

Preserving a Voice

A Digital Edition of The Genuine Jew - דער אָמת'ר יוד

Preserving a voice is a born-digital translation of Yiddish press annotated and contextualised as a scholarly digital edition. The edition and supporting data are displayed on a static, open-source website, situated in the semantic web. The focus lies in transforming scans of the newspaper in Five-Star Open-Source and Linked Data in a universally understood language without losing the essence of Yiddish press sources.

דער אָמת'ר יוד – The Genuine Jew (Salat, 1904) is a Yiddish newspaper that was published from February to October 1904 in Lviv, Ukraine – at that time part of the Austrian-Hungarian monarchy in Galicia. The paper features various articles, news and stories from different writers and the editor. A big focus is on Jewish life at the beginning of the 20th century Eastern Europe focusing on religious questions of Jewish people in the diaspora. Every issue also includes general, political, and Jewish news as well as historical retellings important for Jewish heritage and culture.

The scans, either in PDFs or images, are provided by the Austrian National Library (2023) already including an automated transcription in Hebrew letters in a separate text document. Taking those digitised sources, the text is transliterated, then translated manually as existing automated translation tools are not proficient (enough) to provide a satisfactory translation of multilingual sources featuring Yiddish – a language with numerous dialects and interchangeable letters – Hebrew, and German. The translation is then followed by the annotation in TEI / XML. Using the Cookiecutter library (2024) and XSLT stylesheets, the data is transformed and processed into HTML files which are then deployed on GitHub pages. On the website, the newspapers are published as text and viewable as HTML sources. People, places, Geopolitical Entities, important events, and Jewish concepts / traditions are linked into indices and to external databases like Wikidata or GeoNames.

Special focus is put on creating open, findable, accessible, interoperable, and reusable data by following the Five-Star Principle for Open Data (Berners-Lee, 2009) and the FAIR principles (Gengnagel, Neuber, and Schulz, 2022). The SDE will be made widely accessible by making use of catalogues and databases, enriching metadata, universally usable and machine-readable formats and linking all resources used and code written on the website. Thereby, publishing as Linked Data, it is easier to reuse, align and enrich the sources by accommodating multiple search methods (Hawkings, 2021). Additionally, it will follow the Web Content Accessibility Guidelines (2024) to improve universal access and digital accessibility as academia, research and digital resources widely lack digital accessibility.

The project creates Yiddish sources available to non-Yiddish speakers to ensure remembrance of the cultural heritage of Jewish life before World War II. Before the Shoa, there are estimated to have been between eleven and thirteen million Yiddish speakers. Pinpointing an exact number of native Yiddish speakers has proved difficult to find out, but there are between 350,000 and 600,000 estimated speakers, mostly in educational settings, in an effort to revive and spread Yiddish language usage again (Benor, 2023, pp. 98-99; Benor, 2024). Putting active Yiddish speakers who are also scholars working with Yiddish sources and said sources in relation to each other, several of these sources remain unaccounted for. While databases and projects collaboratively work on Jewish and Yiddish literature (Perego, 2022), there is no translation database found for Yiddish press available despite an abundance of Yiddish press sources available either on paper or digitised. Also, working only with Jewish press in the officially recognised language of the origin country – for

example, English, German, or Polish – is not sufficient to fully understand Jewish history as Yiddish press played a significant role in pre-World War II diasporic Jewish culture and the formation of a Jewish identity (Radosav, 2017, pp. 39-42). The English translation provides insight into Jewish cultural, religious, and social circumstances and traditions available for those not being able to read or understand Yiddish.

Combining all these standards, guidelines, and principles דער אַמתִּיר יוד is an effort to bring minority language primary sources to the wider research and non-research sphere alike.

References

- Benor, S. B. (2023). 'Hebrew and Jewish diaspora languages', in Fraiman, K. and Bell, D. P. (eds.) *The Routledge handbook of Judaism in the 21st Century*. London; New York: Routledge, pp. 89-110.
- Benor, S. B. (2024). *How many speakers are there? Statistics on Jewish Language Vitality*. Available at: <https://www.jewishlanguages.org/stats> (Accessed: 17 June 2024).
- Burners-Lee, T. (2009). *Linked Data*. Available at: <https://www.w3.org/DesignIssues/LinkedData.html> (Accessed: 17 June 2024).
- Cookiecutter (2024). *Cookiecutter Data Science*. Available at: <https://www.cookiecutter.io/templates> (Accessed: 17 June 2024).
- Costiuc Radosav, A. (2017). 'Tsi vos toyg a Yudish blat? Of what use is a Yiddish newspaper: Yiddish as a language of the press in nineteenth-century Romania', *Studia Judaica*, 22, pp. 21-49. <https://doi.org/10.24193/sj.2017.v22.2>.
- Gengnagel, T., Neuber, F., and Schulz, D. (2022). *Criteria for reviewing the application of FAIR principles in digital scholarly editions. Version 1.1*. Available at: <https://ride.i-d-e.de/fair-criteria-editions/> (Accessed: 17 June 2024).
- Hawkins, A. (2022). 'Archives, linked data and the Digital Humanities: Increasing access to digitised and born-digital archives via the semantic web', *Archival Science*, 22(3), pp. 319-344. <https://doi.org/10.1007/s10502-021-09381-0>.
- Österreichische Nationalbibliothek (2023). *Der emeth'r Jud, Organ für alle jüdischen Angelegenheiten*. Available at: <http://data.onb.ac.at/rec/AC09997825> (Accessed: 17 June 2024).
- Perego, S. (2022). *Yiddish: language, culture, and memory from the late 19th century to the present*. Available at: <https://transatlantic-cultures.org/en/catalog/culture-yiddish> (Accessed: 17 June 2024).
- Salat, O. W. (1904). *Der emeth'r Jud, Organ für alle jüdischen Angelegenheiten*. Available at: https://anno.onb.ac.at/info/emj_info.html (Accessed: 17 June 2024).
- World Wide Web Consortium (2024). *WCAG 2 Overview*. Available at: <https://www.w3.org/WAI/standards-guidelines/wcag/> (Accessed: 17 June 2024).