



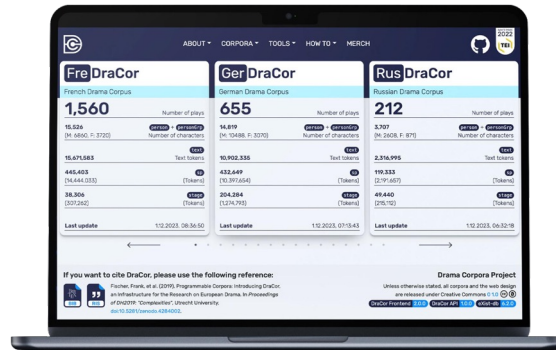
Open Science with DraCor

Making Literary Research Data Accessible and Reusable



Netzwerk Digitale
Geisteswissenschaften

What is DraCor?



DraCor is

- A **web-resource** available at dracor.org
- An open **collection** of (mostly) European **drama corpora**
- .. encoded in a research-oriented & open **machine-readable format** (TEI/XML, a long-standing standard in digital humanities)
- A researcher-friendly **data delivery infrastructure** (API, Python & R packages) which delivers:
 - full texts of the plays
 - semantically segmented texts of play parts (e.g. ‘all female speech from German plays’)
 - plays’ social network data
 - corpora and plays’ metadata, including LOD data
- A **collaboration** between University of Potsdam and Free University of Berlin, which was also temporarily funded as part of the EU project **CLS INFRA**
- A **community** of researchers, educators, developers, data contributors

Drama corpora: 3150 plays in 15 languages/dialects

The screenshot displays the DraCor website interface. At the top, there is a navigation bar with links for ABOUT, CORPORA, TOOLS, HOW TO, and MERCH. On the right, there are logos for GitHub and TEI, along with a 'BAMPTZ PRIZE 2022' badge. The main content area features three panels for different language corpora: Fre DraCor (French), Ger DraCor (German), and Rus DraCor (Russian). Each panel provides a summary of the corpus, including the number of plays, characters, text tokens, and stage tokens, along with the last update date.

| Language | Number of plays | Number of characters | Text tokens | Stage Tokens | Last update |
|----------------------|-----------------|----------------------------|-------------|---------------------|----------------------|
| Fre DraCor (French) | 1,560 | 15,526 (M: 6860, F: 3720) | 15,671,583 | 38,306 (307,262) | 01/12/2023, 08:36:50 |
| Ger DraCor (German) | 678 | 15,042 (M: 10617, F: 3137) | 11,056,425 | 208,470 (1,297,640) | 09/02/2024, 15:28:02 |
| Rus DraCor (Russian) | 212 | 3,707 (M: 2608, F: 871) | 2,316,995 | 49,440 (215,112) | 01/12/2023, 06:32:18 |

If you want to cite DraCor, please use the following reference:


Fischer, Frank, et al. (2019). Programmable Corpora: Introducing DraCor, an Infrastructure for the Research on European Drama. In *Proceedings of DH2019: "Complexities"*, Utrecht University, doi:10.5281/zenodo.4284002.

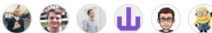
Drama Corpora Project
Unless otherwise stated, all corpora and the web design are released under Creative Commons 0 1.0 CC BY

[DraCor Frontend 2.1.0](#) [DraCor API 1.0.2](#) [eXist-db 6.2.0](#)

DraCor encoding (TEI/XML) — open on GitHub and via the API

main [gerdracor / tei / beer-struensee.xml](#) [Go to file](#) [...](#)

 [lehkost](#) add titlePage to some plays Latest commit bfadf6b on Jun 30 [History](#)

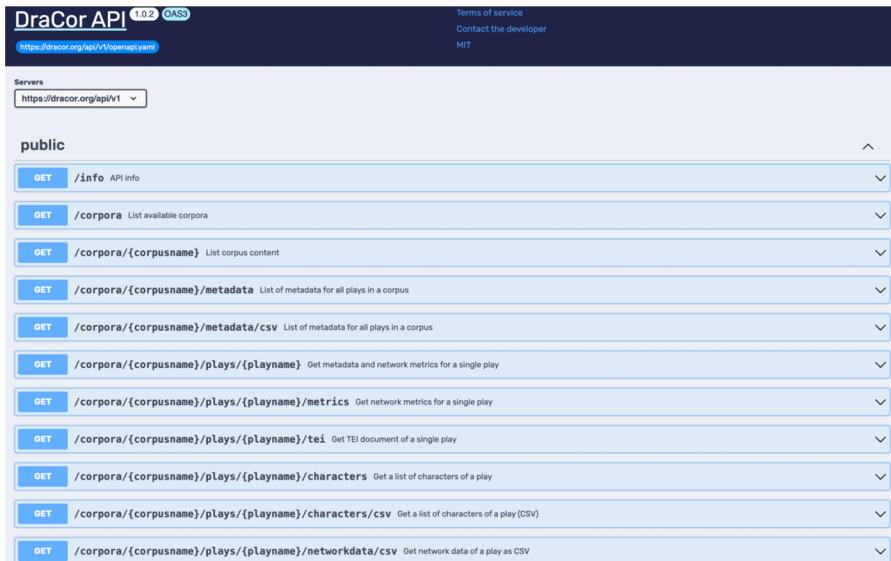
[6 contributors](#) 

9172 lines (9172 slocl) | 404 KB [Raw](#) [Blame](#) [Edit](#) [Delete](#)

```
1 <?xml version="1.0" encoding="utf-8"?>
2 <?xml-stylesheet type="text/css" href="../css/tei.css"?>
3 <?xml-model href="https://dracor.org/schema.rng" type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"?>
4 <TEI xmlns="http://www.tei-c.org/ns/1.0" xml:id="ger000090" xml:lang="ger">
5   <teiHeader>
6     <fileDesc>
7       <titleStmt>
8         <title type="main">Struensee</title>
9         <title type="sub">Trauerspiel in fünf Aufzügen</title>
10      <author>
11        <persName>
12          <forename>Michael</forename>
13          <surname>Beer</surname>
14        </persName>
15        <idno type="wikidata">Q107287</idno>
16        <idno type="pnd">118657887</idno>
17      </author>
18    </titleStmt>
19    <publicationStmt>
20      <publisher xml:id="dracor">DraCor</publisher>
21      <idno type="URL">https://dracor.org</idno>
```

DraCor research data delivery infrastructure

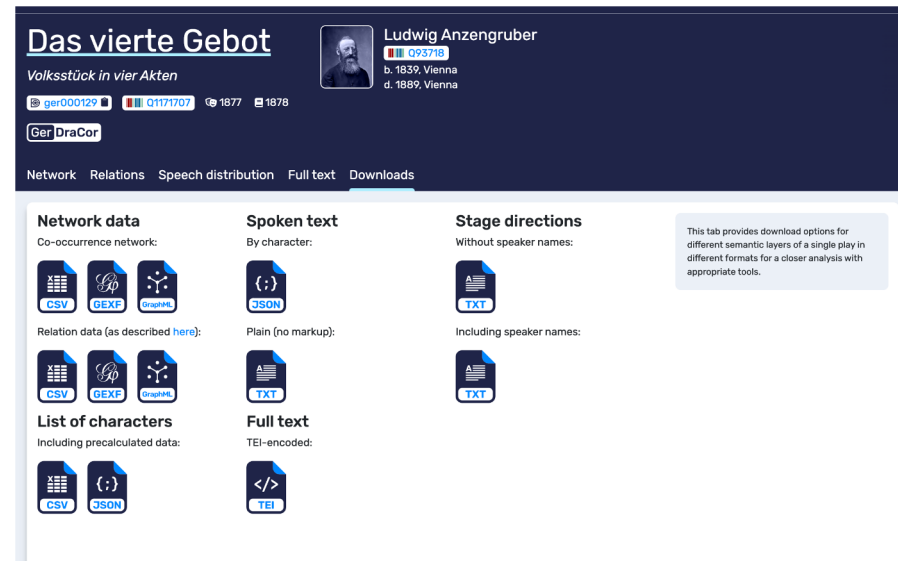
API:



The screenshot shows the DraCor API interface. At the top, it displays 'DraCor API 1.0.2 OAS3' and provides links for 'Terms of service', 'Contact the developer', and 'MIT'. Below this, there is a 'Servers' section with a dropdown menu set to 'https://draacor.org/api/v1'. The main content area is titled 'public' and lists various API endpoints with their methods and descriptions:

- GET /info** API info
- GET /corpora** List available corpora
- GET /corpora/{corpusname}** List corpus content
- GET /corpora/{corpusname}/metadata** List of metadata for all plays in a corpus
- GET /corpora/{corpusname}/metadata/csv** List of metadata for all plays in a corpus
- GET /corpora/{corpusname}/plays/{playname}** Get metadata and network metrics for a single play
- GET /corpora/{corpusname}/plays/{playname}/metrics** Get network metrics for a single play
- GET /corpora/{corpusname}/plays/{playname}/tei** Get TEI document of a single play
- GET /corpora/{corpusname}/plays/{playname}/characters** Get a list of characters of a play
- GET /corpora/{corpusname}/plays/{playname}/characters/csv** Get a list of characters of a play (CSV)
- GET /corpora/{corpusname}/plays/{playname}/networkdata/csv** Get network data of a play as CSV

Downloadable data files:



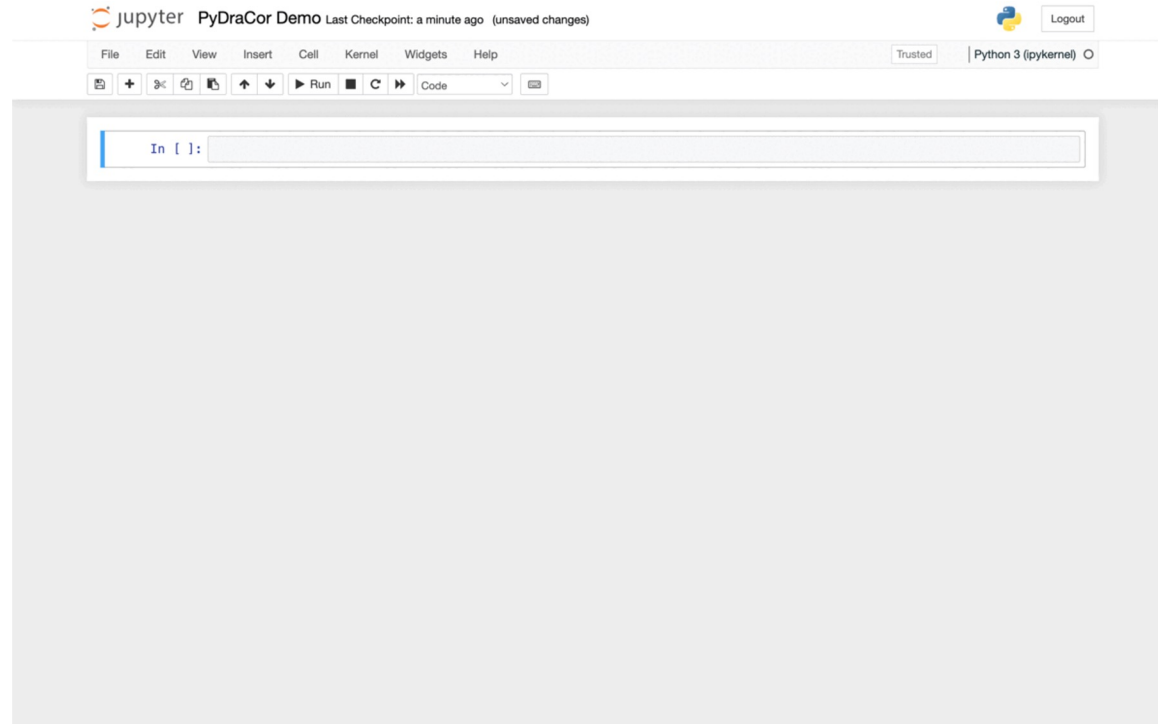
The screenshot shows the DraCor interface for the play 'Das vierte Gebot' by Ludwig Anzengruber. It includes a portrait of the author, his birth and death dates (b. 1839, Vienna; d. 1889, Vienna), and the play's title 'Volksstück in vier Akten'. The interface is in German and shows the play's identifier 'ger000129' and 'Q1171707'. Below this, there are tabs for 'Network', 'Relations', 'Speech distribution', 'Full text', and 'Downloads'. The 'Downloads' tab is active, showing three main categories of data:

- Network data**: Co-occurrence network. Options include CSV, GEXF, and GraphML.
- Spoken text**: By character. Options include JSON and Plain (no markup).
- Stage directions**: Without speaker names. Option includes TXT.

Additional options are provided for 'List of characters' (CSV, JSON) and 'Full text' (TEI). A note on the right states: 'This tab provides download options for different semantic layers of a single play in different formats for a closer analysis with appropriate tools.'

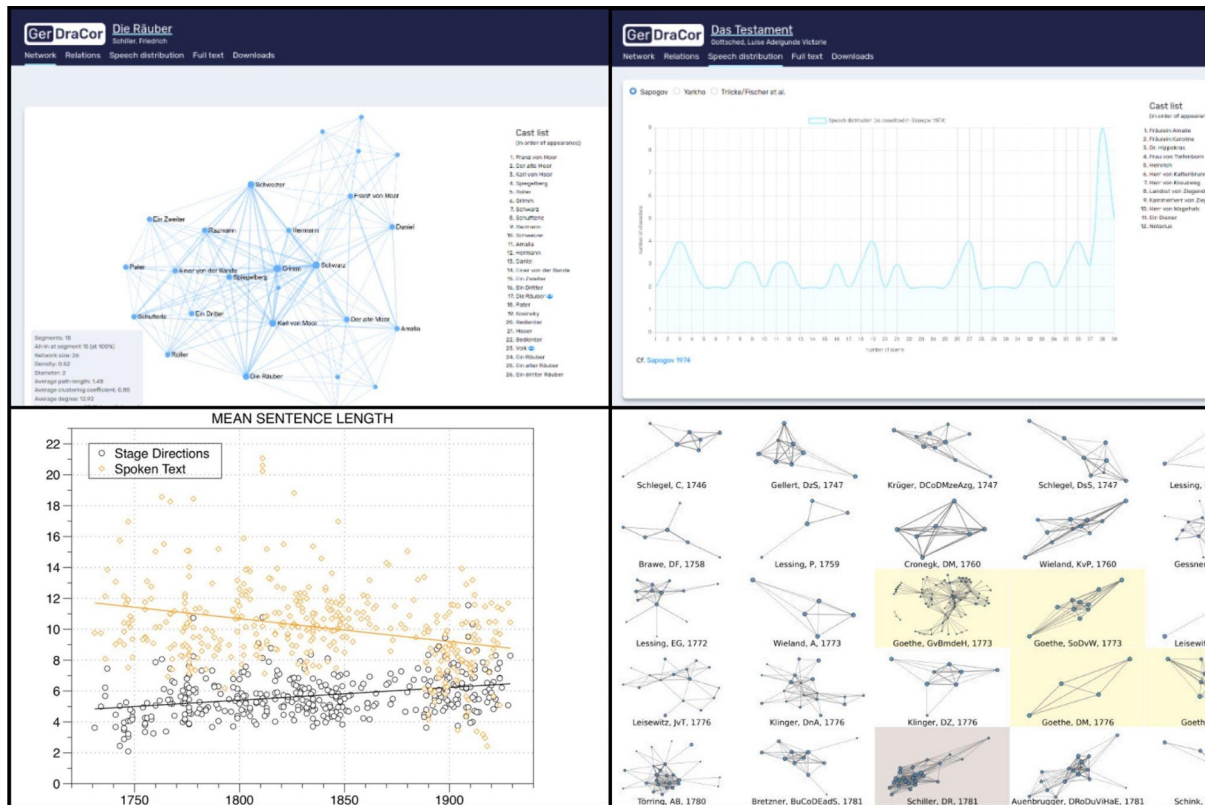
DraCor research data delivery infrastructure

Python package:



pypi.org/project/pydracor

DraCor-based modeling of drama



DraCor-based research



ABOUT ▾ CORPORA ▾ TOOLS ▾ HOW TO ▾ MERCH



Research

An (incomplete) collection of research articles on drama using data provided through DraCor (including papers authored by the [DLINA collective](#), the project preceding DraCor):






2023

- Petr Pořízka: CapekDraCor: A New Contribution to the European Programmable Drama Corpora. In: Jazykovedný časopis 74, 1 (2023), pp. 244–253. ([doi:10.2478/jazcas-2023-0042](https://doi.org/10.2478/jazcas-2023-0042))
- Katrin Dennerlein, Thomas Schmidt, Christian Wolff: EmoDrama. Ein Korpus mit Emotionsinformationen in Dramen von 1650–1815. In: Zeitschrift für digitale Geisteswissenschaften 8 (2023). ([doi:10.17175/2023_010](https://doi.org/10.17175/2023_010))
- Florian Cafiero, Simon Gabay: Dating the Stylistic Turn: The Strength of the Auctorial Signal in Early Modern French Plays. In: CHR2023. 6–8 December 2023. Paris. ([hal-04343340](https://hal.archives-ouvertes.fr/hal-04343340))
- Luca Giovannini, Daniil Skorinkin, Peer Trilcke, Ingo Börner, Frank Fischer, Julia Dudar, Carsten Milling and Petr Pořízka: Distributed Corpus Building in Literary Studies: The DraCor Example. In: DH2023: »Collaboration as opportunity«. 10–14 July 2023. Book of Abstracts. University of Graz. ([doi:10.5281/zenodo.8107457](https://doi.org/10.5281/zenodo.8107457))
- Janis Pagel, Benjamin Krautter, Melanie Andresen, Marcus Willand, Nils Reiter: On Designing Collaboration in a Mixed-Methods Scenario. Reflecting Quantitative Drama Analytics (QuaDrama). In: Birgit Schneider, Beate Löffler, Tino Mager, Carola Hein (eds.): Mixing Methods. Practical Insights from the Humanities in the Digital Age. Bielefeld University Press 2023, pp. 81–102. ([doi:10.14361/9783839469132](https://doi.org/10.14361/9783839469132))
- Luca Giovannini, Ingo Börner, Frank Fischer, Carsten Milling, Daniil Skorinkin, and Peer Trilcke (2023): Quali prospettive per ItaDraCor? Risorse e strumenti per la codifica di testi drammatici in lingua italiana. In: AIUCD2023: »La memoria digitale«. 5–7 June 2023. Book of Abstracts. University of Siena. ([doi:10.5281/zenodo.10009102](https://doi.org/10.5281/zenodo.10009102))
- Ingo Börner, Frank Fischer, Luca Giovannini, Christopher Lu, Carsten Milling, Daniil Skorinkin, Henny Sluyter-Gäthje, and Peer Trilcke: Onboard onto DraCor: Prototyping Workflows to Homogenize Drama Corpora for an Open Infrastructure. In: DHd2023: »Open Humanities, Open Culture«. 13–17 March 2023. Book of Abstracts. University of Luxembourg and University of Trier. ([doi:10.5281/zenodo.7715333](https://doi.org/10.5281/zenodo.7715333))
- Ingo Börner, Peer Trilcke et al.: On Programmable Corpora. Report and Prototype (DraCor). CLS INFRA D7.1. Zenodo, 28 February 2023. ([doi:10.5281/zenodo.7664964](https://doi.org/10.5281/zenodo.7664964))
- Peer Trilcke: Small Worlds, Beat Charts und die Netzwerkanalyse dramatischer Texte. Reflexionen aus dem Rabbit Hole. In: Fotis Jannidis (ed.): Digitale Literaturwissenschaft. Germanistische Symposien. Stuttgart: Metzler 2023, pp. 563–596. ([doi:10.1007/978-3-476-05886-7_23](https://doi.org/10.1007/978-3-476-05886-7_23))
- María Teresa Santa María Fernández, Monika Dabrowska: Análisis comparativo del Coro como personaje en tres tragedias griega y tres dramas españoles del Corpus DraCor. In: Neophilologus (2023). ([doi:10.1007/s11061-023-09768-7](https://doi.org/10.1007/s11061-023-09768-7))
- Katrin Dennerlein, Thomas Schmidt, Christian Wolff: Computational emotion classification for genre corpora of German tragedies and comedies from 17th to early 19th century. In: Digital Scholarship in the Humanities (2023). ([doi:10.1093/lc/fqad046](https://doi.org/10.1093/lc/fqad046))

DraCor community



Latest contributions from the community

- Ukrainian DraCor 
- Family relations markup in corpora 
- Early Modern English Drama Markup cleaning 
- Polish DraCor 
- Dutch DraCor 
- ... (more coming, now with help of YERUN)

The YERUN award will help to strengthen the LOD part



The screenshot shows the Wikidata page for 'The Blue Rose' (Q12082864), a play written by Lesya Ukrainka. The page displays various data fields and statements.

Language table:

| Language | Label | Description | Also known as |
|----------|------------------|--------------------------------|---------------------|
| English | The Blue Rose | play written by Lesya Ukrainka | |
| Russian | Голубая роза | драма / Drama | Українка / Ukrainka |
| German | No label defined | No description defined | |
| Latvian | No label defined | No description defined | |

Statements:

- instance of:** literary work (0 references, + add reference, + add value)
- form of creative work:** play (0 references, + add reference, + add value)
- author:** Lesya Ukrainka (0 references, + add reference, + add value)

Wikimedia projects:

- Wikipedia (1 entry) [edit]
- Wikibooks (0 entries) [edit]
- Wikinews (0 entries) [edit]
- Wikiquote (1 entry) [edit]
- Wikisource (0 entries) [edit]
- Wikiversity (0 entries) [edit]
- Wikivoyage (0 entries) [edit]
- Wiktionary (0 entries) [edit]

Thank you!

Daniil Skorinkin, Digital Humanities coordinator, University of Potsdam

skorinkin@uni-potsdam.de

twitter.com/DH_Potsdam

hcommons.social/@dh_potsdam