recognised by the name provided in the register. In the end, every place mentioned could be matched with an existing place and coordinates to display them on an embedded map as well as linking them to other regions.

88 Monika Adamczyk-Garbowska, 'A Historical Decontextualization of Eastern Europe', in *Yiddish and the Field of Translation: Agents, Strategies, Concepts and Discourses across Time and Space. In Cooperation with Marianne Windsperger*, 1. Auflage, Schriften Des Centrums Für Jüdische Studien, Band 033 (Göttingen: Böhlau Verlag Wien, 2020), 81–82, <https://doi.org/10.7767/9783205210306>.

6 Challenges of Automatisisation

The vast amount of manual labour invested into this project elicits questions about the possible advantages of streamlining automation procedures to facilitate the transformation process and possibly increase a future project scope. Aside from the ethical challenges of using AI and LLMs, Yiddish as a language and Hebrew as a script are not often used in Digital Humanities and Data Science projects. This could be one of the reasons as to why the tools necessary are not up to a usable and satisfactory standard. In this chapter I will discuss current OCR (Optical Character Recognition) and HTR (Handwritten Text Recognition) models for Yiddish, highlight the translation tools or rather dictionaries that were of help. I will also shed light on the cultural components of Jewish languages that makes it harder to translate it to a non-Jewish language.

6.1 Transcription

In a 2015 paper⁸⁹, datasets of the Europeana Newspaper project were tested on how well the OCR workflow performed on historical newspaper pages. While major scripts like Latin or Cyrillic scored satisfactory with around 75 % accuracy, Yiddish script emerged with a success rate of 36.5 %⁹⁰. Since OCR was developed to transcribe predictable printed single characters⁹¹, its usage is limited and barely customisable for more complex scripts. In this context, the low success rate is barely a surprise, even if a model is trained on Hebrew script. Since the Europeana survey,

89 Stefan Pletschacher, Christian Clausner, and Apostolos Antonacopoulos, 'Europeana Newspapers OCR Workflow Evaluation', *Proceedings of the 3rd International Workshop on Historical Document Imaging and Processing*, 22 August 2015, 39–46, <https://doi.org/10.1145/2809544.2809554>.

90 Ibid., 44.

91 Joe Nockels et al., 'Understanding the Application of Handwritten Text Recognition Technology in Heritage Contexts: A Systematic Review of Transkribus in Published Research', *Archival Science* 22, no. 3 (September 2022): 368, <https://doi.org/10.1007/s10502-022-09397-0>.

during the last decade, models have been trained and improved to better recognise Yiddish writing in Hebrew script. Notable projects are the text recognition model for the *Vaybertaytsch* typeface, a semi-cursive Ashkenazi typeface, developed by Ronny Reshef and Mirjam Gutschow⁹² as well as the Jochre3 model by Assaf Uriel, Amber Clooney, Michelle Sigiel, and Grisha Leyferi⁹³. The Vaybertaytsch corpus was trained on a set of 60,063 words from a single source and Jochre3 was trained on a corpus of 157,000 words spanning over 554 pages from the 1890s to the 1980s, eight cities, and seven different genres. The translated and annotated corpus of *Der Emes'r Yud* contains roughly 80,000 words and would be an ideal training and testing set for future research. The provided OCR by the National Library of Austria is unfortunately not of sufficient quality and would therefore benefit from such an endeavour.

In contrast to OCR, HTR can recognise an infinite number of glyphs for the same character at the same time, implementing algorithms to predict patterns instead of recognising each character individually⁹⁴. This development is essential for Yiddish writing. The Yiddish alphabet derives from the Hebrew script which is an abjad⁹⁵ as opposed to an alphabet and does not include vowels in its writing but rather indicates vowels by adding diacritics or *niqqud* to the letters. Hebrew characters were adapted for the Yiddish alphabet to create vowels: *ayen* (װ) turned into E, *alef* (א) for A and O, which can be distinguished by adding diacritical marks, *yud* (י) for either I or Y, and combinations *vov-yud* (װ) and *tsvey yudn* (ײ) for OY and AY

92 Ronny Reshef and Mirjam Gutschow, 'Text Recognition Model for Yiddish in Vaybertaytsh Typeface, Based on Community Regulations', *Journal of Open Humanities Data* 10 (May 2024): 35, <https://doi.org/10.5334/johd.194>.

93 Assaf Urieli et al., 'Jochre 3 and the Yiddish OCR Corpus', version 1, preprint, arXiv, 14 January 2025, <https://doi.org/10.48550/ARXIV.2501.08442>.

94 Beatrice Couture et al., 'The Challenges of HTR Model Training: Feedback from the Project Donner Le Gout de l'archive a l'ere Numerique', *Journal of Data Mining & Digital Humanities Historical Documents and...* (December 2023): 10542, <https://doi.org/10.46298/jdmdh.10542>.

95 Peter T. Daniels, 'Fundamentals of Grammatology', *Journal of the American Oriental Society* 110, no. 4 (December 1990): 730, <https://doi-org/10.2307/602899>.

respectively, all in accordance to the YIVO Transliteration chart⁹⁶. However, the usage of these letters has changed throughout history and the same characters were used to symbolise different sounds depending on time and place⁹⁷. Furthermore, Yiddish was never the official language of a sovereign state and does thus not have a unified orthography nor was the language standardised for the majority of its existence. The first international conference on Yiddish language, after which standardisation of Yiddish began to develop, was in 1908⁹⁸. The use of Yiddish in different regions brought forth a variety of dialects and influences of other languages in turn result in a large number of potential glyphs⁹⁹. Developing a potential HTR model for Yiddish is therefore a harder task and requires extensive training on individual sources to be successfully utilised.

The Jochre3 team point out several complications they had to overcome to train their models accordingly which are mirrored in this project as well. As is the case with *Der Emes'r Yud*, non-fiction Yiddish texts from the 19th and 20th centuries supplement text with other alphabets – mainly Latin and Cyrillic – depending on the publishing origin. This text as well as numbers are then written from left to right contradicting the right to left text flow of Hebrew text, seen in Figure 7.

96 'Yiddish Transliteration and Spelling | YIVO Transliteration Chart for Yiddish Alphabet | Yiddish Spelling Guidelines', accessed 26 January 2026, <https://www.yiddishwit.com/transliteration.html>.

97 Alexander Beider, 'Yiddish in Eastern Europe', in *Languages in Jewish Communities, Past and Present*, First edition, Contributions to the Sociology of Language (CSL), volume 112 (De Gruyter Mouton, 2018), 298–99.

98 Joshua A. Fishman, 'Czernowitz Conference', in *The YIVO Encyclopedia of Jews in Eastern Europe* (New York, NY: YIVO Institute for Jewish Research, 12 February 2010), <https://encyclopedia.yivo.org/article/2221>.

99 Urieli et al., 'Jochre 3 and the Yiddish OCR Corpus'.

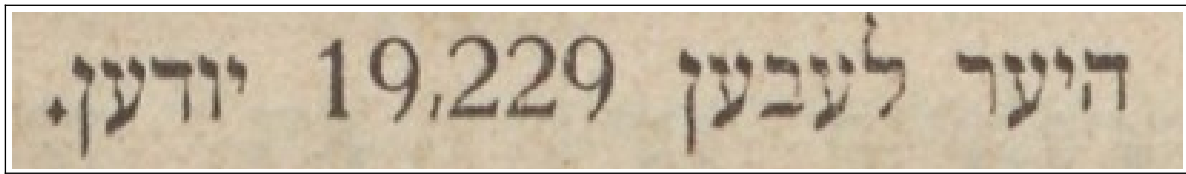


Figure 7: Example of right-to-left text and left-to-right numbers combined: "19,229 Jews live here.", *Der Emes'r Yud* (12.02.1904), 6.

Furthermore, spacing between letters of a word was used to either emphasise a word (Figure 8) or highlight a personal name, i.e. a city or person (Figure 9), equivalent to how italics are used nowadays.

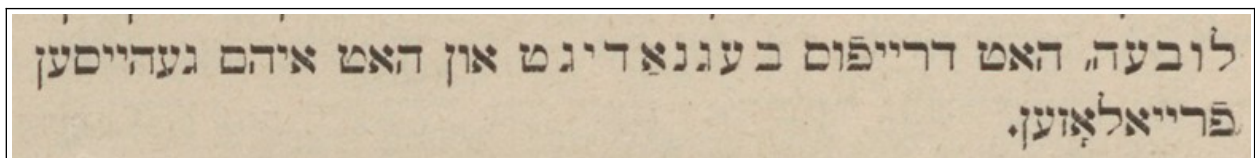


Figure 8: Spacing for emphasis: "Loubet pardoned Dreyfus and acquitted him.": Pardoned is emphasised, *Der Emes'r Yud* (08.03.1904,) 2.

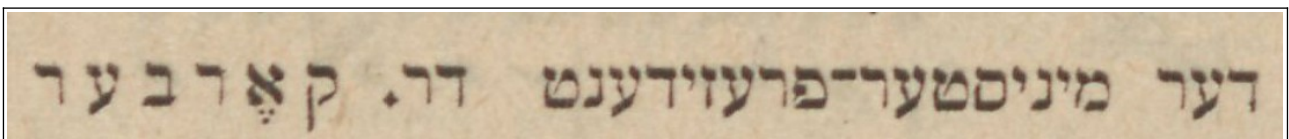


Figure 9: Spacing to indicate a name: "The prime minister Dr Körber", *Der Emes'r Yud* (1.03.1904,) 3.

A challenge for Yiddish script are also the ligatures *vov-yud* (וי), *tsvey yudn* (יי), and *tsvey vovn* (ייו) which compounded symbolise OY, EY, and V respectively, but do also exist on their own as *yud* (י) – I/Y – and *vov* (ו) – U. For Jochre3 the best result was obtained by considering all ligatures as separate glyphs except for *pasekh tsvey yudn* (יִי) – AY – the diacritic *pasekh* signifying the ligature. These peculiarities need to be corrected post-OCR for the source-text to have the right numbers, word-order, and spacing¹⁰⁰, the latter could also cause words to not be recognised as one compound. Irrespective of printing quality, these challenges can exacerbate the training of a Yiddish print HTR model.

100 Urieli et al., 'Jochre 3 and the Yiddish OCR Corpus'.

7 Scholarly Digital Edition

This project is situated in the methodology and theoretical background of the scholarly digital edition (SDE) which combines the requirements of scholarly editions with a digital paradigm. A digital edition describes projects that use digital technologies to (re)produce editions that can include transcriptions as well as markup of historical or literary documents. In these projects, digital technologies are used to grant and facilitate access to source material otherwise restricted, by providing access to an open online database and are therefore themselves part of cultural heritage¹⁰¹. Scholarly editions can serve multiple functions depending on the research area – history, linguistics, musicology, art history, etc. – and intended goals – reconstruction, stylistic analysis, accessibility, textual criticism, etc. As per Patrick Sahle's definition: "A scholarly edition is the critical representation of historic documents"¹⁰², a scholarly edition engages critically with historic documents (which do not have to be textual in nature) which need to be edited to be fully understood by today's audience and thereby recodes and transforms media. Text criticism can include phonological examination as well as identifying structures and naming entities of interest. The digital paradigm all digital scholarly editions should follow make it impossible to print a SDE without significant loss of function. Those aspects include multimodality (i.e. text, image, sound) or modularised structure to connect the material with editorial knowledge¹⁰³. Additional material can also contextualise and enrich the SDE for users. Relating parts within the edition and referring to external data, the born online medium increases transparency of the scientific

101 Greta Franzini, Melissa Terras, and Simon Mahony, 'Digital Editions of Text: Surveying User Requirements in the Digital Humanities', *Journal on Computing and Cultural Heritage* 12, no. 1 (February 2019): 2–8, <https://doi.org/10.1145/3230671>.

102 Patrick Sahle, 'What Is a Scholarly Digital Edition?', in *Digital Scholarly Editing: Theories and Practices*, Digital Humanities Series (Open Book Publishers, 2016), 23.

103 Ibid., 23–26.

process by connecting different parts with hyperlinks instead of footnotes for context and lead to external databases as well as represent the source with images and the facsimile¹⁰⁴. Following these digital paradigms is vital as to not reproduce and legitimise pre-digital practices for the digital edition¹⁰⁵. Therefore, a SDE adheres to the same guidelines as a printed scholarly edition but has also to abide to additional rules specifically for the digital realm. Apart from the digital paradigm, no SDE is like any other as the actual design and functionality is flexible in the interpretation according to the project's objectives. Still, guidelines are in place to support interoperability across disciplines, including the FAIR data guidelines¹⁰⁶, the Linked Open Data principles¹⁰⁷, or the CARE principles for Indigenous Data Governance¹⁰⁸. In the next chapter, I will explain the first two set of guidelines and describe how they apply to my research project.

104 Patrick Sahle, 'Digitale Edition', in *Digital Humanities*, ed. Fotis Jannidis, Hubertus Kohle, and Malte Rehbein (Stuttgart: J.B. Metzler, 2017), 239–41, https://doi.org/10.1007/978-3-476-05446-3_17.

105 Michelle Doran, 'Reflections on Digital Scholarly Editions: From Humanities Computing to Digital Humanities, the Influence of Web 2.0, and the Impact of the Editorial Process', *Variants* 15–16 (2021): 218, <https://doi.org/10.4000/variants.1414>.

106 'FAIR Principles', *GO FAIR*, n.d., accessed 22 January 2026, <https://www.go-fair.org/fair-principles/>.

107 Tim Berners-Lee, 'Linked Data', 18 June 2009, <https://www.w3.org/DesignIssues/LinkedData.html>.

108 'CARE Principles', Global Indigenous Data Alliance, accessed 22 January 2026, <https://www.gida-global.org/care>.

8 Digital Curation

Different guidelines and principles for digital projects can be implemented to standardise and facilitate access and accessibility. Although this project is not an archival practice per se, the guidelines for digital curation can be applied to the process of creating this digital edition. Hence, I will look at archival preserving practices used for born-digital and digitised collections in addition to following common editing and open web principles. The combination of these guidelines and principles aids in providing accessible online content which can be reused and accessed by different means of operation.

In comparison to analogue archival data, it is not sufficient for digital archival data to be available online. To be efficient, digital data needs to be integrated, interoperable linked data in contrast to having unstructured data which cannot be mined to its full potential¹⁰⁹. This digital curation practice preserves, promotes, and provides long term access to digital data of heritage data and supporting materials with current and future technologies¹¹⁰. Hence, it is important to use standardised formats, such as HTML or CSV. Further important aspects to regard when creating digital projects will be explored in the following chapters.

8.1 Linked Data

Linked Data facilitates the usage of research data by providing interoperable, reusable, and integrated data. Furthermore turning data into machine-readable and cross-disciplinary datasets, the benefits for digital humanities scholarship can be

109 Ashleigh Hawkins, 'Archives, Linked Data and the Digital Humanities: Increasing Access to Digitised and Born-Digital Archives via the Semantic Web', *Archival Science* 22, no. 3 (September 2022): 320, <https://doi.org/10.1007/s10502-021-09381-0>.

110 Arjun Sabharwal, *Digital Curation in the Digital Humanities: Preserving and Promoting Archival and Special Collections*, Chandos Information Professional Series (Waltham, MA: Chandos Publishing, 2015), 11.

extended even further¹¹¹. It functions as a means to meet the objective of a semantic web – a web of data – which can be processed by machines as well as humans¹¹². Tim Berners-Lee states four principles for creating linked data:

- (1) The usage of URIs (Uniform Resource Identifiers) – a unique sequence of characters to identify entities.
- (2) The URIs need to be resolvable with HTTP (Hypertext Transfer Protocol) identifiers to be searchable online.
- (3) Using these URIs, useful information should be returned using standards like RDF (Resource Description Framework)
- (4) The data should be linked to other things by reusing URIs¹¹³.

These three standardised technologies – URI, HTTP, and RDF – form the base for Linked Data¹¹⁴. Following these principles means that important entities in the data need to be distinguished by a unique identifier (like place001, W-5379, or 2017a) with which they can then be found on the web and can therefore be distinctly cited by external users.

One possibility to work towards Linked Data, which I have made use of, is to provide the Wikidata ID in registers, linking to this and other free databases like GeoNames and DBpedia. The use of Wikidata is very common for digital databases and editions to identify and enrich the data used, found, and annotated in researchers' projects. Although Wikidata misses historical, local, and specific entries, digital humanists using the database does have a positive impact on Wikidata's quality and in turn then improves the data other researchers use for their projects to situate the

111 Hawkins, 'Archives, Linked Data and the Digital Humanities', 328.

112 Christian Bizer, Tom Heath, and Tim Berners-Lee, 'Linked Data - The Story So Far', *International Journal on Semantic Web and Information Systems* 5, no. 3 (July 2009): 15, <https://doi.org/10.4018/jswis.2009081901>.

113 Berners-Lee, 'Linked Data'.

114 Bizer, Heath, and Berners-Lee, 'Linked Data - The Story So Far', 14.

generated data in the semantic web. Although at times generic, Wikidata is a solid starting point. When used, researchers should compare Wikidata's content with the data of other databases¹¹⁵. Further databases can be implemented and should be used to compare the data from Wikidata to other sources.

8.1.1 Metadata

When curating digital historical data, researchers must also plan to curate and archive relevant metadata, i.e. data of data, as it contains valuable and structured data, which is useful for further research too. Metadata is a source on its own, providing context for the creation of a curated digital product and highlighting the work that went into a product beyond what is visible on a website or project report and should hence be archived side by side with the curated data. Therefore, it is vital to not only archive the representation, i.e. the digital edition, but also the document containing the code and the document's metadata. Data bridges humanities and digital collection by providing metadata for missing contextual information to feed the search algorithm for prioritising the database in search results. For digital editions in particular, using the TEI (Text Encoding Initiative) standard, the metadata is included in the document already as to why not only the representation on a website, but also the document itself containing the code needs to be preserved. To support digital curation and collaboration, metadata standards are indispensable for the authenticity and longevity of the collection¹¹⁶. Examples for metadata standards are – besides TEI – RDF (Resource Description Framework) for web resources or Dublin Core and DOI (Digital Object Identifier) for networked resources.

115 Fudie Zhao, 'A Systematic Review of Wikidata in Digital Humanities Projects', *Digital Scholarship in the Humanities* 38, no. 2 (May 2023): 868–72, <https://doi.org/10.1093/llc/fqac083>.

116 Sabharwal, *Digital Curation in the Digital Humanities*, 55–60.

8.1.2 Linked Open Data

Creating a truly open access database requires the fulfilment of the FAIR principles explained in the next chapter as well as five stars in the Linked Open Data scheme by Berners-Lee¹¹⁷. The Five Stars are represent a ranking system for the openness of an online source, gaining a star each time another aspect of interoperability and universal aspects is met:

- ★ The Linked Data is available online, open and free to use
- ★★ The Data is structured to be machine-readable.
- ★★★ The structure is available in non-proprietary format
- ★★★★ The Data uses open standards.
- ★★★★★ The Data is linked to other data(bases) on the Web

For this project the first star had already been met by the scans of the source provided by the Austrian National Library. Theoretically, it was also already machine-readable by the supplied automated transcription, gaining another star, but in practice, the transcription does not match the source text. As the translated text on the project website is available in XML, HTTP, CSV, and JSON, it is universally accessible without being tied to any specific software, gaining the third star. The universal identifiers designed with open standards connect the data on the website and externally to other databases, thereby gaining the fourth and fifth star. According to this chart, the finished dataset has met the highest standard of open linked data.

8.2 FAIR Data Principles

Once a digital project has been finished and the outcomes have been shared, it stands to lose its importance and impacts if the researchers are not careful to

¹¹⁷ Berners-Lee, 'Linked Data'.

preserve and care for its longevity. The FAIR principles were developed to assist in supporting the infrastructure for reusable scientific data as well as provide foundational guidelines for high quality open (meta)data and its subsequent reuse. It builds on the idea that data should be both human- and machine-readable and was developed with different stakeholders in mind. The term *FAIR* is an acronym consisting of four different aspects of open data publishing: *Findable*, *Accessible*, *Interoperable*, and *Reusable*¹¹⁸. This subchapter will explain the four categories as defined by Wilkinson et al. which are explained in great detail on the Go Fair website¹¹⁹. Furthermore, the guidelines are supplemented with their role for digital scholarly editions in particular as developed by Gengnagel, Neuber, and Schulz¹²⁰. Finally, I will also describe how I have configured the data to meet each of the guideline's aspects.

8.2.1 Findable

The guideline *Findable* describes how openly published data can be found by means of ample and identifiable data and metadata. It requires a universally persistent identifier describing the available data. Additionally, data can be listed, registered, and indexed in a searchable resource catalogue¹²¹, i.e. Zenodo¹²² or Europeana¹²³. Data repositories usually assign a unique identifier after uploading the dataset.

118 Mark D. Wilkinson et al., 'The FAIR Guiding Principles for Scientific Data Management and Stewardship', *Scientific Data* 3, no. 1 (March 2016): 1–2, <https://doi.org/10.1038/sdata.2016.18>.

119 'Fair Principles'.

120 Tessa Gengnagel et al., *FAIR Principles in Digital Scholarly Editions*, 2022, <https://ride.i-d-e.de/fair-criteria-editions/>.

121 Wilkinson et al., 'The FAIR Guiding Principles for Scientific Data Management and Stewardship', 4.

122 <https://zenodo.org/>

123 <https://www.europeana.eu>

Additionally to the required metadata, it is also useful to provide context on the quality, condition, and characteristics of the data to allow for functional filtering¹²⁴.

For a digital edition this means that the edition can be found in a library or edition catalogues as well as in the general data repositories mentioned¹²⁵. For these repositories to work, researchers need to carefully select metadata describing their projects for the interested public to find exactly what they are looking for.

I will register my project on Zenodo and contact the Austrian National Library to provide the translation of one of their items either on their website or linking to a publicly accessible webpage.

8.2.2 Accessible

The *Accessible* aspect defines how data can be retrieved, accessed, and downloaded. The accessible guideline entails a different process than the accessibility of the data which is covered by the Web Content Accessibility Guidelines¹²⁶ which should be considered when planning and publishing a digital project. To be accessible according to the FAIR principles, data should use a standardised communications protocol and preserve the metadata, even if the data itself is no longer available. Such a standardised communications protocol needs to also be open, free, and universally implementable as well as allow for authentication or authorisation if necessary. The first is given by using HTTP which is free and open source, guaranteeing globally universal access to the data. As data storage can be expensive

124 <https://www.go-fair.org/fair-principles/f1-meta-data-assigned-globally-unique-persistent-identifiers/>

125 Gengnagel et al., *FAIR Principles*.

126 'Web Content Accessibility Guidelines (WCAG) 2.1', with Andrew Kirkpatrick et al., accessed 15 February 2026, <https://www.w3.org/TR/WCAG21/>.

in the long run, storing only the metadata can be a sustainable alternative by which the original project can be traced down¹²⁷ and the author contacted.

When creating an edition, it is important to consider if any user can access the edition or if usage is tied to a certain institution. Furthermore, copyright can be an issue if the material is recent and therefore the edition can only be made partially available¹²⁸.

My project has been publicly available since the beginning by deploying the edition via GitHub pages which uses Hypertext Transfer Protocol Secure (HTTPS) to access it. For long-term preservation, the metadata will be made available on Zenodo and the project will remain on a publicly available repository, presumably on GitHub or a similar public repository.

8.2.3 Interoperable

Interoperable data requires a broadly applicable language – or standard – to describe the data and metadata to allow for exchange and interpretation without having to use specialised or proprietary software. If special knowledge is required to interpret the data, these specification should be provided in the documentation as well. This in turn allows the data to be linked to other external datasets with universally unique and persistent identifiers, another goal of the interoperable standard¹²⁹.

For SDEs, the TEI standard is the most widely used standard to format textual digital editions. Although TEI is customisable, the basic framework to structure documents allows for integration and comparison of different editions. As specific software for

127 <https://www.go-fair.org/fair-principles/metadata-retrievable-identifier-standardised-communication-protocol/>

128 Gengnagel et al., *FAIR Principles*.

129 <https://www.go-fair.org/fair-principles/i1-metadata-use-formal-accessible-shared-broadly-applicable-language-knowledge-representation/>

creating and using the corpora of a digital edition is used, the compatibility for different software should also be regarded¹³⁰.

The documents are encoded according to the TEI and XML standard and are then converted in static HTML. The data can be either downloaded in CSV, JSON, and HTML format as well as in the original XML files. External data is linked using the unique identifiers provided by Wikidata and GeoNames. The facsimile can be retrieved via the Austrian National Library catalogue on ANNO in PDF and TXT format.

8.2.4 Reusable

The *Reusable* principle also concerns itself with interoperability. While the technical interoperability is covered by the previous guidelines, reusability is about the legal interoperability. Here, accurate and relevant attributes when describing the data using domain-relevant standards published with a clear and accessible data usage license are needed. Creative Commons¹³¹ aides in providing exact usage licenses readable by both humans and machines. If external users are allowed to reuse the produced data, they should be able to know how the data provided was generated which can be clearly stated in a documentation file. Depending on the area, discipline specific community standards help facilitate cooperation¹³².

For a digital edition, the creation, processing and modelling needs to be documented to allow third parties to work with the edition. The license should be stated in the mark-up as well as documentation files. The aggregation of data should be detailed in the documentation file as well¹³³.

130 Gengnagel et al., *FAIR Principles*.

131 https://wiki.creativecommons.org/wiki/License_RDF

132 <https://www.go-fair.org/fair-principles/r1-metadata-richly-described-plurality-accurate-relevant-attributes/>

133 Gengnagel et al., *FAIR Principles*.

The edition and complementing files are all open for reuse and interpretation. The licensing is stated as follows in each file:

```
<availability>
  <licence
    target="https://creativecommons.org/licenses/by/4.0/deed.en">
    <p>
      You are free to: Share – copy and redistribute the
      material in any medium or format; adapt – remix,
      transform, and build upon the material for any
      purpose, even commercially.
    </p>
    <p>
      The licensor cannot revoke these freedoms as long as
      you follow the license terms. Under the following
      terms:
    </p>
    <p>
      Attribution – You must give appropriate credit,
      provide a link to the license, and indicate if
      changes were made. You may do so in any reasonable
      manner, but not in any way that suggests the licensor
      endorses you or your use. No additional restrictions
      – You may not apply legal terms or technological
      measures that legally restrict others from doing
      anything the license permits.
    </p>
    <p>Notices:</p>
    <p>
      You do not have to comply with the license for
      elements of the material in the public domain or
      where your use is permitted by an applicable
      exception or limitation. No warranties are given. The
      license may not give you all of the permissions
      necessary for your intended use. For example, other
      rights such as publicity, privacy, or moral rights
      may limit how you use the material.
    </p>
  </licence>
```

<p>Copyright 2021 Österreichische Nationalbibliothek</p>
</availability>

The notice allows for reuse of the material as long as credit is given to the original work. The copyright remains with the Austrian National Library.

Following these standards, principles, and guidelines ensures a wide usage and accessibility of the created data to be a high-quality source for further research. Making your data open and fair supports an interlinked network of machine-readable resources, beneficial for common users and researchers alike.

9 Scholarly Annotation Methodology

The project is built on a three-level modular structure to visualise the XML-files on an easy to navigate static website. The website was created using the Digital Scholarly Edition Static Cookiecutter Template by the Austrian Centre for Digital Humanities (ACDH), an institution of the Austrian Academy of Sciences¹³⁴. The first layer are the XML-files which are linked to the corresponding second layer, XSLT (Extensible Stylesheet Language Transformation) and CSS (Cascading Stylesheets), to then display a HTML website which constitutes the third layer.

The first layer consists of the translated archival data in XML files which contains the semantic information and annotation to be transformed into visual information. On its own, the documents cannot be displayed on a website without a transformation script. This layer is thoroughly explained in Chapter 9. The second layer is the dynamic engine responsible for the visualisation transformation. It is the subject of Chapter 10. The first and second layer contain corresponding code to convert the semantic XML into presentational HTML. The third and final layer makes up the user interface constructed by HTML and minor CSS and JavaScript contributions. It is responsible for the styling, layout, and interactivity of the website. This will be further explored in Chapter 11.

9.1 Encoding Rationale

Digital editions are often created using markup language, which is a human- and machine-readable text-encoding system. XML (Extensible Markup Language) and more specific TEI (Text Encoding Initiative) have been the tools of choice for most modern textual scholarly digital editions¹³⁵. XML is a descriptive markup language in

134 *Acdh-Oeaw/Dse-Static-Cookiecutter*, XSLT, 11 October 2021; Austrian Centre for Digital Humanities, released 9 December 2025, <https://github.com/acdh-oeaw/dse-static-cookiecutter>.

135 Sahle, 'Digitale Edition', 245.

contrast to presentational or procedural markup used in word processing and programming constructions respectively. A descriptive language describes what a piece of text is, rather than stating how that piece of text should be processed¹³⁶. The TEI Guidelines are an XML based format and semantic markup language used to represent “the structural, renditional, and conceptual features of texts”¹³⁷. The structure is highly customisable and has therefore proved to be popular for different projects in the humanities and social sciences. Despite some shortcomings noted by its users, TEI/XML remains the standard for digital editing, albeit not for a lack of alternatives¹³⁸. For this project, no alternatives were sought out as TEI/XML proved to be sufficient to meet the objectives set out. However, if I had chosen not to translate the sources and instead create a Yiddish edition, I would have encountered an array of obstacles. Non-Latin scripts are not yet as integrated in the standards, already the direction in which the text is read can lead to difficulties¹³⁹. By translating and transliterating I was able to avoid any problems faced when working with a minority language in common digital knowledge infrastructures.

9.2 Encoding Scheme

The TEI scheme is a conceptual framework consisting of multiple modules which can be combined to meet the project’s needs and the editor’s objectives. According to the TEI best practices, this project follows the encoding level 5. This means that the text is generated by humans, it is transcribed completely and able to be used and

136 Tim Bray et al., ‘Extensible Markup Language (XML) 1.0 (Fifth Edition)’, W3C, 26 November 2008, <https://www.w3.org/TR/REC-xml/>.

137 ‘TEI Guidelines’, Text Encoding Initiative, accessed 21 November 2025, <https://www.tei-c.org/guidelines/>.

138 Thomas Stäcker, ›A Digital Edition Is Not Visible‹ - Some Thoughts on the Nature and Persistence of Digital Editions, version 1.0, 2020, Html,Xml,PDF, https://doi.org/10.17175/2020_005.

139 Alíz Horváth et al., ‘Towards Multilingually Enabled Digital Knowledge Infrastructures’, in *Multilingual Digital Humanities*, 1st ed., by Lorella Viola and Paul Spence (London: Routledge, 2023), 204, <https://doi.org/10.4324/9781003393696-17>.

displayed without the facsimile or the source material accompanying the transcribed text. Additionally, it requires human intervention including subject knowledge to be annotated accordingly¹⁴⁰. While the schema can be constructed flexible to the project's needs, some modules are fundamental and should rarely be excluded from the structure: the *header*, *core*, and *textstructure* models should be used in every TEI file with few exceptions. The *header* module describes the different aspects of the encoding process such as the text itself, the source material, or encoding rationale. The *core* module, subtitled "Elements Available in All TEI Documents" entails descriptions for elements and attributes which are likely to be used globally in any kind of texts. The *structure* module includes elements to annotate punctuation, quotations, simple editorial changes, numbers, references, and non-textual components. The *textstructure* module compliments the *core* module by defining the structure for TEI documents and describes how the tags should be used, how to structure the document in order, how to group texts, and what elements are common to all divisions¹⁴¹. Since the project's objectives do not branch out into text criticism, linguistic analysis, or textual interpretation, only few further modules were needed to annotate to the preferred extent. The other two modules used for the edition are *transc* and *namesdate*.

The module *transc* or "Representation of Primary Sources" is required to describe the digital facsimile. In this project, the facsimile is provided via a link to the Austrian National Library to avoid possible copyright issues and especially to keep the data storage size at a minimum. By linking to the original source location, the picture is not loaded on the website itself, but in an additional browser tab or window.

Supplementing a URL in the <facsimile> tag is not usual practice, but needed in this

140 Kevin Hawkins and Michelle Dalmau, 'Best Practices for TEI in Libraries', Text, November 2017, https://tei-c.org/extra/teiinlibraries/3.1.0a/main-driver.html#index.xml-body.1_div.3_div.5.

141 '1 The TEI Infrastructure - The TEI Guidelines', Text Encoding Initiative, accessed 28 November 2025, <https://tei-c.org/release/doc/tei-p5-doc/en/html/ST.html#STIN>.

case to automatically supply a connection to the source on the website. The `<surface>` element contains information about the size given in pixels by referencing the upper left corner with the x coordinate (@ulx) and the y coordinate (@uly) and the lower right corner with the x coordinate (@lrx) and y coordinate (@lry) respectively. While the @source attribute in the `<facsimile>` element provides the link for the whole source file, the @url attribute in the `<graphic>` element references the specific page¹⁴². For example, the facsimile of an issue released on March 15th 1904, including four pages is formatted like this:

```
<facsimile source="https://anno.onb.ac.at/cgi-content/anno?
aid=emj&datum=19040315">
  <surface ulx="0" uly="0" lrx="4257" lry="6222" n="1">
    <graphic url="https://anno.onb.ac.at/cgi-content/
annoshow?call=emj|19040315|1|100.0|0" />
  </surface>
  <surface ulx="0" uly="0" lrx="4305" lry="6222" n="2">
    <graphic url="https://anno.onb.ac.at/cgi-content/
annoshow?call=emj|19040315|2|100.0|0" />
  </surface>
  <surface ulx="0" uly="0" lrx="4257" lry="6222" n="3">
    <graphic url="https://anno.onb.ac.at/cgi-content/
annoshow?call=emj|19040315|3|100.0|0" />
  </surface>
  <surface ulx="0" uly="0" lrx="4305" lry="6222" n="4">
    <graphic url="https://anno.onb.ac.at/cgi-content/
annoshow?call=emj|19040315|4|100.0|0" />
  </surface>
</facsimile>
```

The other module I have used to format this project is the *namesdate* module, which describes the encoding of names, dates, people, and places. This module allows for distinct annotation of different entities. Entities can be described in great detail including honorifics, generational epitaphs, distinction between the type and size of

142 '12 Representation of Primary Sources - The TEI Guidelines', Text Encoding Initiative, accessed 28 November 2025, <https://tei-c.org/release/doc/tei-p5-doc/en/html/PH.html>.

places as well as prosopographical data and are therefore highly customisable depending on the context and objectives of a project¹⁴³. For the purpose of possible re-use I decided to keep a simple structure of first name(s) and surname(s) as well as a unified typology of marking places. The usage of this module will be further explored in the next subchapter.

9.3 Annotation Methodology

The project contains two different types of XML files: the edition file in which the translated source text is annotated, and the register files in which people (listperson.xml) and places (listplace.xml) are compiled. For each newspaper a separate XML file is created and named after the date the source was published (YYYY-MM-DD.xml, e.g. 1904-02-05.xml).

The <teiHeader> for each newspaper issue contains metadata of the created file and the source it is based on. Therein included is the title of the file, its authors as well as editors, the responsibility statement, the copyright statement, languages used, and information about the source regarding its origin and current repository. The header closes with the <facsimile> element as described above.

The textual annotation of the translated source text fulfils two purposes: The first is to structure the textual layout and stylise the text in combination with the stylesheets. This will be further contextualised in Chapter 10. The second purpose is to mark up named entities as well as link to the indices and therefore append further information to the text.

143 '14 Names, Dates, People, and Places - The TEI Guidelines', Text Encoding Initiative, accessed 28 November 2025, <https://tei-c.org/release/doc/tei-p5-doc/en/html/ND.html>.

9.3.1 Layout

The text is formatted in a tree hierarchy with nested XML elements which are transformed via the stylesheets to create a valid HTML syntax. The translation is contained within the `<text>` and `<body>` element wherein every text segment, e.g. a newspaper article, is enclosed by a text division element (`<div>`) and structured by paragraphs (`<p>`). The newspaper headlines (`<head>`) can be divided into up to three levels by supplementing the type attribute "h2" or "h3". The divisions are numbered with the attribute *n*.

```
<text>
  <body>
    <div n="1">
      <head/>
      <head type="h2" />
      <p/>
    </div>
  </body>
</text>
```

Header, footer, and page numbers are tagged with the *forme work* element (`<fw>`). This element can be placed within or outside a division element without invalidating the syntax. The `<fw>` element is further specified by the type attribute ("header", "footer", and "pageNum") depending on it's content. Breaks in the text can be indicated by line beginning (`<lb>`), column beginning (`<cb>`) and page beginning (`<pb>`). Line beginnings are visualised on the website, but column beginnings are not displayed. The original source text is structured in two columns which is not visualised on the website and is therefore rendered unnecessary. A new page is indicated by the page number, but I have decided to display the entire newspaper on one page for the sake of simplicity. However, since the XML files should also be able to stand alone without the transformation and website, it is vital for the layout to be representative of the source itself.

Some text is formatted to provide visual cues for readers. Titles, headers, footers, and page numbers are aligned in the centre and the titles are also displayed in bold lettering. If a word or phrase is emphasised in the source material by spaced out letters, the highlighting element (`<hi>`) is used to render the passage in italic letters:

the improvements are `<hi rend="italic">not enough</hi>`

As most of the elements are not personalised and are used as the TEI guidelines suggest, the layout annotation can be universally reused and reproduced.

9.3.2 Linguistic Characteristics

Although the majority of the newspaper is written in Yiddish, passages are also written in Hebrew and German. The element `<foreign>` is used in combination with the global, i.e. usable in every element, attribute `xml:lang` and the corresponding ISO language code, in this case “de” for German and “he” for Hebrew:

`<foreign xml:lang="de">1. Jahrgang - Nr. 8 -</foreign>`

If Hebrew is not transliterated and therefore written in another script, the attribute also needs to signal this:

`<foreign xml:lang="he-Hebr">הדת' קי' מחזיק</foreign>`

While the language usage and distribution is not of importance in this project, it may be used in the future and should therefore be compatible with existing structures.

9.3.3 References

References are used to provide further context by linking to additional information either within the edition project or to external sources. In both cases, the element `<ref>` is used along with the attribute `source` to include the corresponding URI or

URL. Referencing internally, a unique URI, *place###* or *person###*, create a direct link to the register entry:

```
<ref source="place078">Greece</ref>
```

```
<ref source="person101">John of Kronstadt</ref>
```

The numbering was given to each entity in ascending order of their appearance in the source material.

The other instance for including references is to provide additional information for potentially unknown definitions (e.g. *kashrut*, Jewish dietary laws), historical events (e.g. the Spanish-American war), laws (e.g. decree of March 21st 1890), or Jewish books of prayer and law (e.g. Shulchan Aruch), amongst other things:

```
<ref source="https://www.wikidata.org/wiki/Q211361">matzah</ref>
```

```
<ref source="https://www.wikidata.org/wiki/Q1713492">Young Czech Party</ref>
```

```
<ref source="https://www.wikidata.org/wiki/Q473670">January Uprising</ref> of <date when-iso="1863">1863</date>
```

These references are not further highlighted other than being formatted as a link in blue colour and underlined. In most cases, users are lead to Wikidata except for specific laws which lead to a database collecting past constitutions and codices in one instance¹⁴⁴ and the website of the Austrian parliament in another instance¹⁴⁵. These cases of references are primarily front-end functionality and do not provide further information in the annotation since it does not link to any glossary or

144 'Verfassungen in Der Europäischen Union', accessed 20 March 2026, <https://www.verfassungen.eu/>.

145 'Startseite | Parlament Österreich', accessed 20 March 2026, <https://www.parlament.gv.at/>.

become subject to a possible further analysis down the line. This referencing is for the readers of the finished edition on the website first and foremost.

9.3.4 Named Entities

Named Entities help transform the source into contextualised and structured data. In this case, people and places are tagged in the translated text and linked to their register entry. The elements `<placeName>` and `<persName>` respectively mark up the named entities to enable coloured highlighting on the website and aid in further computational analysis. The referencing element `<ref>` is used in addition to include the entity's URI to link to the register entry. Alternatively, both elements could be combined by supplementing the URI in `<persName>/<placeName>` and therefore avoid the double tagging. However, it was important for me to keep them separated. While only the name itself is wrapped by `<ref>`, the entity's tag also includes further information such as academic titles or house numbers. By using both elements, the text could be marked up to meet the project's intended purposes.

People

The names captured by `<persName>` do not always consist of a first and last name, but also include honorifics, titles, epitaphs, etc. People could be referenced by their job title or last name alone as well which necessitates a flexible mark up practice to include any mentions of people found in the text. The TEI guidelines allow for a detailed annotation distinguishing the different parts a name could consist of by using different elements. However, I chose a different route by annotating every person regardless of which part of their name is used with `<persName>`:

```
<persName><ref source="person005">István Tisza</ref></persName>
```

Further information can be given in the index entry. Still, a person's full name is seldom used in the text. Rather people are referred to by only their last name. Hence, often people had to be deducted by the surrounding contextual information:

```
<persName>Professor <ref source="person102">Billroth</ref>  
</persName> is the first who managed to cut out cancer of the  
innermost parts of the body.
```

This passage refers to Theodor Billroth, a German surgeon, who performed the first successful gastrectomy for gastric cancer. A large portion of insufficiently named people could be found due to similar contextualisation.

In some instances, the newspaper authors have completely omitted a person's name but rather refer to them by their title, function, or occupation. I was able to work out the majority of their identities with the given time frame and supplemented information. These people include religious figures (e.g. the Pope), rulers (e.g. Tsar, Afghan ruler, Japanese Emperor), politicians (e.g. Austrian's minister president, Grand Duke of Mecklenburg), and family of named people (e.g. Austrian Crown Princess Stéphanie's mother). In every case, the person is annotated with <persName> and is linked with their register entry in which their names and if possible a link to a Wikidata entry is given. However, their name is not added in the text itself as I decided against editorial supplementing the source text by adding commentary. If the person is named and then subsequently referred to by their occupation or role (e.g. "the rabbi") I also decided against highlighting every instance, especially in the historical retelling of rabbis' lives as these sections are often written in a dialogical format and it should be evident to the reader as to who is being referred to.

Places

The element <placeName> includes all locations mentioned without distinguishing between different types of places, neither in the register nor in the text. For

commonly occurring obstacles during the annotation process such as territory changes and multilingual names for places standardised solutions could be implemented.

The spelling in the text is normalised to Latin script of the contemporary name. If the current name and the contemporary name do not match, a note is supplemented in the index entry. This rule has but a few exceptions. For example, the Moldovan city Chişinău is also written as Kishinev as it was formerly known when it is mentioned in the context of the Kishinev pogroms:

```
<placeName><ref source="place015">Chişinău</ref></placeName>
```

That is also the root for the `<placeName><ref source="place015">Kishinev</ref></placeName>` pogrom.

Street names including house numbers are annotated with `<placeName>`, but the house number is excluded from the `<ref>` element which links back to the register. Only the street name without number is indexed to accurately provide the location and link to corresponding external databases:

```
<placeName><ref source="place003">Sonnengasse</ref> 27</placeName>
```

Throughout the source, country names are also used as acting subjects. For example, the war between Russia and Japan is covered extensively in the newspaper and the two countries are named in lieu of politicians or writing "the Japanese/Russian government". Every instance of a country being named, independent of functioning as a place or as an acting subject representing actions in war or trade, is annotated as a country. Another possibility would have been to differentiate place and actor via annotating the latter as an organisation and creating a third register. However, as there is no benefit in this practise and would rather complicate the process, every place mentioned is annotated as a place.

9.3.5 Temporal Information

Annotating temporal information and subsequent normalisation adds helps to further contextualise the source content. On every first page, the date of publication is given in Hebrew and German.

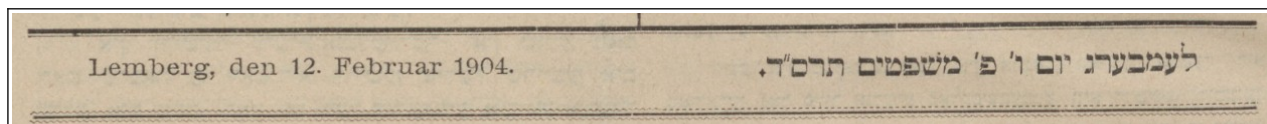


Figure 10: Publication date on the first page in German and Hebrew, *Der Emes'r Yud* (12.02.1904), 1.

A date is also given in the newspaper article either referring to recent events in the news section or occurrences long ago, especially in the historical retelling of rabbi's lives. Except for the Hebrew one, the date is translated into English and formatted according to the Gregorian calendar. This then is annotated with the `<date>` element and normalised within the *when-iso* attribute:

On the `<date when-iso="1903-12-31">31st</date>` day

The Hebrew version of the publishing date given on the first page is a peculiar case as the date is not the converted Hebrew calendar, but rather it is an entirely different system. This date is made up of the day of the week, the weekly Torah portion being read during this week in the Jewish calendar, and the year according to the Hebrew calendar. The Hebrew equivalent of the Gregorian date of March 8th 1904 is pictured in Figure 11. It reads from right to left: **Lemberg** (Lviv) **yom 3** (third day of the week = Tuesday) **Vayakhel Pekudei** (weekly Torah portion; in this case, two portions are read to fit all the portions in one lunisolar year) **5664**. This is equal to the Hebrew date of 21st of Adar 5664 or March 8th 1904 in the Gregorian calendar.



Figure 11: The Hebrew date of publication, *Der Emes'r Yud* (08.03.1904), 1.

To annotate this date correctly, it is necessary to indicate that another calendar system is used. The *calendar* attribute and *when-custom* aid in the accurate mark up of the transliterated date. Additionally, the `<foreign>` tag is needed as it is not translated, but transliterated with the value therefore reflecting the Hebrew language in Latin script. The dating according to the Hebrew calendar is normalised in the attribute, but in the text, the dating as in the source is used. The Torah portions are linked to Wikidata:

```
<date calendar="Hebrew" when-custom="5664-12-21">
<foreign xml:lang="he-Latin">Yom Shlishi
<ref source="https://www.wikidata.org/wiki/Q927246">Vayakhel</ref>
<ref source="https://www.wikidata.org/wiki/Q1164194">Pekudei</ref>
5664</foreign></date>
```

In the history articles, the authors at times provide the date in the Jewish calendar and Gregorian calendar format. Both dates are annotated and supplemented with according attributes as seen above: *calendar* and *when-custom* for the Hebrew dating and *when-iso* for the Gregorian.

9.4 Index Structure

The indices are lists of the people and places mentioned in the text, one for each entity type. Every entry contains one entity in their respective register. The XML for the index is a simple file consisting of a `<teiHeader>` and the entity entries. The schema for the places consists of the following elements and attributes:

```

<place xml:id="place###">
  <placeName></placeName>
  <location type="coords">
    <geo></geo>
  </location>
  <location type="located_in_place">
    <placeName key="place###"></placeName>
  </location>
  <idno type="Geonames">https://www.geonames.org/</idno>
  <idno type="Wikidata">https://www.wikidata.org/wiki/Q</idno>
  <note></note>
</place>

```

Mandatory elements are `<place xml:id="place###">`, `<placeName>`, and `<location type="coords">`. Ideally, the Geonames and Wikidata links are given in `<idno>` as external sources. Because geographical data on Wikidata can at times lack information and provide incorrect coordinates¹⁴⁶, it is vital to link to another geographical database as well. Whereas Wikidata contains information about the place and can link to further sources, Geonames is used for the coordinates which indicate the location in the incorporated map and possibly lead to further location-specific sources apart from Wikipedia¹⁴⁷. If the indexed place is smaller than a continent or ocean, `<location type="located_in_place">` links to the corresponding place entity. In some cases, further information was necessary to contextualise historical names or places that do not exist anymore, provided in `<note>`. A complete place entry could look like this:

```

<place xml:id="place204">
  <placeName>Livonia</placeName>
  <location type="coords">
    <geo>58 26</geo>
  </location>

```

146 Zhao, 'A Systematic Review of Wikidata', 870.

147 Dirk Ahlers, 'Linkage Quality Analysis of GeoNames in the Semantic Web', *Proceedings of the 11th Workshop on Geographic Information Retrieval*, 30 November 2017, 1–2, <https://doi.org/10.1145/3155902.3155904>.

```

<location type="located_in_place">
  <placeName key="place203">Latvia</placeName>
</location>
<location type="located_in_place">
  <placeName key="place206">Estonia</placeName>
</location>
<location type="located_in_place">
  <placeName key="place100">Europe</placeName>
</location>
<idno type="Geonames">
https://www.geonames.org/457850/livonija.html</idno>
<idno type="Wikidata">
https://www.wikidata.org/wiki/Q183464</idno>
<note>Historical region in today's Latvia and Estonia.</note>
</place>

```

Similar, the XML file for the person index begins with the TEI header and lists the people following this schema:

```

<person xml:id="person###">
  <persName>
    <forename></forename>
    <surname></surname>
  </persName>
  <birth>
    <date when-iso="YYYY-MM-DD">YYYY-MM-DD</date>
  </birth>
  <death>
    <date when-iso="YYYY-MM-DD">YYYY-MM-DD</date>
  </death>
  <idno type="URL" subtype="Wikidata">
https://www.wikidata.org/wiki/Q</idno>
  <occupation></occupation>
</person>

```

Every element in the entity entry contains information, except for the identifier <idno>. However, there are difficulties obtaining the same information for every person. The source also includes people who lived before the establishment of last names. In these cases, I decided to incorporate the epithets (e.g. "of Germany", "the

Great") or a forward slash ("/") if a name does not include any further identification. Two thirds of the entries contain a link to a Wikidata page or a similar database in case the person is not featured on Wikidata. Wikidata provides multilingual, crowdsourced content, but it lacks historical and specific local data which can also lack in quality. However it was then possible to substitute with other sources¹⁴⁸. Other sources I have chosen to link to include the Jewish Encyclopedia¹⁴⁹ and a crowd sourced genealogy website called Geni¹⁵⁰. Although Geni is not a scientific ideal open source database it contains knowledge which could later aid in adding new data to other open source databases. Therefore, I decided to include it if no entries in other sources exist. Still, a third of the people mentioned in the source, for example, advertisers or locally known people, cannot be traced back to an actual person and therefore could not be linked to external sources.

The life data is normalised to the Gregorian calendar in every instance. Unknown birth and death dates are supplemented by "Unknown" to avoid leaving empty spaces. The dates are also supplied in the attribute *when-iso*, provided that precise dates are given. Approximate birth and death dates are supplemented by *Not-Before* and *Not-After* attributes such as the biodata for the biblical figure Abraham:

```
<persName>
  <forename>Abraham</forename>
  <surname>/</surname>
</persName>
<birth>
  <date notBefore="-2167" notAfter="-1792">Between 2167 BCE
    and 1792 BCE</date>
</birth>
<death>
  <date notBefore="-2050" notAfter="-1638">Between 2050 BCE
```

148 Zhao, 'A Systematic Review of Wikidata'.

149 Jew. Encycl., 'Jewish Encyclopedia'.

150 'Geni.Com', accessed 2 December 2025, <https://www.geni.com/>.

```
        and 1638 BCE</date>
    </death>
```

Information about every person is provided in the <occupation> element. This includes their job, life dates, or why they are mentioned in the newspaper. An entry with all information provided would look like this:

```
<person xml:id="person131">
  <persName>
    <forename>Carl Julius</forename>
    <surname>Späth</surname>
  </persName>
  <birth>
    <date when-iso="1838-04-12">1838-04-12</date>
  </birth>
  <death>
    <date when-iso="1919-04-02">1919-04-02</date>
  </death>
  <idno type="URL" subtype="Wikidata">
    https://www.wikidata.org/wiki/Q1731850</idno>
  <occupation>German watchmaker, built astronomical
    clock.</occupation>
</person>
```

About one third of all people mentioned could not be found in any external sources. If the only information available stems from the source itself, meaning that I could not find further external data to verify a person, only the data in the source could be used to fill in the gaps. Every appearance of a name is highlighted and leads to their index independent of how often a person is mentioned – once in the entire source collection or seven times in one paragraph. Therefore, by not creating an entry for an unknown person, the edition would have been incomplete. It is also important to create an entry for each entity in case future research reveals more information and allows to link back to other datasets, like this edition, to further enrich the available data. An example for a person with little information is Dr Garfunkel, a Jewish doctor in Odessa who was sent to the battlefield to care for wounded soldiers in the battle

between Russia and Japan instead of another Jewish doctor who had before been picked at random by drawing a lot. Given that not a lot of local doctors are featured in Wikidata and that only a last name was given, it is almost impossible to verify who this person was without diving into local archives. His entry looks as follows:

```
<person xml:id="person206">
  <persName>
    <forename>/</forename>
    <surname>Garfunkel</surname>
  </persName>
  <birth>
    <date>Unknown</date>
  </birth>
  <death>
    <date>Unknown</date>
  </death>
  <occupation>Jewish doctor.</occupation>
</person>
```

A case could be made to create Wikidata entries to contribute information to the Linked Open Data web space. However, I decided not to compile this information given that the resources for additional research about each person was limited and too time consuming as an additional task for this project.

9.5 Challenges in Annotation

Some challenges encountered while transforming the material from a scanned Yiddish source to an interoperable English translated file system have already been touched upon. This section is meant to summarise these obstacles as well as discuss further difficulties that occurred during this process. These include the website's layout, lacking information in external sources, and creating a valid XML syntax.

The layouts of the source material and the website differ in appearance. Since the focus was on the translation and annotation to create a functional database, I did

not prioritise providing the correct layout by switching between one and two columns like the source material but instead I kept one column throughout the text. The page break and column break elements are kept in the XML files in case future endeavours call for accurate layout, but they do not transform into a valid HTML syntax in the current implementation. Furthermore, the punctuation follows a different scheme in the translation. The sentences in the original source could span over multiple lines, often interspersed with commas which functioned as a sentence breaker. The differing English and Yiddish syntax furthermore complicated writing readable sentence structures. Therefore, paragraphs function as the orienting text section wherein the punctuation is shifted to accommodate easy and fluent reading, prioritising readability over exact language structures.

The division elements are numbered for structure and order. However, the advertisements on the last page are the only section with a deeper text division level. Therefore, I decided to add the attribute *type* and the value "ad" to every advertisement entry as well as add the "ads" value to indicate that the last <div> segment contains the advertisements:

```
<div n="12">  
  <div type="ad" n="1">
```

Ambiguity was another problem I kept encountering throughout the completion of this project. Information about two thirds of all people mentioned could be found in external sources while the remaining third remained unknown. This was solved by still creating a reference and entry in the index while leaving the identification element empty. However, all places mentioned could be identified and located, but not every location's information is present on any database due to their size (i.e. a street) or due to them not existing anymore. For example, Austria-Hungary is not indexed in Geonames, but on Wikidata and therefore provides the coordinates used

back when it still existed. If neither Wikidata nor Geonames has an entry, the coordinates from OpenStreetMap were used to locate the place on a map.

Furthermore, due to the translation process, places or people could be incorrectly translated and referenced, especially if little further information about an entity was provided. I am secure in the decisions I made to append knowledge and link further databases but errors in the tagging process cannot be ruled out.

10 Transformation Pipeline

During the transformation process into HTML, the XML files are fed to the XSL transformation stylesheets via the cookiecutter scripts. The scripts work as a processor to transfer the documents and replace the template tags with names found in the cookiecutter¹⁵¹ files. Crucially, the stylesheets replace the XML/TEI elements with HTML elements. This structure can then display the files as an interactive website. In this case, the documents on GitHub are transformed by XSLT into HTML to build a cohesive webpage, including index, toolbar, and navigation. Theoretically, the XML files do not need to follow a standard as long as the stylesheets translate the tags used into universally understood and processable HTML. Still, using standards throughout the project is important for interoperability and reusability purposes. The workflow transforming XML via XSLT into HTML is depicted in Figure 12.

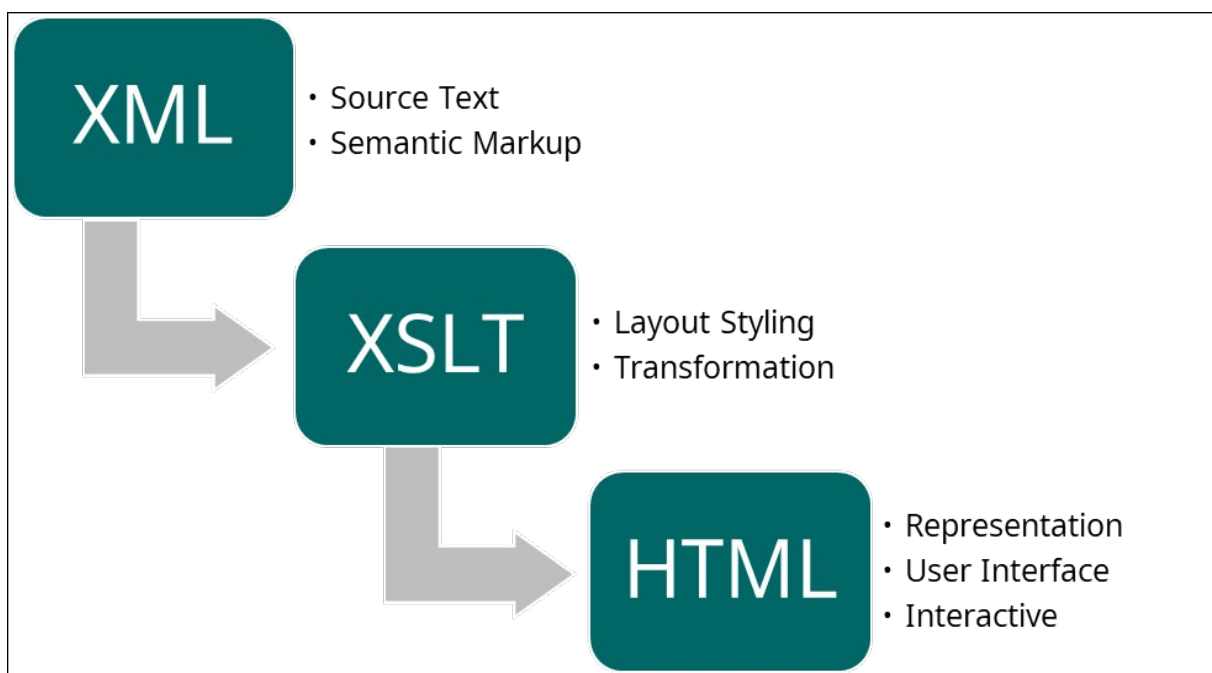


Figure 12: The flow of data from XML to HTML.

151 'Cookiecutter — Cookiecutter 1.7.2 Documentation', accessed 20 March 2026, <https://cookiecutter.readthedocs.io/en/1.7.2/README.html>.

This chapter describes the steps and infrastructure needed for the transformation pipeline: the cookiecutter library, the content of the transformation scripts, the directory structure including dependencies for deployment, the command line functionality, and the front-end implementation.

10.1 Cookiecutter

Cookiecutter is a Python templating library created to facilitate data science projects by providing a skeletal framework for building custom templates – called cookiecutters – for a myriad of different uses. To set up a template, no knowledge of Python is required due to its setup using the Jinja templating engine, hence why basic skills in operating Jinja are required¹⁵². These templates provide an organisational structure for data, documentation, code, reports etc. but remain flexible for adaptation¹⁵³. For this project, the template was already available and did not call for additional coding experience in Python or Jinja. The template provided by the ACDH, which is based on the cookiecutter library, is beginner- and user friendly. Therefore, the core concepts and execution semantics of Cookiecutter will not be explored further, but are described in detail in Dingel et al¹⁵⁴. Rather, the focus hereby lies in the created directory structure and workflow logic that trigger the transformation and website deployment.

10.2 Transformation Script

The XSLT scripts translate a language universally used for edition projects to a language constructing a usable static website in HTML. Each element used in the

152 Juergen Dingel et al., 'Project Templating and Onboarding With Cookiecutter: Foundations, Uses, and Guidelines', *Software: Practice and Experience* 56, no. 1 (January 2026): 44, <https://doi.org/10.1002/spe.70024>.

153 Tim Rooker et al., 'A Standardised Framework for Data Science in Advanced Manufacturing Systems', *Procedia CIRP* 130 (2024): 1670, <https://doi.org/10.1016/j.procir.2024.10.298>.

154 Dingel et al., 'Project Templating'.

TEI/XML files need to have a corresponding code to transform the XML element into a valid HTML element. This procedure splits the workload into smaller tasks that acknowledges labour divisions in group projects. Members can concentrate on their field of expertise without having to deepen their knowledge in the technical implementation of every language. At the same time, stylesheets allow for well-structured XML documents on one side as well as a valid HTML website on the other side without compromising either functionality or overcomplicating the process.

There are three distinct usages of the stylesheets for this project: style and layout, referencing named entities, and implementation of metadata. All cases will be explained and example codes will be given to illustrate how the stylesheets interact with the annotated source text.

10.2.1 Formatting and Layout

Formatting

The overarching stylesheet for the entire website is `shared.xml`. This file includes common elements which are universally used throughout the project. Additionally, each section of the website is based on a separate stylesheet that formats the XML contents. This way, contradicting formatting preferences can be implemented without over-complicating the code. In addition to translating the element, further styling can be implemented to override the standard HTML element formatting. The transformation script for the row of a table does not require additional styling for this purpose and the script hence only transforms:

```
<xsl:template match="tei:row">
  <xsl:element name="tr">
    <xsl:apply-templates/>
  </xsl:element>
</xsl:template>
```

The template matches the TEI element <row> and transforms it into the HTML element <tr> with <xsl:element> without any further formatting. The third line then applies to the tags that are nested in-between. This is the basic syntax for a transformation code.

In some cases, the TEI element does not match with a corresponding valid HTML element. Here, the element is transformed into a element and corresponding class attribute. Hence, the annotated text can still be distinguished in HTML without any formatting the text on the website surface:

```
<xsl:template match="tei:date">
  <span class="date">
    <xsl:apply-templates/>
  </span>
</xsl:template>
```

Styling can also be implemented if the same TEI element is used for different formats. In this case, the type attribute specifies the intended purpose and look of the text. The text is annotated with the TEI element <hi> and supplemented by an attribute:

```
<xsl:template match="tei:hi[@rend='italic']">
  <em>
    <xsl:apply-templates/>
  </em>
</xsl:template>

<xsl:template match="tei:hi[@rend='bold']">
  <strong>
    <xsl:apply-templates/>
  </strong>
</xsl:template>
```

With this template, text tagged with

```
<hi rend="bold">Jewish Encyclopedia</hi>
```

in the XML document, is transformed to

```
<strong>Jewish Encyclopedia</strong>
```

on the website and displayed like this:

Jewish Encyclopedia.

It is vital that the source text is consistently annotated with the corresponding elements for the transformation to work. Inconsistencies or typos are not flagged as errors. Instead the incorrect parts are ignored by the stylesheets and therefore not transformed to HTML, as long as the syntax is valid. Therefore, knowing your code and understanding HTML is necessary in addition to using the correct XSLT code for this project.

Layout

Each subpage on the website is based on a stylesheet that creates the needed layout and builds the page. Other stylesheets are imported to supplement consistently used building blocks that are needed for each section. Then, HTML and occasionally CSS elements are integrated to provide the tree structure including columns and container sizes. The following code is a simplified and cut version of the XSLT file structuring the page for the translated and annotated source text:

```
<xsl:template match="/">
  <head>
    <xsl:call-template name="html_head" />
  </head>
  <body class="d-flex flex-column h-100">
    <xsl:call-template name="nav_bar" />
    <main class="flex-shrink-0">
      <div class="container">
        <div class="head">
          <div class="col-md-8 col-lg-8 col-sm-12"/>
        </div>
      </div>
    </main>
  </body>
</template>
```

```

<div id="editor-widget">
  <xsl:call-template name="annotation-
    options"/>
</div>
<div data-index="true">
  <div class="col-md-8 col-md-8 col-md-10
    text" style="flex-grow:1;text-align:left;">
    <xsl:apply-templates select="
      .//tei:body" />
  </div>
</div>
<head align="center">
  <xsl:for-each select="tei:head">
    <xsl:value-of select="text()" />
  </xsl:for-each>
</head>
<p align="left;" />
</div>
</main>
<xsl:call-template name="html_footer" />
</body>
</xsl:template>

```

Everything contained by `<xsl:template/>` generates the HTML for the website, but not everything is visible on the website. The metadata in `<head>` can only be viewed in the console. Everything contained by `<body/>` creates the visible website. The class defines the column size, in this case the entire length of the window, with a flexible container which reacts to de- and increasing the window size. This means that the website can also be viewed on other vertical devices by adjusting to a different size. After the navigation bar follows the header, extracting metadata for display. This process will be explained in Chapter 10.2.3. Then follows the editor widget and the text to window width ratio of the rest of the text is defined. The referenced templates are separately defined and called to style the text. Finally, the biggest section content-wise is introduced with the `<head>` and `<p>` elements before closing the main part of the body and ending the page with the page's footer.

Each page created with a different stylesheet can therefore consist of different structures, sizes, and layout choices, depending on what the editor and creator sees fit for the purpose. The other stylesheets in this project follow the same structure. Therefore, one explanatory layout code suffices to visualise how the structure transformation works.

10.2.2 Named Entity Reference

The named entities – people and places – also have to fulfil two different roles in the XML and HTML respectively. In XML consistent annotation is key for interoperability and further analysis. For the website, the entities have to be transformed into valid HTML including links to the registers. They can also be highlighted in colour, if the option is toggled in the micro editor as seen in Figure 13 and Figure 14.

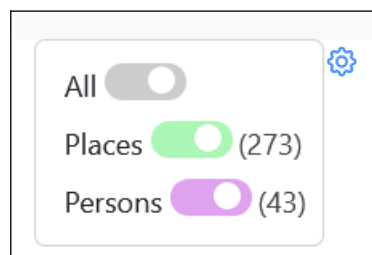


Figure 13: The micro editor.

(Vienna). This Monday, Emperor Franz Joseph felt something in his lung. He has complained about rheumatism and stayed in bed all day. On Thursday, he already felt better, but your majesty will stay inside for some days and not go outside. - This week the grand duke of Mecklenburg visited our emperor and was received with a big parade.

Figure 14: Highlighted people and places.

The TEI elements <persName> and <placeName> do not have matching HTML pendants. The transformation script thus turns the elements into elements and uses the class attribute to classify and format them:

```
<xsl:template match="tei:persName">
  <span class="persons entity" data-bs-toggle="modal">
    <xsl:apply-templates/>
```

```

    </span>
</xsl:template>
<xsl:template match="tei:placeName">
    <span class="places entity" data-bs-toggle="modal">
        <xsl:apply-templates/>
    </span>
</xsl:template>

```

This script matches one part of the entities markup. Another transformation script is responsible for linking them to their register entry:

```

<xsl:template match="tei:ref">
    <xsl:variable name="source">
        <xsl:value-of select="./@source" />
        <xsl:value-of select="./@key.html" />
    </xsl:variable>
    <a href="{ $source }" target="_blank">
        <xsl:apply-templates/>
    </a>
</xsl:template>

```

The `<ref>` element is turned into an `<a>` tag which defines a hyperlink and the *source* attribute becomes an *href* attribute. The attribute *target="blank"* opens a new window allowing the user to remain on the page as well as following the link to the register entry. Together, the transformation turns this XML snippet

```

<persName><ref source="person065">Moses</ref></persName>

```

into this HTML:

```

<span class="persons entity badge-item" data-bs-toggle="modal">
    <a xmlns="http://www.w3.org/1999/xhtml" href="person065"
        target="_blank">
        Moses
    </a>
</span>

```

In combination with the JavaScript and CSS code responsible for the micro editor, the named entities can be toggled on and off by the click of a button:

```
<span class="persons entity badge-item pers purple" data-bs-  
toggle="modal">  
  <a xmlns="http://www.w3.org/1999/xhtml" href="person065"  
    target="_blank">  
    Moses  
  </a>  
</span>
```

The collaboration between the XML, XSLT, and JS files allows the TEI annotation to be precise for future research while also supporting a richer website usage, enhancing the benefits for human and machine-supported experiences.

10.2.3 Metadata Implementation

Most metadata is not visible on the website. However, some information is vital and needs to be viewable for the general user. Besides the metadata in each XML file, saved in the <teiHeader>, the project's parameters are listed in params.xsl. This contains information about the directory, project link, title, and logo. The parameters are imported into the header, footer, and navigation bar by stylesheets. Hence, possible alteration of the project title and GitHub URL is easily changeable across the website by editing the params.xsl file only. One parameter, the short project title is implemented in the navigation bar with this code:

```
<a class="navbar-brand" href="index.html">  
  <xsl:value-of select="$project_short_title" />  
</a>
```

```
<a class="navbar-brand" href="index.html">The Genuine Jew</a>
```

Another parameter, the link to the GitHub repository, is linked in the website footer:

```

<div class="float-end me-3">
  <a href="{ $github_url }">
    <i class="bi bi-github" />
  </a>
</div>

<div class="float-end me-3">
  <a href="https://github.com/lug-robin/tgj-data/">
    <i class="bi bi-github"></i>
  </a>
</div>

```

Variables are created by fetching metadata stored in the TEI header to convey and link information usually stored away. The edition.xsl stylesheet creates a section with metadata to be displayed before the contents of the XML documents themselves follow. At the top of every newspaper entry, the title of the corresponding file is followed by a link to the facsimile and a download button using variables:

```

<xsl:variable name="teiSource">
  <xsl:value-of select="data(tei:TEI/@source)" />
</xsl:variable>
<xsl:variable name="doc_title">
  <xsl:value-of select=".//tei:titleStmt/tei:title[2]/text()" />
</xsl:variable>
<xsl:variable name="teiFacsimile">
  <xsl:value-of select="data(.//tei:facsimile/@source)" />
</xsl:variable>

```

These placeholders are then set at the wanted position in the structure:

```

<div class="col-md-8 col-lg-8 col-sm-12">
  <h1 align="center">
    <xsl:value-of select="$doc_title" />
  </h1>
  <h2 align="center">
    <a href="{ $teiFacsimile }" target="blank">Facsimile</a>
  </h2>
  <h3 align="center">
    <a href="{ $teiSource }" target="blank"

```

```

        download="{${doc_title}}">
        <i class="bi bi-download" title="TEI/XML" />
    </a>
</h3>
</div>

```

During the transformation process, the placeholders retrieve the data from the TEI header to create the following HTML code:

```

<div class="col-md-8 col-lg-8 col-sm-12">
    <h1 align="center">1904-02-19</h1>
    <h2 align="center">
        <a href="https://anno.onb.ac.at/cgi-content/anno?
        aid=emj&datum=19040219" target="blank">Facsimile</a>
    </h2>
    <h3 align="center">
        <a href="1904-02-19.xml" target="blank" download="1904-02-
        19">
            <i class="bi bi-download" title="TEI/XML"></i>
        </a>
    </h3>
</div>

```

On the website, this code, created from the XML metadata, displays the variables as shown in Figure 15:

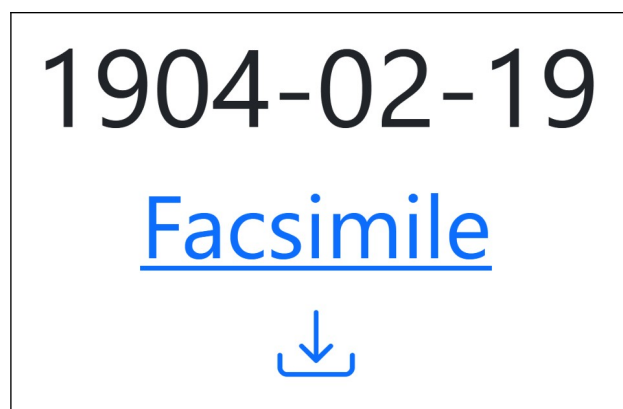


Figure 15: Metadata variables on the website

Stylesheets are a multifaceted tool which can be used in various ways. I demonstrated the three ways they were implemented on the website. Not only does

XSLT transform XML into HTML, but it also adds universally unified data, information, and function adjustable by altering only one XSLT file instead of every XML file.

10.3 Directory Structure

The project is stored in a local repository in which all changes are then pushed to the remote repository. The directory contains four main folders: data, html, shellscripts, and xslt as seen in Figure 16. The folder structure was given by the provided template and needed little alteration to fit the project purpose. Therefore, the dependencies and deployment triggers within and across the folders were already constructed after the initial setup.

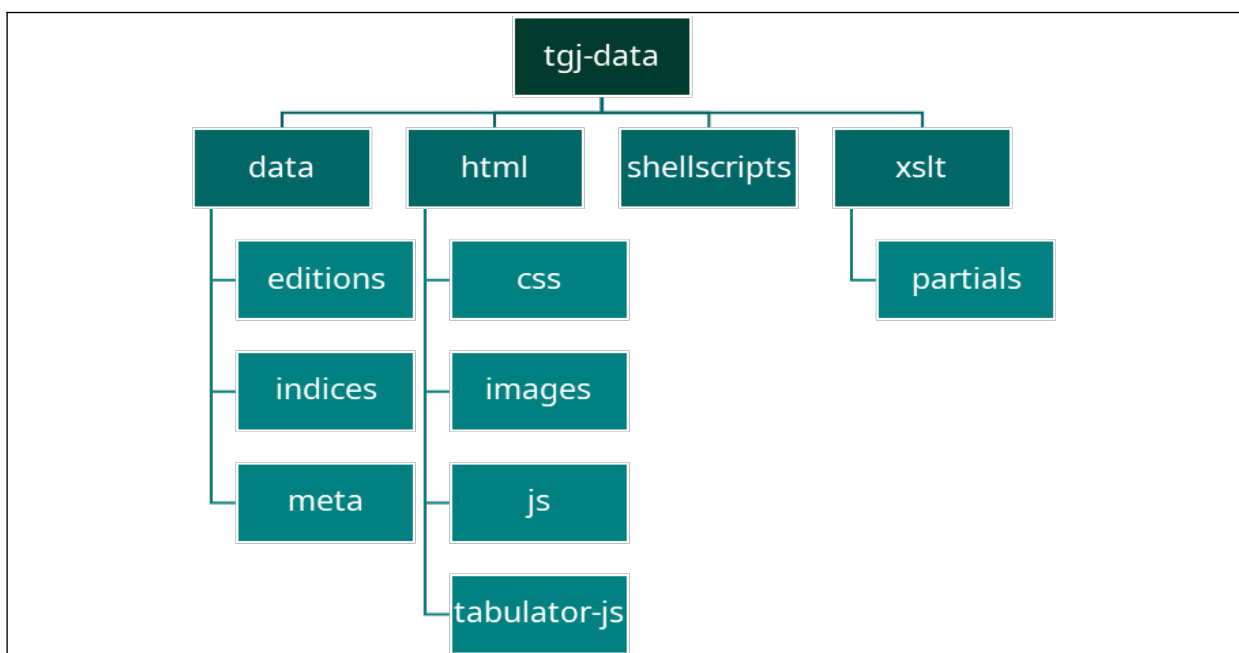


Figure 16: The project's directory structure.

In the base folder tgj-data are two files: build.xml and ss_config.xml. The latter is responsible for the static search nested in its own tab on the website. It interacts with two TXT files also saved in the same folder: words.txt functioning as a dictionary and stopwords.txt which can be used to exclude certain or common words in further

linguistic analyses. Build.xml executes the website build. It works like a spine connecting the different parts to construct a functioning, interconnected website.

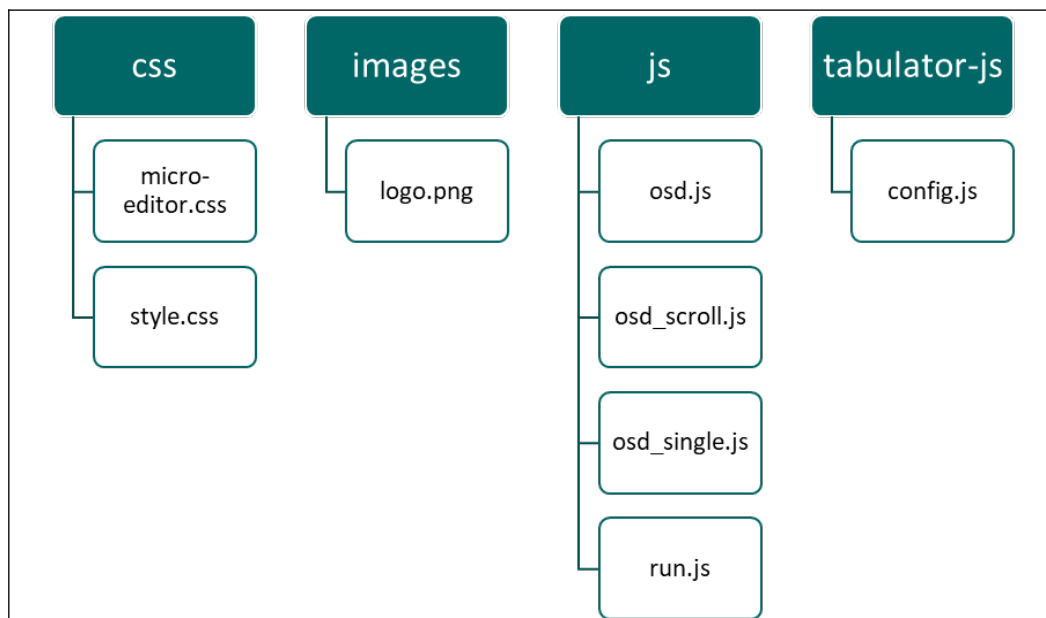


Figure 17: The *html* folder directory

The XSLT, JavaScript, and CSS code in the `xslt` and `html` folders respectively build the framework for the XML files stored in the data folders. The XSLT code can be understood as the building material while the scripts are the means to set it into place. The JavaScript and CSS code files are complimentary building blocks not needed for the basic outline, but for additional features. These files include OpenSeadragon¹⁵⁵, an open-source, web-based viewer for the layout coded in JavaScript, and a micro editor in CSS to highlight named entities. Finally, the website icon displayed in the browser tab is also saved in this directory in the folder *images*. Connecting this directory with a GitHub account can already create a website since all the building blocks are connected with and depend on each other.

The XML stylesheets in the `xslt` folder are divided in two groups: one for the website layout, the templates for every page of the website and the other containing the

155 'OpenSeadragon', accessed 21 December 2025, <https://openseadragon.github.io/>.

styles of the different parts of the website. The templates are stored in the folder partials.

File Name	Description
aot-options.xsl	XML implementation of the micro editor coded in JavaScript (run.js) and CSS (micro-editor.css)
html_footer.xsl	Styles the GitHub URL and the dark mode widget.
html_head.xsl	Creates SEO metadata for website and mobile and collects the Cascading Stylesheets implemented on each page.
html_navbar.xsl	Menu on the top of the page to switch between the edition, registers, search function, and further resources.
params.xsl	Parameter for the project implementing what was entered at the start of the project using the ACDH cookiecutter.
person.xsl	Register entry for people. Creates the layout as well as changes the semantic to create functioning links to external databases.
place.xsl	Register entry for places. Creates the layout as well as changes the semantic to create functioning links to external databases.
shared.xsl	Template for the whole project. Includes the transformation styles for universally used tags, layouts, and the highlighting options for named entities.

File Name	Description
tabulator_dl_buttons.xsl	The "Download Table". Creates buttons to download the tabular register data of the editions and indices in CSV, JSON, and HTML. Works together with tabulator_js.xsl
tabulator_js.xsl	The JavaScript code for the "Download Table" to download the tabular register data of the editions and indices in CSV, JSON, and HTML. Works together with tabulator_dl_buttons.xsl

Table 2: Stylesheet templates

The templates saved in the partials folder are imported into other documents by their ID. The name of the templating file is written in the code with which it can be imported into other documents by the following lines:

```
<xsl:template name="annotation-options"/>

<xsl:import href="./partials/aot-options.xsl"/>

<xsl:call-template name="annotation-options"/>
```

The first line is the template name given in the source file with which it can be referenced. The second line is written at the beginning of the target file doing the importing. The final line is written at the position where the template would be located in the later HTML website file. By importing other stylesheets, the code does not have to be repeated in each file. This would have resulted in long, confusing scripts and was thereby avoided. For example, the head, footer, and navigation bar remain stable and can be altered on the whole website by changing only one file.

The other stylesheets contain not only the transformation style for elements, but additionally the layout of the specific page, including margin and position of the different elements:

File Name	Description
editions.xml	Stylesheet for the newspaper text segment.
index.xml	Stylesheet for the landing page of the website
listperson.xml	Stylesheet for the tabular register page for people linking to the register entry styled in person.xml
listplace.xml	Stylesheet for the tabular register page for places linking to the register entry styled in place.xml
meta.xml	Stylesheets for all other subpages not included in other stylesheets: the landing page / about section and the further resources.
search.xml	Stylesheet for the search page
toc.xml	Stylesheet for the tabular register page for editions linking to the text entry styled in editions.xml

Table 3: Layout Stylesheets

These stylesheets are not imported into other files. Instead they are referenced in the build.xml file which connects the content to the layout:

```
<xslt in="./data/indices/listperson.xml" out="$
{target}/listperson.html" style="./xslt/listperson.xsl">
    <factory name="net.sf.saxon.TransformerFactoryImpl" />
    <classpath location="{basedir}/saxon/saxon9he.jar" />
</xslt>
```

The content of the website is stored in the data folder. The translated source text is saved in the editions folder, one XML file for each published paper issue that was translated, which makes 13 files in total. The people and places registers are stored in the indices folder. The meta folder contains the landing page describing the project and a file listing tools, books, and websites used to create this project called resources.xml.

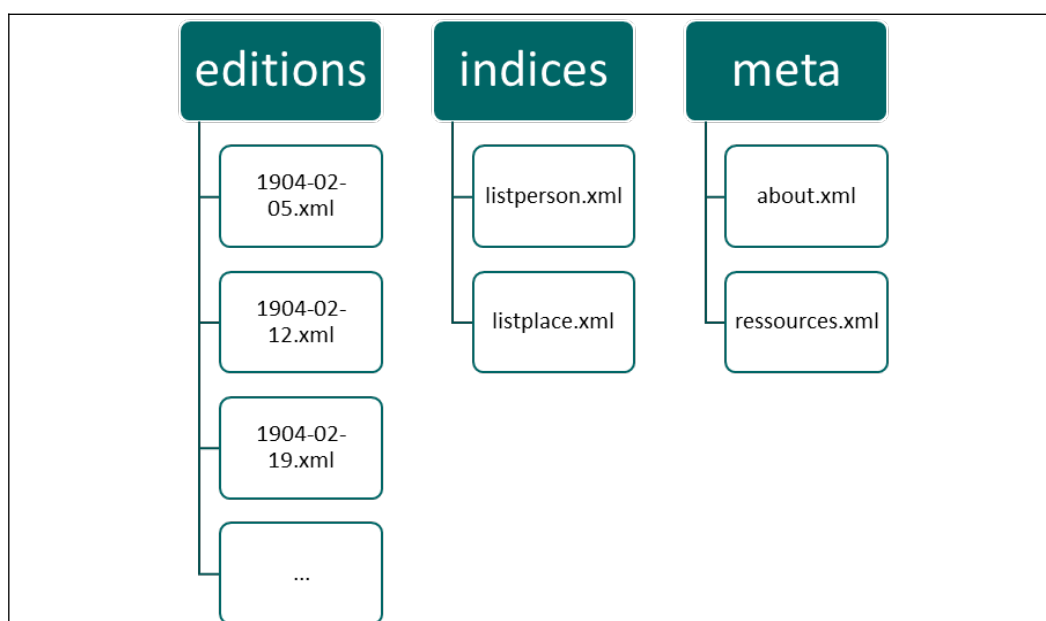


Figure 18: The data directory

The inherited framework included more files than were needed for this project, namely additional content or different layout options. Some of those files had dependencies to other documents which need to be deleted before removal was possible. For example, the structure included a two additional registers for more named entities. These files were also included in the build.xml file and page deployment would fail if the line deploying the registers was not removed from the code, resulting in an error. Therefore, great care was needed for restructuring the framework in order to maintain a functioning website.

10.4 Command Line Interface

The ACDH template provided a pre-configured command line workflow with shell scripts and a Linux based environment for which I chose Ubuntu. It is set up to interact with the local Linux drive I had installed on my Windows system. The shell scripts interpret the command lines given in the Ubuntu interface which in turn pushes the changes made to GitHub and the website. Instead of manual conversion, the framework makes use of automation scripts which are triggered by a series of commands. Apart from the initial setup in the shell, there are only a handful of commands which allow for an iterative workflow. The following commands are used regularly to inspect changes, add potential new files, and commit changes to the remote repository automatically triggering the changes to be built and deployed onto GitHub pages:

```
~/tgj-data$ git status
```

```
~/tgj-data$ git add .
```

```
~/tgj-data$ git commit -a
```

```
~/tgj-data$ git push origin master
```

After the changes have been pushed to the GitHub repository, the subsequent *pages build and deployment* workflow is triggered and the website is automatically updated. To see the changes that were made, the browser needs to be refreshed. Summarised, changes to the website are made in three steps:

1. Edit the code and save the files
2. Commit the changes in the Ubuntu interface to push to the remote repository
3. Refresh the browser after the workflow has completed.

This approach guarantees a consistent transformation pipeline and traceable changes due to the mandatory comment written when committing a change.

10.5 Frontend Implementation

At the end of the transformation pipeline emerges a static HTML5 website as a result. The majority of today's web is comprised of dynamic websites which need servers to process and databases to respond to user requests. Contrary, static websites are built using HTML, CSS, and JS which provide stable results without relying on server performance. Therefore, displaying the files as a static HTML website was a conscious choice to avoid high cost and time-consuming ongoing maintenance. After the initial setup, this website will not need to be maintained regularly. Furthermore, the data is automatically preserved, as a byproduct. Hence, even if the website is no longer hosted on GitHub pages, the HTML, XML, and XSLT are open standard formats and are not bound to any platform. Therefore, the data can be preserved migrated, reused, and archived without much loss of data¹⁵⁶. This matrix of low cost and high performance usage provides the ideal project template for projects in small teams or by students. The website is also split into several HTML files which allows a simple exchange of contents by exchanging one file instead one massive code document.

Creating a simple, user-friendly page was one of the main objectives of this project. Therefore, I also avoided implementing further gimmicks beyond a dark mode – which was already available in the template – and highlighting named entities. The code for changing fonts and font sizes is still available to be changed in the micro-editor code, but currently disabled to concentrate on the text itself. This code was also provided by the template. The next chapter focuses on the website, its usage and navigation.

156 Olivia Wikle and Evan Peter Williamson, 'Static Web Methodology as a Sustainable Approach to Digital Humanities Projects', *The Code4Lib Journal*, no. 60 (April 2025), <https://journal.code4lib.org/articles/18372>.

11 Digital Infrastructure

The visible end product of this project is the website on which the encoded and marked up files are executed for readers on a user-friendly surface. Most of the layout and design has been taken from the template with minimal alteration as not the design, but the content is the project's priority. This chapter explores the design, functions, infrastructure and hosting of the website as it is deployed on GitHub pages.

11.1 Navigation

The website's layout is kept simple and ensures that the user can visit any part of the website by navigating through the menu on the top of the site or follow the links embedded in the text on the landing page. The landing page titled "About the project" contains text that shortly explains the website and its origin. The navigation bar consists of four tabs as seen in Figure 19: Edition, Register – which drops down to reveal the People and Places indices – Search, and Further Resources.

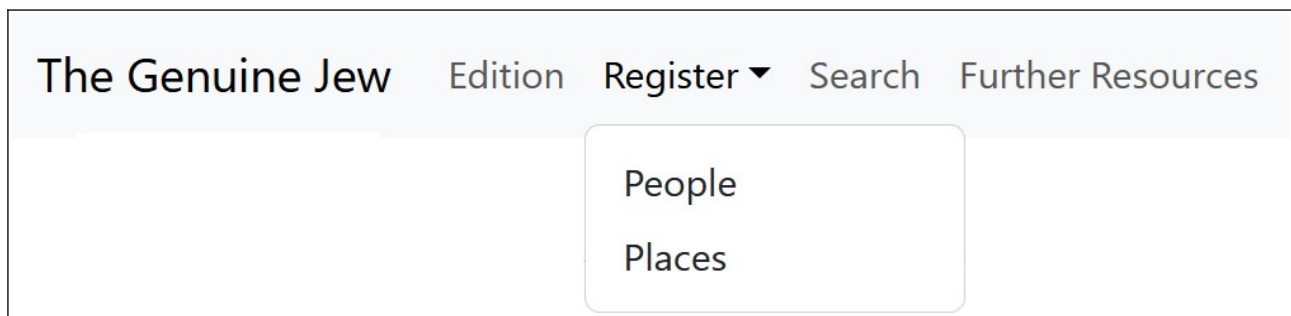
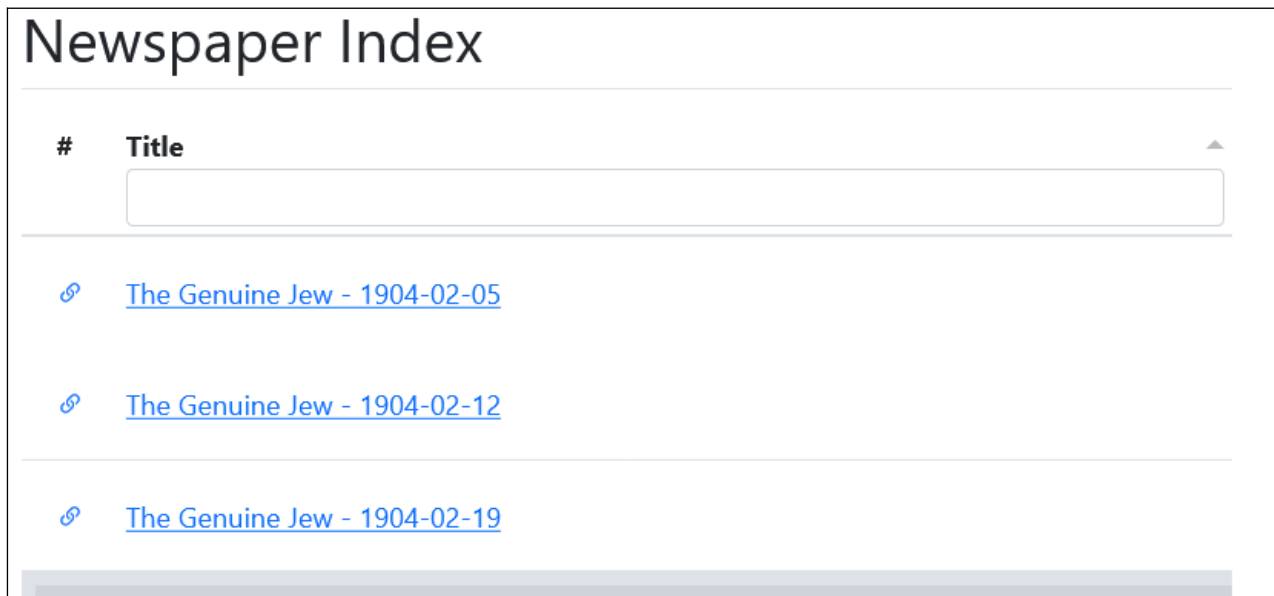


Figure 19: Navigation Bar

In total, there are six pages on the website: the landing page to which users can navigate by clicking on "The Genuine Jew" on the left, the edition register, organised in a table, the two registers, the search function, and finally the further resources which is a simple text like the landing page. Hence, every part of the website can be

found by two clicks at most and the page allows for fluent navigation between the different parts.

The editions are listed in a searchable and scrollable table of contents. The file names consist of the title of the newspaper and the publication date.



#	Title
	The Genuine Jew - 1904-02-05
	The Genuine Jew - 1904-02-12
	The Genuine Jew - 1904-02-19

Figure 20: Newspaper Index

This table links every paper translated in a periodically sorted list, immediately linking to the text itself. Both the text and the icon next to it function as links to the text. All three indices can be downloaded as CSV, JSON, or HTML files at the bottom of the list.

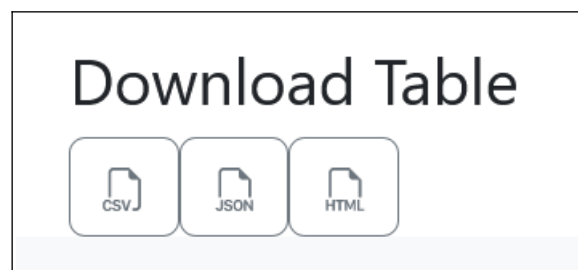


Figure 21: Download option for the indices.

The layout and font are both deliberately chosen for readability purposes. The chosen font family is a default sans serif and the layout aligned left since justification

alignment and serif fonts can hinder readability¹⁵⁷. The website partly mimics the source material for orientation purposes such as text alignment, page breaks, and font size for headings. The most notable difference however is, that the source material is printed in two columns for the majority of the text. Since keeping two columns for the website would have negatively impacted the reading experience because the source material is read right to left and the translation is read left to right, the website displays the text in one column only. Lines and borders used to arrange the text in the source have also not been added to the website. However, some data, for example the pricing options, have been adapted into a table layout for readability purposes. In conclusion, the text's layout has been kept to minimum styling as to not complicate the reading experience.

The page layout adapts to the size of the window via flex containers. Therefore, the page can be viewed on different devices in horizontal or vertical orientation. The layout also adapts to manual zooming and allows the text size to increase without causing awkward formatting errors. It is important for the website to be adaptable in this way. For easier navigation between the newspaper issues, arrows can be used to jump to the previous or next paper in line as seen in Figure 22.

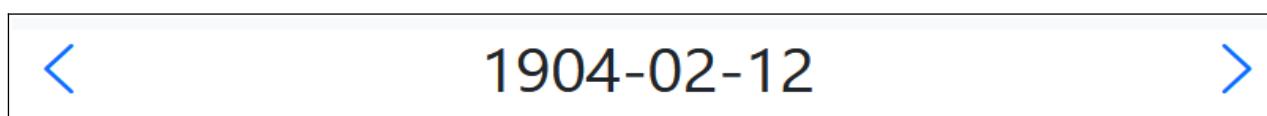


Figure 22: Newspaper header with navigation arrows

The entities registers are organised by a table including a search tab at the top and the possibility to order the items by first name, surname, and note or name, latitude or longitude respectively. Their standard order is by their UID which is not visible on

157 Veröffentlicht von domingos, *Barrierefreie Schrift-Gestaltung – Digitale Barrierefreiheit*, n.d., accessed 23 January 2026, <https://www.netz-barrierefrei.de/wordpress/barrierefreie-schrift-gestaltung/>.

the website. The UID was given in order of appearance in the text. Therefore, the people and places appearing in the first publications are the on the top of the list. In both cases, the data in the first column are used as links to go to the entity entry. For people, this is the surname, as pictured in Figure 23, and for places it is the name of the place.

“About the project” and “Further Resources” are both simple texts including links to parts of the website or external sources and otherwise follow the layout scheme of the editions.

Index of People		
Surname ▲	Name ▲	Notes ▲
Žáček	Jan	Vice President of the imperi...
Urusov	Sergey Dmitriyevich	Prince, govenor of Bessabar...
of Austria	Franz Joseph I	Emperor of Austria-Hungar...

Figure 23: Index of People

11.2 Search Function

Like any website on the internet, the content is searchable by a simple browser-based search. However, to forage through every edition at once, a full-text search bar powered by StaticSearch¹⁵⁸ was implemented. The results can contain variations of the sought out word as shown in Figure 24. In addition to the results, the documents in which the word appears and how often it is found are counted. The blue arrow

¹⁵⁸ Craig Buckler, ‘StaticSearch’, Publican.Dev, 17 June 2025, <https://publican.dev/tools/staticsearch/>.

next to the context should lead to the position in the text with the paragraph containing the word being highlighted. Unfortunately, the search does not allow for fuzzy results. Results are only shown if the word is written in the correct spelling, using the exact letters and including the right diacritics. Proper names of people and places can therefore only be found if they are spelled in the same manner as I have spelled them. It is also important to be able to look up temporal name variations of a place to receive all information needed. In these cases, other used names are listed in the register entry notes and can therefore be traced down as well.

Full-Text Search

Documents found: 7

[1904-03-29](#) Score: 4

Germany. The German emperor who is travelling in the Mediterranean for relaxation met with the King...

...Prague tried with constant effort for the rabbi to travel as a free man to Vienna, to face...

...known his orderliness, agreed that the Rabbi should travel alone, without a guard and without shackles....

Semitic tribes travelled closer and closer to Egypt described by an image...

↗

↗

↗

↗

Figure 24: Search function demonstrated with the word "travel"

The source material can also be explored in a non-linear fashion by making use of the indices. As every place and person is given its own register entry, users can look explore the corpus by location and people of interest. Although it could have been interesting, I decided against linking the entities out of the register with their appearances in the text because of the high volume of source material it would entail for some subjects. For example, looking for Russia in the source text via the

search bar yields almost 300 results – a number not feasible to include in the register entry. Instead, users can look for the entities indirectly by copying and pasting the name in the correct spelling from the index into the search bar and retrieve all instances where the word was mentioned in the translated text. Users can also click on entities in the text to go to their index entry and subsequently further external databases such as Wikidata and Geonames. An important feature of the index entry is the linkage to region the chosen place is a part of as demonstrated by Figure 25.

Lviv	
Part of	• Galicia • Austria-Hungary • Ukraine • Europe
Geonames	702550
Wikidata ID	Q36036

Figure 25: Place register entry for Lviv

These simple features therefore allow users to explore the material in a linear as well as non-linear fashion by subsequent and structured linkage between the different parts of the website and external databases.

11.3 Infrastructure and Hosting

The remote repository is on GitHub and the website is hosted by GitHub pages. The platform functions as a backup tool with version control as well as a methodological tool. Every change made can be traced in the GitHub workflows, thereby creating the website's documentation and providing a scholarly auditing trail. Hosting the data on an open access platform like GitHub permits computational reproducibility.

Additionally to storing the data locally, keeping the data safe on a remote page falls

in line with the sustainable practice of (self-)archiving produced data¹⁵⁹. GitHub pages deplores directly from the repository itself which allows to easily rebuild the website as long as the code exists. Furthermore, the service is free to use. By having both the XML and HTML on the same platform, the website can function as the visualised view of the data stored in the repository.

By the collaboration of the technical architecture explored in Chapter 9 and the user interface, the text has been made accessible to scholars and non-scholars alike. Without knowledge about TEI, XML, or HTML, the data can be explored via integrated links, but it can also be investigated further by means of the source code and the workflows documented by version control. Throughout the different means of encoding and transformation, the scan of a newspaper has been opened up into an interoperable, open, accessible, and searchable edition free to use for everyone interested.

159 Emily Escamilla et al., 'The Rise of GitHub in Scholarly Publications', in *Linking Theory and Practice of Digital Libraries*, vol. 13541, ed. Gianmaria Silvello et al., Lecture Notes in Computer Science (Cham: Springer International Publishing, 2022), 188–90, https://doi.org/10.1007/978-3-031-16802-4_15.

12 Conclusion

In this thesis I set out to explore the possibilities and boundaries of translating and transforming a Yiddish periodical into an interoperable digital edition. Yiddish press in Eastern Europe was an important agent of Jewish culture and politics, transmitting more than daily news and supporting the formation of a modern Jewish identity. The linguistic translation makes these sources available to a wider non-Yiddish speaking audience. The digital translation – the encoding and transformation – expands the accessibility by making the text both readable to humans and machines. I also showcased tools and methods that can provide assistance and facilitate this process. Ultimately, automatic text recognition and translation do not yet fulfil the qualitative standards for creating publishable results. Furthermore, Yiddish script and language varies greatly from region, time, context, and source and always needs human interpretation to provide contextual and cultural knowledge for an authentic understanding of a source text. At this point, human control and correction of an automatically transcribed and translated text would take as much if not more time than fully omitting computational assistance and doing everything manually. The translation and annotation in this project has hence been done manually.

The translated and annotated dataset is openly available for external researchers who could further examine *Der Emes'r Yud* through a historical, linguistic, social, or political lens. A comparison between the news coverage of *Der Emes'r Yud* and other local newspapers published in other languages could yield interesting results. One could also look at societal norms concerning education and the role of women as was written about multiple times in the source. As *Der Emes'r Yud* can also be clearly ascribed to a political movement, the paper could also fit in a wider analysis including socialist or Zionist tendencies in Eastern Europe or the wider Jewish diaspora. As the dataset is provided in non-proprietary formats, it could also be used

as training data for a specialised OCR model to eventually translate the entire corpus.

Future research could also venture farther into the development of standards for processing non-Latin scripts in scholarly digital editions and guidelines for annotation, especially for multi-lingual datasets using multiple fonts. Resources for digital editions of translated text should also be expanded and include more minority languages. Developments for Yiddish text recognition have already been made in the last few years, but I see great potential in the future to also include regional and temporal differences. In a next step this could lead to a broader study of Yiddish writing development throughout Europe.

When I started this project, I wanted to find out if transforming a Yiddish source into a digital edition for further research and wider accessibility is a worthwhile endeavour. The answer to this question is yes and no. For further usage, only translating the sources could be sufficient. Defining named entities and situating the findings in linked open web is significantly more work which can be omitted if a project does not call for it. A possible setting for a similar project could, for example, be executed with student collaboration. With the right tools (e.g. micro edition, a pre-existing website, or templates like a cookiecutter etc.) and guidelines, the greatest hurdle in completing a project like this is time and effort needed. Although I was able to gain valuable insights in every of the project's aspects, I would not conduct this project in the same manner again. However, undertaking a similar project in a team or lessening the project outcomes to only a translation without annotation to produce a linked open dataset would still yield comprehensive data with significantly lower workload. This master's project has therefore been a vital contribution in understanding what is available and what still needs to be done to work with minority languages in the Digital Humanities.

Bibliography

- Abugov, Netta, and Dorit Ravid. 'Noun Plurals in Israeli Hasidic Yiddish: A Psycholinguistic Perspective'. In *Yiddish Language Structures*, 9–37. Empirical Approaches to Language Typology 52. Berlin Boston, Mass: De Gruyter Mouton, 2014.
- Adamczyk-Garbowska, Monika. 'A Historical Decontextualization of Eastern Europe'. In *Yiddish and the Field of Translation: Agents, Strategies, Concepts and Discourses across Time and Space. In Cooperation with Marianne Windsperger*, 1. Auflage, 77–95. Schriften Des Centrums Für Jüdische Studien, Band 033. Göttingen: Böhlau Verlag Wien, 2020. <https://doi.org/10.7767/9783205210306>.
- Ahlers, Dirk. 'Linkage Quality Analysis of GeoNames in the Semantic Web'. *Proceedings of the 11th Workshop on Geographic Information Retrieval*, 30 November 2017, 1–2. <https://doi.org/10.1145/3155902.3155904>.
- Aptroot, Marion, and Roland Gruschka. *Jiddisch: Geschichte und Kultur einer Weltsprache*. Originalausgabe. Beck'sche Reihe 1621. München: Beck, 2010.
- Bauer, Ela. 'Commemoration and Jewish National Identity: Lviv in the Last Decades of the Nineteenth Century'. *Journal of Jewish Identities* 14, no. 1 (2021): 1–20.
- Benor, Sarah Bunin. 'Hebrew and Jewish Diaspora Languages'. In *The Routledge Handbook of Judaism in the 21st Century*, 89–110. Routledge Handbooks in Religion, Book 18. London, New York: Routledge, Taylor and Francis Group, 2023.
- . 'How Many Speakers Are There? Statistics on Jewish Language Vitality.' *Jewish Language Website*, 2024. <https://www.jewishlanguages.org/stats>.
- Berger, Shlomo. 'Translation, Interpretation and Transformation: Yiddish between Languages and Cultures'. In *Trans-lation, Trans-nation, Trans-formation: Übersetzen und jüdische Kulturen*, 155–66. Schriften des Centrums für Jüdische Studien, Band 21. Innsbruck: StudienVerlag, 2012.
- Berners-Lee, Tim. 'Linked Data'. 18 June 2009. <https://www.w3.org/DesignIssues/LinkedData.html>.
- Bizer, Christian, Tom Heath, and Tim Berners-Lee. 'Linked Data - The Story So Far'. *International Journal on Semantic Web and Information Systems* 5, no. 3 (July 2009): 1–22. <https://doi.org/10.4018/jswis.2009081901>.

- Bray, Tim, Jean Paoli, C. M. Sperberg-McQueen, Eve Maler, and François Yegau. 'Extensible Markup Language (XML) 1.0 (Fifth Edition)'. W3C, 26 November 2008. <https://www.w3.org/TR/REC-xml/>.
- Buckler, Craig. 'StaticSearch'. Publican.Dev, 17 June 2025. <https://publican.dev/tools/staticsearch/>.
- Cohen, Nathan. 'The Yiddish Press and Yiddish Literature: A Fertile but Complex Relationship'. *Modern Judaism - A Journal of Jewish Ideas and Experience* 28, no. 2 (April 2008): 149–72. <https://doi.org/10.1093/mj/kjn004>.
- . 'Yiddish Publishing Strategies in Eastern Europe (1880–1914)'. *East European Jewish Affairs*, 26 February 2025, 113–24. <https://doi.org/10.1080/13501674.2024.2400340>.
- Cohen Roman, Oren. 'When Yiddish Was Written in Latin Letters'. *Journal of Jewish Languages* 12, no. 1 (April 2024): 27–60. <https://doi.org/10.1163/22134638-bja10043>.
- 'Cookiecutter — Cookiecutter 1.7.2 Documentation'. Accessed 20 March 2026. <https://cookiecutter.readthedocs.io/en/1.7.2/README.html>.
- Costiuc Radosav, Augusta. "'Tsi Vos Toygt a Yudish Blat?' Of What Use Is a Yiddish Newspaper? Yiddish as a Language of the Press in Nineteenth-Century Romania'. *Studia Judaica* 22 (2017): 21–49. <https://doi.org/10.24193/sj.2017.v22.2>.
- Couture, Beatrice, Farah Verret, Maxime Gohier, and Dominique Deslandres. 'The Challenges of HTR Model Training: Feedback from the Project Donner Le Gout de l'archive a l'ere Numerique'. *Journal of Data Mining & Digital Humanities Historical Documents and...* (December 2023): 10542. <https://doi.org/10.46298/jdmdh.10542>.
- Daniels, Peter T. 'Fundamentals of Grammatology'. *Journal of the American Oriental Society* 110, no. 4 (December 1990): 727–31. <https://doi-org/10.2307/602899>.
- Dingel, Juergen, Alex Sun, Hesham Elabd, Nathanael Yao, Andrew Boulos, and Ali Tizghadam. 'Project Templating and Onboarding With Cookiecutter: Foundations, Uses, and Guidelines'. *Software: Practice and Experience* 56, no. 1 (January 2026): 42–63. <https://doi.org/10.1002/spe.70024>.

domingos, Veröffentlicht von. *Barrierefreie Schrift-Gestaltung – Digitale Barrierefreiheit*. n.d. Accessed 23 January 2026. <https://www.netz-barrierefrei.de/wordpress/barrierefreie-schrift-gestaltung/>.

Doran, Michelle. 'Reflections on Digital Scholarly Editions: From Humanities Computing to Digital Humanities, the Influence of Web 2.0, and the Impact of the Editorial Process'. *Variants* 15–16 (2021): 213–30. <https://doi.org/10.4000/variants.1414>.

Escamilla, Emily, Martin Klein, Talya Cooper, Vicky Rampin, Michele C. Weigle, and Michael L. Nelson. 'The Rise of GitHub in Scholarly Publications'. In *Linking Theory and Practice of Digital Libraries*, vol. 13541, edited by Gianmaria Silvello, Oscar Corcho, Paolo Manghi, Giorgio Maria Di Nunzio, Koraljka Golub, Nicola Ferro, and Antonella Poggi, 187–200. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2022. https://doi.org/10.1007/978-3-031-16802-4_15.

Everett, Leila P. 'The Rise of Jewish National Politics in Galicia, 1905 - 1907'. In *Nationbuilding and the Politics of Nationalism: Essays on Austrian Galicia*, 3rd ed., 109–77. Monograph Series / Harvard Ukrainian Research Institute. Cambridge: Harvard Univ. Press, 1995.

'FAIR Principles'. *GO FAIR*, n.d. Accessed 22 January 2026. <https://www.go-fair.org/fair-principles/>.

Fishman, Joshua A. 'Czernowitz Conference'. In *The YIVO Encyclopedia of Jews in Eastern Europe*. New York, NY: YIVO Institute for Jewish Research, 12 February 2010. <https://encyclopedia.yivo.org/article/2221>.

Fleischer, Jürg. 'The (Original) Unity of Western and Eastern Yiddish: An Assessment Based on Morphosyntactic Phenomena'. In *Yiddish Language Structures*, 107–23. Empirical Approaches to Language Typology 52. Berlin Boston, Mass: De Gruyter Mouton, 2014.

Franzini, Greta, Melissa Terras, and Simon Mahony. 'Digital Editions of Text: Surveying User Requirements in the Digital Humanities'. *Journal on Computing and Cultural Heritage* 12, no. 1 (February 2019): 1–23. <https://doi.org/10.1145/3230671>.

Gengnagel, Tessa, Frederike Neuber, Daniela Schulz, and Patrick Sahle. *FAIR Principles in Digital Scholarly Editions*. 2022. <https://ride.i-d-e.de/fair-criteria-editions/>.

'Geni.Com'. Accessed 2 December 2025. <https://www.geni.com/>.

- 'GeoNames'. Accessed 26 January 2026. <https://www.geonames.org/>.
- Gillman, Abigail. 'The Task of Jewish Translation Revisited'. In *The Routledge Handbook of Translation History*, 1st ed., by Christopher Rundle, 250–66. London: Routledge, 2021. <https://doi.org/10.4324/9781315640129-19>.
- Global Indigenous Data Alliance. 'CARE Principles'. Accessed 22 January 2026. <https://www.gida-global.org/care>.
- Hawkins, Ashleigh. 'Archives, Linked Data and the Digital Humanities: Increasing Access to Digitised and Born-Digital Archives via the Semantic Web'. *Archival Science* 22, no. 3 (September 2022): 319–44. <https://doi.org/10.1007/s10502-021-09381-0>.
- Hawkins, Kevin, and Michelle Dalmau. 'Best Practices for TEI in Libraries'. Text. November 2017. https://tei-c.org/extra/teiinlibraries/3.1.0a/main-driver.html#index.xml-body.1_div.3_div.5.
- HebrewNamer. 'HebrewNamer'. Accessed 26 January 2026. <https://hebrewnamer.com/>.
- Horel, Catherine. 'Midsize Cities in Austria-Hungary'. In *Multicultural Cities of the Habsburg Empire, 1880-1914*, 23–62. Imagined Communities and Conflictual Encounters. Central European University Press, 2023. <https://doi.org/10.7829/j.ctv2kzv0mg.8>.
- Horváth, Alíz, Cornelis Van Lit, Cosima Wagner, and David Joseph Wrisley. 'Towards Multilingually Enabled Digital Knowledge Infrastructures'. In *Multilingual Digital Humanities*, 1st ed., by Lorella Viola and Paul Spence, 197–212. London: Routledge, 2023. <https://doi.org/10.4324/9781003393696-17>.
- In Geveb. 'Latest Yiddish Translations, 2024'. Accessed 8 March 2026. <https://ingeveb.org/blog/latest-yiddish-translations-2024>.
- Jewish Encyclopedia. 'JewishEncyclopedia.Com'. Accessed 2 December 2025. <https://jewishencyclopedia.com/>.
- Koller, Werner, and Kjetil Berg Henjum. *Einführung in die Übersetzungswissenschaft*. 9. überarb. u. erw. Aufl. Stuttgart: Utb GmbH, 2020.
- Kotlerman, Boris. 'Kol Mevasser'. In *The YIVO Encyclopedia of Jews in Eastern Europe*. New York, NY: YIVO Institute for Jewish Research, 12 February 2010. <https://encyclopedia.yivo.org/article/1761>.

- Łapot, Mirosław. 'Development of the Education of Galician Jews at the Turn of the 19th and 20th Centuries as Exemplified by the City of Lviv'. *Scripta Judaica Cracoviensia* 18 (2021): 1–14. <https://doi.org/10.4467/20843925SJ.20.001.13869>.
- Luger, Robin. 'The Genuine Jew'. The Genuine Jew. Accessed 14 February 2026. <https://lug-robin.github.io/tgj-data/>.
- lug-robin. *Lug-Robin/Tgj-Data*. XSLT. 24 February 2024, released 11 February 2026. <https://github.com/lug-robin/tgj-data>.
- Manekin, Rachel. 'Die hebräische und jiddische Presse in Galizien'. In *Politische Öffentlichkeit und Zivilgesellschaft, Teilbd. 2, Die Presse als der Faktor der politischen Mobilisierung*, with Kommission für die Geschichte der österreichisch-ungarischen Monarchie 1848-1918, 8:2341–65. Die Habsburgermonarchie 1848-1918. Wien: Verlag der Österreichischen Akademie der Wissenschaften, 2006.
- Masel, Roni. 'Who Is a Yid? Reading the Journal *Der Yid* beyond the Hebraist – Yiddishist Binary'. *Journal of Modern Jewish Studies* 20, no. 3 (July 2021): 361–83. <https://doi.org/10.1080/14725886.2020.1809127>.
- Mendelsohn, Ezra. 'Jewish Assimilation in L'viv: The Case of Wilhelm Feldman'. In *Nationbuilding and the Politics of Nationalism: Essays on Austrian Galicia*, 3rd ed., 94–110. Monograph Series / Harvard Ukrainian Research Institute. Cambridge: Harvard Univ. Press, 1995.
- Nockels, Joe, Paul Gooding, Sarah Ames, and Melissa Terras. 'Understanding the Application of Handwritten Text Recognition Technology in Heritage Contexts: A Systematic Review of Transkribus in Published Research'. *Archival Science* 22, no. 3 (September 2022): 367–92. <https://doi.org/10.1007/s10502-022-09397-0>.
- Norich, Anita. *Writing in Tongues: Translating Yiddish in the Twentieth Century*. 1st ed. Samuel and Althea Stroum Lectures in Jewish Studies. Seattle: University of Washington Press, 2013.
- 'OpenSeadragon'. Accessed 21 December 2025. <https://openseadragon.github.io/>.
- 'Parlamentarierinnen und Parlamentarier ab 1848 | Parlament Österreich'. Accessed 26 January 2026. <https://www.parlament.gv.at/recherchieren/personen/parlamentarierinnen-ab-1848>.

- Perego, Simon. *Yiddish: Language, Culture and Memory from the Late 19th Century to the Present*. 15 April 2022. <https://doi.org/10.35008/TRACS-0123>.
- Pletschacher, Stefan, Christian Clausner, and Apostolos Antonacopoulos. 'Europeana Newspapers OCR Workflow Evaluation'. *Proceedings of the 3rd International Workshop on Historical Document Imaging and Processing*, 22 August 2015, 39–46. <https://doi.org/10.1145/2809544.2809554>.
- Portnoy, Edward. *Bad Rabbi: And Other Strange but True Stories from the Yiddish Press*. Stanford Studies in Jewish History and Culture. Stanford, California: Stanford university press, 2018.
- Reiner, Elchanan. 'The Yeshiva before 1800'. In *The YIVO Encyclopedia of Jews in Eastern Europe*. New York, NY: YIVO Institute for Jewish Research, 12 February 2010. <https://encyclopedia.yivo.org/article/47>.
- Reshef, Ronny, and Mirjam Gutschow. 'Text Recognition Model for Yiddish in Vaybertaytsh Typeface, Based on Community Regulations'. *Journal of Open Humanities Data* 10 (May 2024): 35. <https://doi.org/10.5334/johd.194>.
- Robinson, Abraham, and Osias H. Wilf. *Neues hebräisches Wörterbuch: verfasst von hervorragenden und besten Schriftstellern = Millon ibrihadas me-ibrit le-germanit Nosafot we-hasamtot*. Stanislau, Lemberg: E. Salat, 1912. <https://repozytorium.biblos.pk.edu.pl/resources/40492>.
- Rooker, Tim, Erica Smith, Javier Dominguez-Caballero, Lindsay Lee, and Sabino Ayvar-Soberanis. 'A Standardised Framework for Data Science in Advanced Manufacturing Systems'. *Procedia CIRP* 130 (2024): 1667–73. <https://doi.org/10.1016/j.procir.2024.10.298>.
- Sabharwal, Arjun. *Digital Curation in the Digital Humanities: Preserving and Promoting Archival and Special Collections*. Chandos Information Professional Series. Waltham, MA: Chandos Publishing, 2015.
- Sahle, Patrick. 'Digitale Edition'. In *Digital Humanities*, edited by Fotis Jannidis, Hubertus Kohle, and Malte Rehbein, 234–49. Stuttgart: J.B. Metzler, 2017. https://doi.org/10.1007/978-3-476-05446-3_17.
- . 'What Is a Scholarly Digital Edition?' In *Digital Scholarly Editing: Theories and Practices*, 19–40. Digital Humanities Series. Open Book Publishers, 2016.

- Schäfer, Lea. 'Yiddish'. In *Oxford Research Encyclopedia of Linguistics*, by Lea Schäfer. Oxford University Press, 2023.
<https://doi.org/10.1093/acrefore/9780199384655.013.946>.
- Shandler, Jeffrey. *Adventures in Yiddishland: Postvernacular Language & Culture*. Berkeley: University of California Press, 2006.
- Shanes, Joshua. *Diaspora Nationalism and Jewish Identity in Habsburg Galicia*. 1st ed. Cambridge University Press, 2012. <https://doi.org/10.1017/CBO9781139013642>.
- . 'Jewish Press and Periodicals'. In *The Routledge Handbook of Judaism in the 21st Century*, 451–57. Routledge Handbooks in Religion, Book 18. London; New York: Routledge, Taylor and Francis Group, 2023.
- Stäcker, Thomas. »A Digital Edition Is Not Visible« - Some Thoughts on the Nature and Persistence of Digital Editions. Version 1.0. 2020. Html,Xml,PDF.
https://doi.org/10.17175/2020_005.
- 'Startseite | Parlament Österreich'. Accessed 20 March 2026.
<https://www.parlament.gv.at/>.
- Stern, Heidi. 'Yiddish Studies: Language, Linguistics, and Literature'. *Year's Work in Modern Language Studies* 83, no. 1 (April 2023): 655–67.
- Text Encoding Initiative. '1 The TEI Infrastructure - The TEI Guidelines'. Accessed 28 November 2025. <https://tei-c.org/release/doc/tei-p5-doc/en/html/ST.html#STIN>.
- . '12 Representation of Primary Sources - The TEI Guidelines'. Accessed 28 November 2025. <https://tei-c.org/release/doc/tei-p5-doc/en/html/PH.html>.
- . '14 Names, Dates, People, and Places - The TEI Guidelines'. Accessed 28 November 2025. <https://tei-c.org/release/doc/tei-p5-doc/en/html/ND.html>.
- . 'TEI Guidelines'. Accessed 21 November 2025.
<https://www.tei-c.org/guidelines/>.
- 'The National Library Newspaper Collection'. Accessed 3 February 2026.
<https://www.nli.org.il/en/discover/newspapers>.
- The New York Public Library. 'Yiddish Women in Translation'. Accessed 8 March 2026.
<https://www.nypl.org/blog/2020/08/17/other-words-yiddish-women-translation-month>.

- 'The Streets of Lviv - List | Project "Lviv Streets"'. Accessed 26 January 2026.
<https://streets.lvivcenter.org/en/streets/>.
- 'The YIVO Encyclopedia of Jews in Eastern Europe'. Accessed 26 January 2026.
<https://encyclopedia.yivo.org/>.
- Toury, Jacob. *Die jüdische Presse im Österreichischen Kaiserreich: Ein Beitrag zur Problematik der Akkulturation 1802-1918*. Schriftenreihe wissenschaftlicher Abhandlungen des Leo Baeck Instituts 41. Tübingen: Mohr Siebeck, 1983.
- Urieli, Assaf, Amber Clooney, Michelle Sigiel, and Grisha Leyfer. 'Jochre 3 and the Yiddish OCR Corpus'. Version 1. Preprint, arXiv, 14 January 2025.
<https://doi.org/10.48550/ARXIV.2501.08442>.
- 'Verfassungen in Der Europäischen Union'. Accessed 20 March 2026.
<https://www.verfassungen.eu/>.
- 'Web Content Accessibility Guidelines (WCAG) 2.1'. With Andrew Kirkpatrick, Joshue O Connor, Alastair Campell, and Michael Cooper. Accessed 15 February 2026.
<https://www.w3.org/TR/WCAG21/>.
- Wikle, Olivia, and Evan Peter Williamson. 'Static Web Methodology as a Sustainable Approach to Digital Humanities Projects'. *The Code4Lib Journal*, no. 60 (April 2025). <https://journal.code4lib.org/articles/18372>.
- Wilf, Osias H. *Der Emes'r Yud, Organ für alle jüdischen Angelegenheiten*. L'viv, 1904. Reprint, Lemberg: Salat, 1904. <http://anno.onb.ac.at/cgi-content/anno?aid=emj>.
- . *ha- Boneh: ' iton yotse be-arba' ah peraḳim ba-shanah le-' inyene ha-dat, ha-yahadut, ha-ḥinukh ve-ha-ḥayim*. Lemberg: A. Goldman, 1904.
<http://digital.ub.uni-potsdam.de/503435>.
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 'The FAIR Guiding Principles for Scientific Data Management and Stewardship'. *Scientific Data* 3, no. 1 (March 2016): 9. <https://doi.org/10.1038/sdata.2016.18>.
- 'Yiddish in Translation | Yiddish Book Center'. Accessed 8 March 2026.
<https://www.yiddishbookcenter.org/discover/yiddish-translation>.
- 'Yiddish Transliteration and Spelling | YIVO Transliteration Chart for Yiddish Alphabet | Yiddish Spelling Guidelines'. Accessed 26 January 2026.
<https://www.yiddishwit.com/transliteration.html>.

- Zalkin, Mordechai. 'Heder'. In *The YIVO Encyclopedia of Jews in Eastern Europe*. New York, NY: YIVO Institute for Jewish Research, 12 February 2010.
<https://encyclopedia.yivo.org/article/31>.
- Zaritt, Saul. 'A Taytsh Manifesto: Yiddish, Translation, and the Making of Modern Jewish Culture'. *Jewish Social Studies* 26, no. 3 (2021): 186–222.
<https://doi.org/10.2979/jewisocistud.26.3.07>.
- Zhao, Fudie. 'A Systematic Review of Wikidata in Digital Humanities Projects'. *Digital Scholarship in the Humanities* 38, no. 2 (May 2023): 852–74.
<https://doi.org/10.1093/llc/fqac083>.

Table of Figures

Figure 1: Printed and available papers of "Der Emes'r Yud"	5
Figure 2: Articles with authors in "Der Emes'r Yud" sorted by topic.....	11
Figure 3: Cities in Eastern and Central Europe from where advertisements were sent to "Der Emes'r Yud"	13
Figure 4: "Japan" written without diacritics; Der Emes'r Yud (26.02.1904), 7.....	25
Figure 6: Map of Europe with the places mentioned in "Der Emes'r Yid"	30
Figure 5: Places mentioned in "Der Emes'r Yid"	29
Figure 7: Example of right-to-left text and left-to-right numbers combined: "19,229 Jews live here.", Der Emes'r Yud (12.02.1904), 6.....	34
Figure 8: Spacing for emphasis: "Loubet pardoned Dreyfus and acquitted him.": Pardoned is emphasised, Der Emes'r Yud (08.03.1904,) 2.....	34
Figure 9: Spacing to indicate a name: "The prime minister Dr Körber", Der Emes'r Yud (1.03.1904,) 3.....	34
Figure 10: Publication date on the first page in German and Hebrew, Der Emes'r Yud (12.02.1904), 1.....	58
Figure 11: The Hebrew date of publication, Der Emes'r Yud (08.03.1904), 1.....	59
Figure 12: The flow of data from XML to HTML.....	67
Figure 13: The micro editor.....	73
Figure 14: Highlighted people and places.....	73
Figure 15: Metadata variables on the website.....	77
Figure 16: The <i>project's</i> directory structure.....	78
Figure 17: The html folder directory.....	79
Figure 18: The data directory.....	83
Figure 19: Navigation Bar.....	86
Figure 20: Newspaper Index.....	87
Figure 21: Download option for the indices.....	87
Figure 22: Newspaper header with navigation arrows.....	88
Figure 23: Index of People.....	89
Figure 24: Search function demonstrated with the word "travel".....	90
Figure 25: Place register entry for Lviv.....	91

Index of Tables

Table 1: Subscription prices of Der Emes'r Yud..... 12

Table 2: Stylesheet templates..... 81

Table 3: Layout Stylesheets..... 82