

# TextGrid Python Clients: Making the Repository Programmable

## Hynek, Stefan

stefan.hynek@uni-goettingen.de  
Niedersächsische Staats- und Universitätsbibliothek  
Göttingen, Deutschland  
ORCID: 0000-0002-3485-0746

## Veentjer, Ubbo

veentjer@sub.uni-goettingen.de  
Niedersächsische Staats- und Universitätsbibliothek  
Göttingen, Deutschland  
ORCID: 0000-0002-9726-3135

## Calvo Tello, José

calvotello@sub.uni-goettingen.de  
Niedersächsische Staats- und Universitätsbibliothek  
Göttingen, Deutschland  
ORCID: 0000-0002-1129-5604

## Barth, Florian

florian.barth@uni-goettingen.de  
Niedersächsische Staats- und Universitätsbibliothek  
Göttingen, Deutschland  
ORCID: 0000-0003-3408-7311

## Funk, Stefan

funk@sub.uni-goettingen.de  
Niedersächsische Staats- und Universitätsbibliothek  
Göttingen, Deutschland  
ORCID: 0000-0003-1259-2288

## Goebel, Mathias

goebel@sub.uni-goettingen.de  
Niedersächsische Staats- und Universitätsbibliothek  
Göttingen, Deutschland  
ORCID: 0000-0002-1102-5284

## Kurzawe, Daniel

kurzawe@sub.uni-goettingen.de  
Niedersächsische Staats- und Universitätsbibliothek  
Göttingen, Deutschland  
ORCID: 0000-0001-5027-7313

## Weimer, Lukas

weimer@sub.uni-goettingen.de  
Niedersächsische Staats- und Universitätsbibliothek  
Göttingen, Deutschland

ORCID: 0000-0001-6919-3646

## Current developments within Text-Grid

Started in 2006, the TextGrid infrastructure was developed jointly with humanities scholars and focuses on texts encoded as TEI. Since 2015, TextGrid has been operated by DARIAH-DE and is now part of the services offered by the Association for Research Infrastructures in the Humanities and Cultural Studies (GKFI) and the NFDI consortium Text+. As a partner institution, the Göttingen State and University Library contributes TextGrid's offerings to the latter, which ensures further demand-oriented development.<sup>1</sup> TextGrid offers interfaces via REST and SOAP and client libraries for Java and XQuery.

In this poster, we present an accessible and streamlined interface which is built on the programming language highly utilised in Digital Humanities: The library *TextGrid Python Clients* (in short: *tgclients*).

## Towards programmable interfaces

Some pioneer projects in the Digital Humanities (Digital Library, Théâtre classique, Biblioteca Italiana, etc.) dealt with the compilation of literary corpora. Researchers began to modify these corpora for their research through new features, tools (de la Rosa et al. 2022) and a new generation of corpora such as KOLIMO (Herrmann and Lauer 2017), ELTeC (Schöch et al. 2021), or DraCor (Börner and Trilcke 2023; Fischer et al. 2019).

The next step to this corpus recombination is the API-driven approach of programmable corpora in the DraCor platform (Börner and Trilcke 2023, 7), inspired by Aaron Swartz's concept of the Programmable Web, in which applications are "part of the ecology— a section of the programmable web" (Swartz 2022, 7). This idea of networked and distributed resources finds resonance in the NFDI consortia such as Text+ (Hinrichs et al. 2022), where the integration of existing resources is one of the main motivations.

## Modules of *tgclients*

Concerning the before-mentioned developments, the *tgclients* library is made to provide a future-proof interface with unified access to the APIs of TextGrid. The clients are capable of solving current needs and are highly scalable to adapt future feature requests. The TextGrid repository is orientated on the FRBR model (Functional Requirements for Bibliographic Records), a very widespread model for library records and publications, and stores data within TextGrid objects which consists of content and metadata (IFLA Study Group 2009). These objects can be or-

ganised as `aggregation` (`collection`, `edition`), `work`, or individual `item`.

**TG-crud:** The TG-crud service is responsible for creating, retrieving, updating, and deleting TextGrid resources, i.e. TextGrid objects including TextGrid metadata (TextGrid-Konsortium 2023b). *tgclients* provides the full functionality of TG-crud with a Python interface and therefore allows creating and editing individual TextGrid objects.

**TG-search:** TG-search is TextGrid's central search index combining semantic and technical metadata and can be used in conjunction with access conditions. In addition to fulltext search including filters and facets, the index holds specific information required to organise objects and their relations. *tgclients* uses this interface to query for all objects of a single project and allows for applying filters such as genre, language, author gender, or period of time. Users can combine these functions with the Aggregator.

**Aggregator:** The TextGrid Aggregator is the export and conversion tool for data from the TextGrid repository (TextGrid-Konsortium 2023a). The aggregator collects resources in one step and converts them into relevant output formats. *tgclients* allows users to access, convert and combine the content of single objects or complete TextGrid aggregations. For example, users can convert all TEI/XML of a nested TextGrid project into plain text.

## Use cases and perspectives

*tgclients* and its API is optimised for the usage in Jupyter Notebooks and provides an advanced user experience in documentation and autocompletion while working on a notebook. We further provide notebooks for educational purposes: Users can access several notebooks in our documentation that explain the usage of all modules with examples in TextGrid projects.<sup>2</sup>

The development team is curious to implement new use cases. Some ideas we have focus on the ex- and import functionality to down- and upload complete projects to TextGrid and to interact with the service for publishing data.

## Fußnoten

1. <https://de.dariah.eu> , GKFI: <https://forschungsinfrastrukturen.de> , <https://www.text-plus.org>.
2. <https://dariah-de.pages.gwdg.de/textgridrep/textgrid-python-clients/docs/getting-started.html#example-jupyter-notebooks>.

## Bibliographie

**Börner, Ingo, and Peer Trilcke.** 2023. "CLS INFRA D7.1 On Programmable Corpora", February. <https://zenodo.org/record/7664964> .

**Fischer, Frank, Ingo Börner, Mathias Göbel, Angelika Hechtl, Christopher Kittel, Carsten Milling, and Peer Trilcke.** 2019. "Programmable Corpora: Introducing DraCor, an Infrastructure for the Research on European Drama". In *Proceedings of DH2019: 'Complexities', Utrecht, July 9–12, 2019* . Utrecht University. <https://doi.org/10.5281/zenodo.4284002> .

**Herrmann, J. Berenike, and Gerhard Lauer.** 2017. "Das 'Was-Bisher-Geschah' von KOLIMO. Ein Update Zum Korpus Der Literarischen Moderne". In *Digitale Nachhaltigkeit* . Bern: ADHO. <https://dh-abstracts.library.cmu.edu/works/10644> .

**Hinrichs, Erhard, Peter Leinen, Alexander Geyken, Andreas Speer, and Regine Stein.** 2022. "Text+: Language- and Text-Based Research Data Infrastructure". Zenodo. <https://doi.org/10.5281/zenodo.6452002>. <https://zenodo.org/record/6452002> .

**IFLA Study Group on the Functional Requirements for Bibliographic Records.** 2009. "Functional Requirements for Bibliographic Records". Accessed July 1, 2023. <https://repository.ifla.org/bitstream/123456789/811/2/ifla-functional-requirements-for-bibliographic-records-frbr.pdf> .

**Rosa, Javier de la, Aitor Díaz, Álvaro Pérez, Salvador Ros, and Elena González-Blanco.** 2022. "Democratizing Poetry Corpora with Averell". In *Responding to Asian Diversity* . Tokyo: ADHO. <https://dh2022.dhii.asia/abstracts/414> .

**Schöch, Christof, Tomaz Erjavec, Roxana Patras, and Diana Santos.** 2021. "Creating the European Literary Text Collection (ELTeC): Challenges and Perspectives". *Modern Languages Open* , no. 1 (December): 25. <https://doi.org/10.3828/mlo.v0i0.364> .

**Swartz, Aaron.** 2022. *Aaron Swartz's The Programmable Web: An Unfinished Work* . Synthesis Lectures on Data, Semantics, and Knowledge Series. Cham: Springer International Publishing AG.

**TextGrid-Konsortium.** 2023a. "TextGrid Aggregator". Accessed July 1, 2023. <https://textgridlab.org/doc/services/submodules/aggregator/docs/index.html>

**TextGrid-Konsortium.** 2023b. "TG-crud". Accessed July 1, 2023. <https://textgridlab.org/doc/services/submodules/tg-crud/tgcrud-webapp/docs/index.html> .

**TextGrid-Konsortium.** 2023c. "TG-search". Accessed July 1, 2023. <https://textgridlab.org/doc/services/submodules/tg-search/docs/index.html> .