

<||/> **CLSIINFRA** COMPUTATIONAL  
LITERARY STUDIES  
INFRASTRUCTURE

# Dockerizing DraCor

**A Container-based Approach to  
Reproducibility in Computational Literary  
Studies**



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004984

**This presentation:**

[bit.ly/dockerizing-dracor](https://bit.ly/dockerizing-dracor)

The background features a large, faint, light blue graphic consisting of a left-pointing chevron on the left, a central vertical bar with a slash and angle brackets, and a right-pointing chevron on the right, mirroring the CLSIINFRA logo.

# Reproducibility as a Challenge for CLS

# A reproducibility problem

"The humanities have a 'reproducibility' problem"

[O'Sullivan 2019](#)



# A reproducibility problem

"The humanities have a 'reproducibility' problem"

[O'Sullivan 2019](#)

“ ...some serious and relevant challenges for the field of CLS: notably, the difficulty to reproduce work in this field, starting with issues of access to data and code, but also concerning questions of lacking reporting standards, limited scholarly recognition, and missing community commitment and capacity that would all be needed to foster a culture of [Repetitive Research] in [Computational Literary Studies] and beyond.”

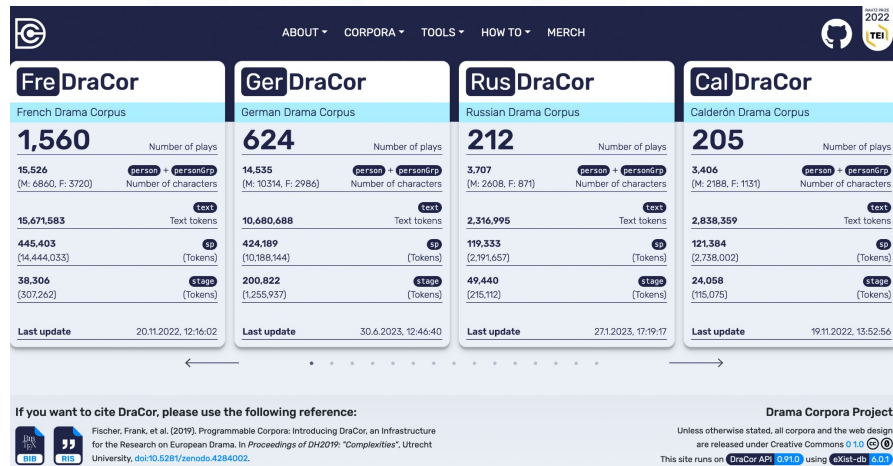
[Schöch \[Preprint\] 2023: 8](#)

- Community and research project [“CLS INRA. Computational Literary Infrastructure”](#), funded by European Union’s Horizon 2020 research and innovation programme
- Work Package 7: “Building the Ecosystem of and for Programmable Corpora”,
  - Cf. the report Börner, Trilcke et al. (2023): [“On Programmable Corpora”](#): “... corpora that expose an open, transparently documented and (at least partly) research-driven API to make texts machine-actionable.”

- Currently working on Task 7.3: “Technical stability of APIs (and Programmable Corpora) for versioning and reproducibility”
- Focus: replication (of research)
  - "... the term *replication* (of research) designates practices of [Repetitive Research] in which the research question, the dataset and the method of analysis of the repeating study are all identical to the original study." ([Schöch \[Preprint\] 2023](#): 14)
  - "The requirements for enabling replication (of research) are quite high, as the complete dataset and all code need to be available in a form that allows running the code without modifications or re-implementations." ([Schöch \[Preprint\] 2023](#): 15)

# Our Approach: Dockerizing DraCor

- Inspired by the concept of *research artifacts* from computer science (cf. e.g. [Arvan et al. 2022](#): 2351-2352)
- Based on [DraCor](#) (Drama Corpora Plattform)
- Self-set Challenge: Make the study [Trilcke et al. \[Preprint, 2022\] “Detecting Small Worlds in a Corpus of Thousands of Theater Plays”](#) replicable.



The screenshot displays the DraCor website interface, which provides detailed statistics for four drama corpora: French (Fre), German (Ger), Russian (Rus), and Calderón (Cal). Each corpus card lists the number of plays, characters, text tokens, and stage tokens, along with the last update date.

Corpus	Plays	Characters	Text Tokens	Stage Tokens	Last Update
Fre DraCor (French Drama Corpus)	1,560	15,526 (M: 6850, F: 3720)	15,671,583	38,306 (307,262)	20.11.2022, 12:16:02
Ger DraCor (German Drama Corpus)	624	14,535 (M: 10314, F: 2986)	10,680,688	200,822 (1,255,937)	30.6.2023, 12:46:40
Rus DraCor (Russian Drama Corpus)	212	3,707 (M: 2608, F: 871)	2,316,995	49,440 (215,112)	27.1.2023, 17:19:17
Cal DraCor (Calderón Drama Corpus)	205	3,406 (M: 2188, F: 1131)	2,838,359	24,058 (115,075)	19.11.2022, 13:52:56

At the bottom of the page, there is a citation reference and a Creative Commons license notice.

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 0.10 © CC  
This site runs on [DraCor API 0.91D](#) using [eXist-db 8.0.1](#)

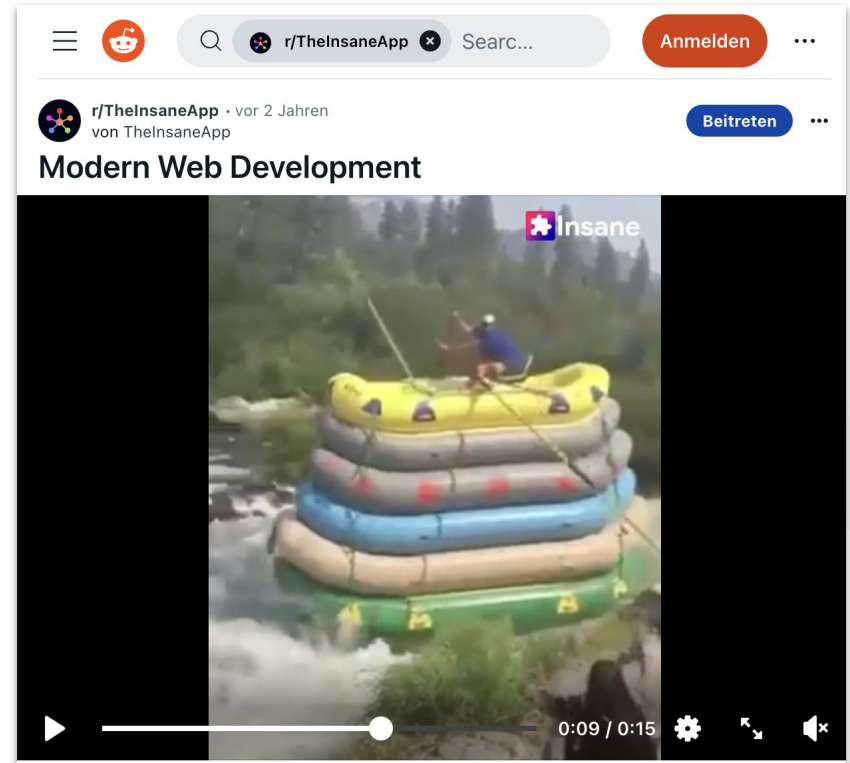


The background features several large, light gray, semi-transparent geometric shapes: a large left-pointing chevron on the left, a large right-pointing chevron on the right, and a central graphic consisting of three vertical bars of varying heights, resembling a bar chart or a stylized 'M' shape.

# Using Docker to Enhance Reproducibility

# Advantages of Docker in IT / “DevOps”

- streamline communication processes between Development (“Dev”) and Operations (“Ops”) Teams
- software dependencies are handled by developers – provide run-able apps



# Key Components of Docker



- **Dockerfile:** Text file (YAML) defines steps how to create an image.
- **Docker Image:** contains instructions for creating a container; built on top of other images
- **Docker Container:** running instance of an image
- **Docker compose** file: defines system consisting of multiple services run as containers in a single YAML file

## From *DevOps* to CLS Research



“You are free to use the code for your own purposes or to try to reproduce the results I put forth here. I make no claims to elegance in programming, but I am confident that **the scripts work, at least as of today**. Durability has taken on a new scale of meaning when seen against the long timescales of bibliographic preservation.” (Piper. Enumerations, xii)

The background features several large, light gray, semi-transparent geometric shapes: a large left-pointing chevron on the left, a large right-pointing chevron on the right, and a central graphic consisting of three vertical bars of varying heights and a diagonal bar, resembling a stylized 'W' or a data visualization element.

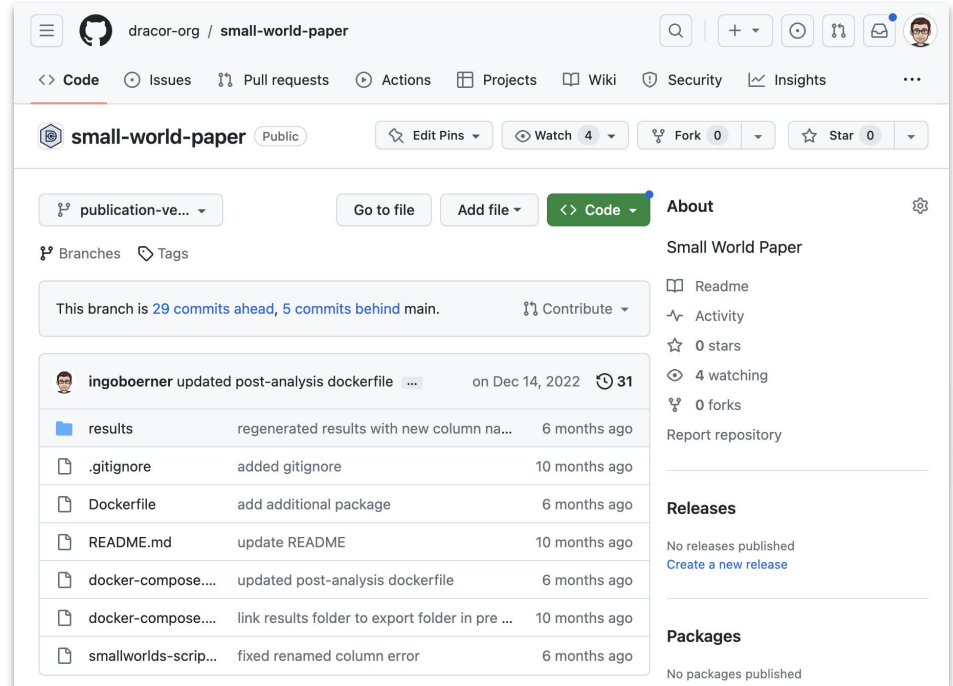
# **Use Case:** **“Small World” study in containers**

# Study on “Detecting Small Worlds in a Corpus of Thousands of Theater Plays”

Study: [Trilcke et al. \[Preprint, 2022\]](#)  
[“Detecting Small Worlds in a Corpus of Thousands of Theater Plays”](#)

- based on a Very Big Drama Corpus (“**VeBiDraCor**”) of approx. 3.000 plays,
- analyzed with an R script fetching data from a locally running DraCor API

stabilize not only the ”data” – the ‘living’ DraCor corpora – but all necessary infrastructural components: DraCor API, Research Environment (RStudio)

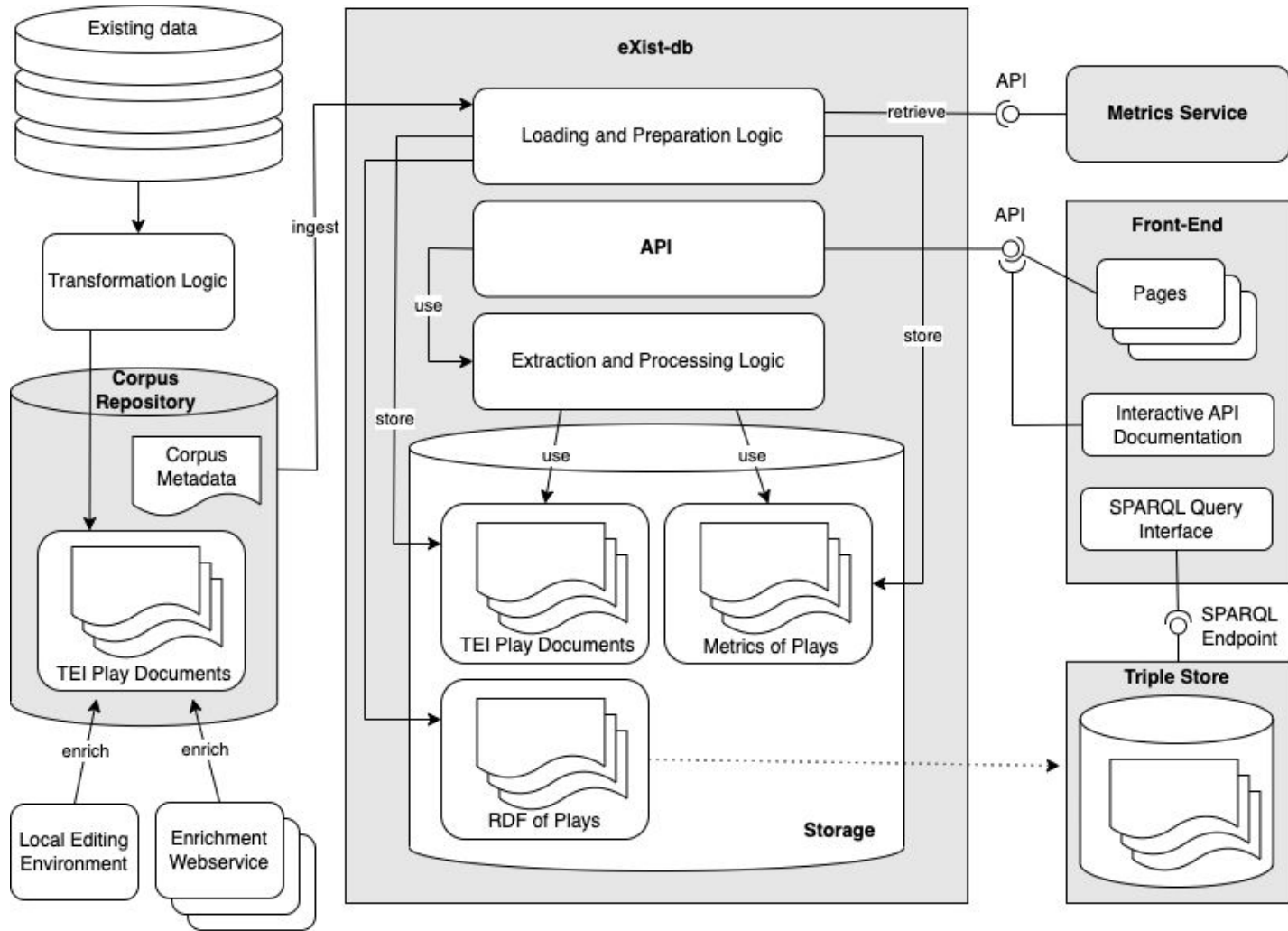


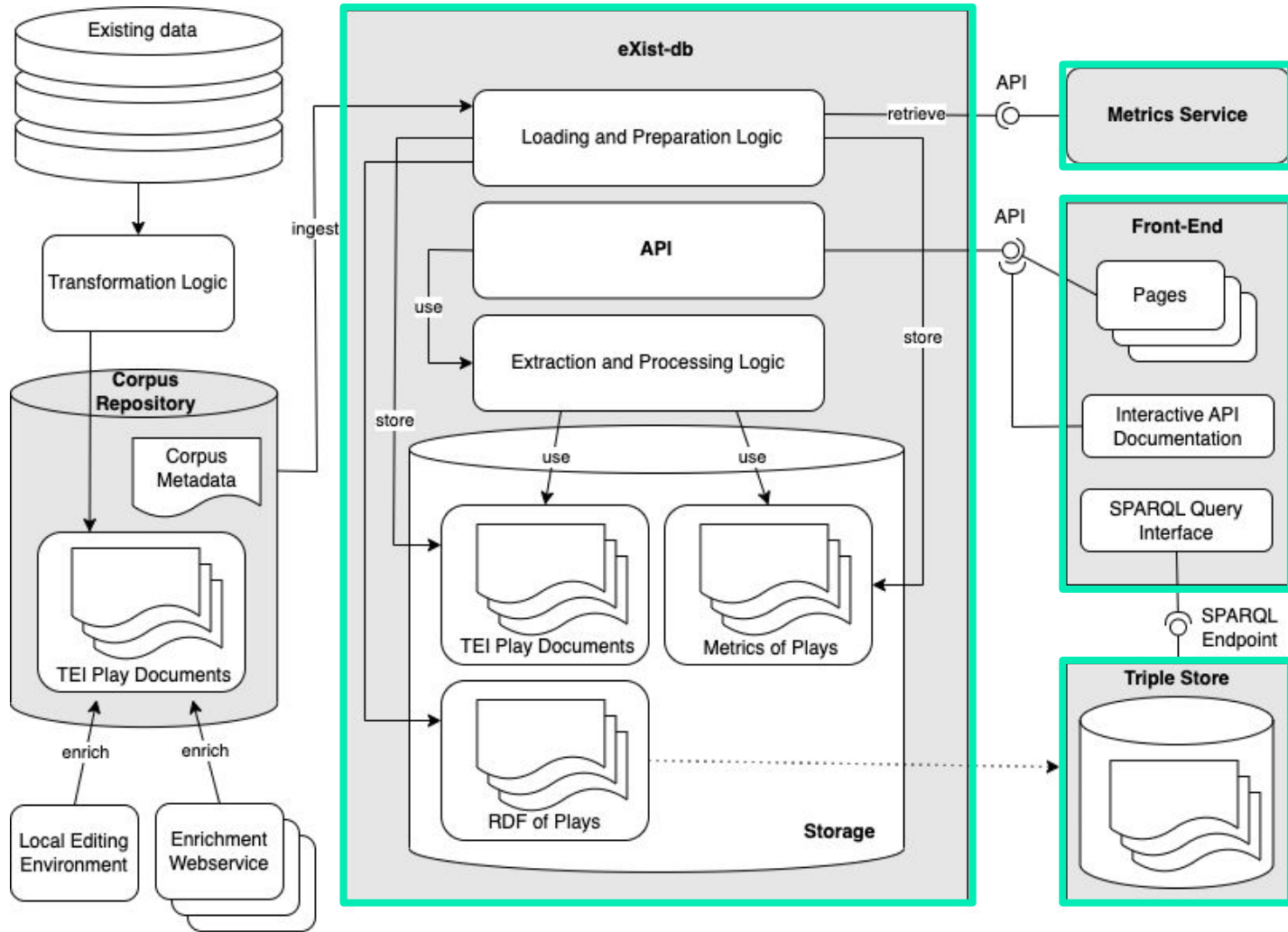
The screenshot shows the GitHub repository page for 'small-world-paper' by dracor-org. The repository is public and has 4 watchers, 0 forks, and 0 stars. The main content area shows a commit history table with the following entries:













Commit	Message	Time
ingoborner updated post-analysis dockerfile ...		on Dec 14, 2022 31
results	regenerated results with new column na...	6 months ago
.gitignore	added gitignore	10 months ago
Dockerfile	add additional package	6 months ago
README.md	update README	10 months ago
docker-compose....	updated post-analysis dockerfile	6 months ago
docker-compose....	link results folder to export folder in pre ...	10 months ago
smallworlds-scrip...	fixed renamed column error	6 months ago

The right sidebar contains sections for 'About' (Small World Paper), 'Releases' (No releases published), and 'Packages' (No packages published).

[GitHub Repository of the Study](#)





- 
 dracor/dracor-api ·  10 ·  0  
 By [dracor](#) · Updated 21 days ago  
[Image](#)
- 
 dracor/dracor-frontend ·  11 ·  0  
 By [dracor](#) · Updated 21 days ago  
[Image](#)
- 
 dracor/dracor-metrics ·  7 ·  0  
 By [dracor](#) · Updated 21 days ago  
[Image](#)
- 
 dracor/dracor-fuseki ·  7 ·  0  
 By [dracor](#) · Updated 21 days ago  
[Image](#)

Docker images of  
 DraCor Core  
 Services  
[\[DockerHub\]](#)



# VeBiDraCor – A Very Big Drama Corpus



dracor-org / vebidracor

Code Pull requests Actions Projects Wiki Security Insights Settings

vebidracor Public

forked from ingoerner/vebidracor

main 2 branches 0 tags

Go to file Add file Code

**Your main branch isn't protected**  
Protect this branch from force pushing or deletion, or require status checks before merging. [Learn more](#)

This branch is 10 commits ahead, 1 commit behind ingoerner:main. [Contribute](#) [Sync fork](#)

ingoerner updated workflow to create vebidracor v.3 3c3495d on Aug 9, 2022 15 commits

File	Commit	Time
.gitignore	added docker-compose to use pre-built images	last year
README.md	added compose file to reuse pre-built vebidracor	last year
docker-compose.empty.y...	updated workflow to create vebidracor v.3	last year
docker-compose.vebidra...	updated workflow to create vebidracor v.3	last year
vebidracor-workflow.ipynb	updated workflow to create vebidracor v.3	last year

README.md

## VeBiDraCor – Very Big Drama Corpus

For VeBiDraCor plays published on [DraCor](#) have been assembled to a Very Big Drama Corpus containing some 2.900 plays.

You can run a local instance of this corpus with docker `docker-compose -f docker-compose.vebidracor.yml up`. The database will have to reindex the file and starting the infrastructure will take some time, but ultimately, you can access it at <http://localhost:8088>

```
#list of corpora to include
corpora_to_include = [
  {
    "corpusname": "als",
    "repository": "https://github.com/dracor-org/alsdracor",
    "commit": "c87ea41aac9412e4bd84a28e9c7632c53904f77c"
  },
  {
    "corpusname": "bash",
    "repository": "https://github.com/dracor-org/bashdracor",
    "commit": "c16b58ef3726a63c431bb9575b682c165c9c0cbd"
  },
  {
    "corpusname": "cal",
    "repository": "https://github.com/dracor-org/caldracor",
    "commit": "6cb804d415051d5f18bc4841fa1ce4343a7f0ab5"
  },
  {
    "corpusname": "fre",
    "repository": "https://github.com/dracor-org/fredracor",
    "commit": "65e93f6ff632b367cdc7e16e3e390956856c4b98",
    "exclude": ["fre00038", "fre00057", "fre00065", "fre00099"]
  },
  {
    "corpusname": "ger",
    "repository": "https://github.com/dracor-org/gerdracor",
    "commit": "9135bd4598f54133f23df6edfc983b79f1616fb5",
    "exclude": ["ger000480"]
  },
  {
    "corpusname": "greek",
    "repository": "https://github.com/dracor-org/greekdracor",
    "commit": "7397aafa1927c3e0a0720bf3c00bf367ab679f26"
  },
]
```

<https://github.com/dracor-org/vebidracor>

# From Corpus-Container to Image: *docker commit*

```
In [30]: !docker container ls -a | grep vebidracor

13ada6df6ea8  ingoboerner/dracor-frontend:v1.4.3_local          "/docker-entrypoint..." 22 hou
rs ago      Up 22 hours                                0.0.0.0:8088->80/tcp        vebidracor_frontend_1
fb0395a7e6d3  ingoboerner/dracor-api:v0.86.3_local             "java org.exist.star..." 22 hou
rs ago      Exited (0) 2 minutes ago                  vebidracor_api_1
5eed285d8f57  ingoboerner/dracor-metrics:v1.2.1               "pipenv run hug -f m..." 22 hou
rs ago      Up 22 hours                                0.0.0.0:8030->8030/tcp     vebidracor_metrics_1
517897c784a1  stain/jena-fuseki                                "/docker-entrypoint..." 22 hou
rs ago      Up 22 hours                                0.0.0.0:3030->3030/tcp     vebidracor_fuseki_1
```

```
In [31]: !docker commit -m "prepared VeBiDraCor based on pre-built images, loaded corpora, added metrics" $vebidracor_api_c
sha256:24305e7b9509418e24b411cbef5e63702b25f95020b9ceb6d247d8cd1689dd45
```

```
In [32]: #!list
!docker images | grep vebidracor

ingoboerner/vebidracor-api          3.0.0          24305e7b9509  9 seconds ago
2.26GB
```

Pushed the images to Dockerhub `docker push ingoboerner/vebidracor-api:3.0.0`. Had to create the repos first on Dockerhub, e.g. <https://hub.docker.com/repository/docker/ingoboerner/vebidracor-api>

```
In [33]: !docker push ingoboerner/vebidracor-api:3.0.0

The push refers to repository [docker.io/ingoboerner/vebidracor-api]


bb6a64ae: Preparing
7f4d745d: Preparing
abele1b1: Preparing
d1cd4f1d: Preparing
5dle4382: Preparing
b2c72703: Preparing
```

# VeBiDraCor Docker Image



dockerhub  Explore Pricing Sign In [Register](#)

Explore > [ingoboerner/vebidracor-api](#)



## ingoboerner/vebidracor-api ☆

By [ingoboerner](#) · Updated a year ago  
Very Big Drama Corpus


Image

↓ Pulls 17

Overview **Tags**

Sort by  Filter Tags

TAG	DIGEST	OS/ARCH	SCANNED	COMPRESSED SIZE <sup>ⓘ</sup>
<a href="#">3.0.0</a> Last pushed a year ago by <a href="#">ingoboerner</a>	<a href="#">0b39f125534f</a>	linux/amd64	---	852.27 MB

[docker pull ingoboerner/vebidrac...](#) 

# Dockerizing Research Environment (RStudio Image)



small-world-paper / Dockerfile

ingoboerner add additional package

5622479 · 7 months ago History

Code Blame 14 lines (14 loc) · 545 Bytes

Raw Copy Download Edit History

```
1 #need to specify version of dockerized R
2 FROM rocker/verse:4
3 # set the API base
4 ENV DRACOR_APIBASE=http://localhost:8088/api/
5 # set environment variable corpusname
6 ENV CORPUSNAME=vebi
7 #install the packages; should also later use designated versions!
8 RUN R -e "install.packages(c('purrr', 'igraph', 'stringr', 'jsonlite', 'dplyr'))"
9 #add the script to use in R Studio server instance
10 ADD smallworlds-script.R /home/rstudio/smallworlds-script.R
11 #create an export folder for the results
12 RUN mkdir /home/rstudio/export
13 #expose the port
14 EXPOSE 8787
```



The Rocker Project  
Docker Containers for the R Environment

The screenshot shows the GitHub repository page for "ingoboerner/smallworld-rstudio". The repository is categorized as an "Image" and has 9 pulls. The repository was updated 7 months ago by the user "ingoboerner". The page includes a search bar, navigation links for "Explore", "Repositories", "Organizations", and "Help", and a blue header with the "dracor" logo.

# Docker Compose Files



```
small-world-paper / docker-compose.pre.yml
ingoboerner link results folder to export folder in pre analysis state container
Code Blame 27 lines (27 loc) · 572 Bytes
1 version: '2'
2 services:
3   api:
4     image: ingoboerner/vebidracor-api:3.0.0
5     ports:
6       - "8080:8080"
7   frontend:
8     image: ingoboerner/dracor-frontend:v1.4.3_local
9     ports:
10      - "8088:80"
11    depends_on:
12      - api
13  rstudio:
14    image: "dracor/smallworld-rstudio"
15    build:
16      context: .
17      dockerfile: Dockerfile
18    environment:
19      - PASSWORD=smallworld
20      - DRACOR_APIBASE=http://api:8080/exist/restxq/
21      - DRACOR_CORPUSNAME=vebi
22    volumes:
23      - ./results:/home/rstudio/export
24    ports:
25      - "8787:8787"
26    depends_on:
27      - api
```

ingoboerner updated post-analysis dockerfile	149862b on Dec 14, 2022	31 commits
results	regenerated results with new column names	7 months ago
.gitignore	added gitignore	last year
Dockerfile	add additional package	7 months ago
README.md	update README	last year
docker-compose.post.yml	updated post-analysis dockerfile	7 months ago
docker-compose.pre.yml	link results folder to export folder in pre analysis st...	last year
smallworlds-script.R	fixed renamed column error	7 months ago

Docker compose files allow to recreate

- pre-analysis state
- post-analysis state

Single command to start infrastructure:

```
docker compose -f
docker-compose.post.yml up
```

The background features several large, light gray, semi-transparent decorative elements: a large left-pointing chevron on the left, a bar chart with three bars of different heights in the center, and a large right-pointing chevron on the right.

# Outlook: StableDraCor Client

# StableDraCor Client

A screenshot of the GitHub repository page for 'dracor-org / stabledracor'. The page shows the repository name, navigation tabs (Code, Pull requests, Actions, Projects, Wiki, Security, Insights, Settings), and repository statistics (Watch: 0, Fork: 3, Star: 0). The main content area displays the 'master' branch, a file list, and a commit history table. The file list includes 'configurations', 'src/stabledracor', '.gitignore', 'LICENSE.txt', 'README.md', 'pyproject.toml', and 'requirements.txt'. The commit history table shows updates by 'ingoberner' over the last week and 4 days ago.

dracor-org / stabledracor

<> Code Pull requests Actions Projects Wiki Security Insights Settings

stabledracor Public  
forked from ingoberner/stable-dracor

master  
Go to file Add file <> Code About

Python package to simplify setting up a local DraCor system using Docker

Readme  
GPL-3.0 license  
Activity  
0 stars  
0 watching  
3 forks  
Report repository

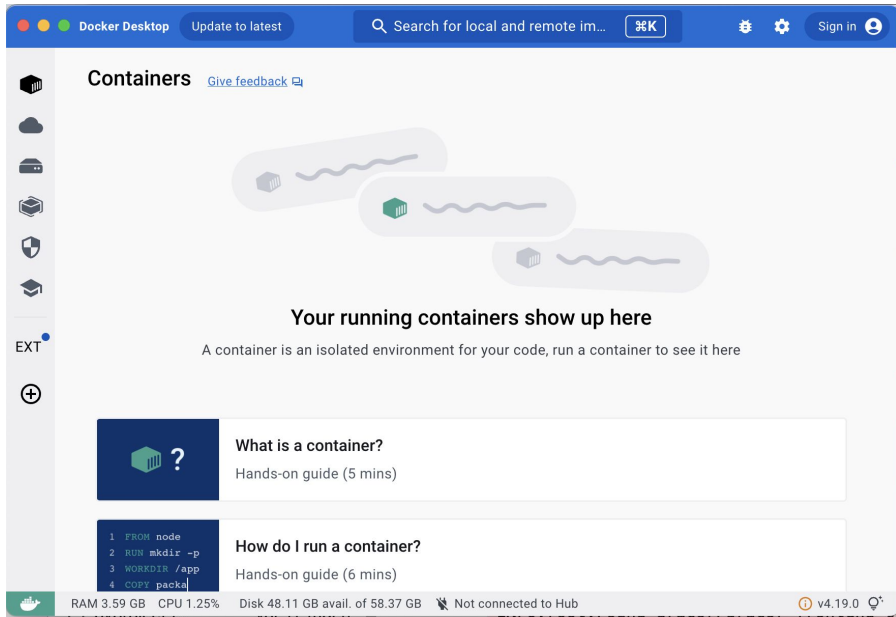
Releases  
No releases published  
Create a new release

This branch is up to date with ingoberner/stable-dracor:master. Contribute Sync fork

ingoberner updated dev notebook ... 4 days ago 54

configurations	moved fullstack compose file	last week
src/stabledracor	removed accidentally added file	4 days ago
.gitignore	updated gitignore	4 days ago
LICENSE.txt	added files for packaging	last week
README.md	updated readme	4 days ago
pyproject.toml	minor fix	4 days ago
requirements.txt	added pyyaml to requirements, method ...	last week

<https://github.com/dracor-org/stabledracor>



Official **Docker Desktop** app to manage images and containers [\[Download\]](#)

Initialize a StableDraCor instance  
(Demo in Jupyter notebook)

```
[7]: local_api = StableDraCor(name="my-stable-dracor",  
                             description="My custom DraCor")
```

```
INFO:root:Initialized new StableDraCor instance: 'my-stable-dracor' (ID: a93e0efb-0088-45fc-a641-bc16d6c50918).  
WARNING:root:Local DraCor API is not available at http://localhost:8088/api/.  
WARNING:root:Personal GitHub Access Token is not supplied. Requests to the GitHub API might be affected by rate limiting.  
INFO:root: Docker is available.  
WARNING:root:Could not detect a running Docker container derived from a dracor/dracor-api image.  
WARNING:root:Could not detect a running Docker container derived from a dracor/dracor-frontend image.  
WARNING:root:Could not detect a running Docker container derived from a dracor/dracor-metrics image.  
WARNING:root:Could not detect a running Docker container derived from a dracor/dracor-fuseki image.  
WARNING:root:Could not detect a running Docker container derived from a dracor/stable-dracor image.
```



# Running containers from within the client



```
[8]: local_api.run()
```

```
INFO:root:Fetched default compose file (configuration) from https://raw.githubusercontent.com/dracor-org/stabledracor/master/configurations/compose.fullstack.empty.yml.
```

```
Network my-stable-dracor_default Creating
Network my-stable-dracor_default Created
Container my-stable-dracor-fuseki-1 Creating
Container my-stable-dracor-metrics-1 Creating
Container my-stable-dracor-metrics-1 Created
Container my-stable-dracor-fuseki-1 Created
Container my-stable-dracor-api-1 Creating
Container my-stable-dracor-api-1 Created
Container my-stable-dracor-frontend-1 Creating
Container my-stable-dracor-frontend-1 Created
Container my-stable-dracor-metrics-1 Starting
Container my-stable-dracor-fuseki-1 Starting
Container my-stable-dracor-fuseki-1 Started
Container my-stable-dracor-metrics-1 Started
Container my-stable-dracor-api-1 Starting
Container my-stable-dracor-api-1 Started
Container my-stable-dracor-frontend-1 Starting
Container my-stable-dracor-frontend-1 Started
```

```
INFO:root:Started with downloaded docker compose file.
```

```
INFO:root:Found dracor/dracor-api container with ID 1931031ecfda. Image is: dracor/dracor-api:v0.90.1-local
```

```
INFO:root:Found dracor/dracor-frontend container with ID 0271fc83ddc3. Image is: dracor/dracor-frontend:v1.6.0-dirty
```

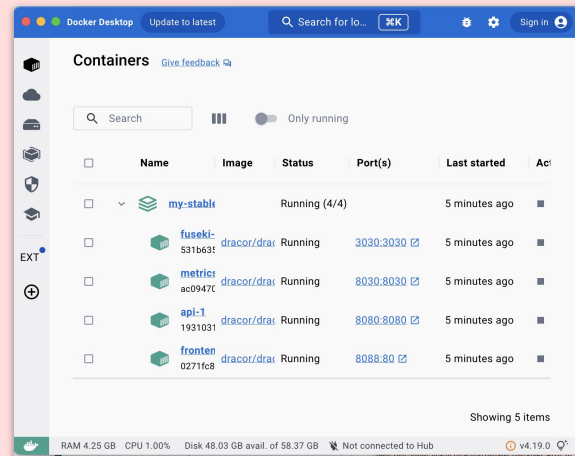
```
INFO:root:Found dracor/dracor-metrics container with ID ac09470773c9. Image is: dracor/dracor-metrics:v1.2.0
```

```
INFO:root:Found dracor/dracor-fuseki container with ID 531b6351a8b9. Image is: dracor/dracor-fuseki:v1.0.0
```

```
INFO:root:Trying to connect to the local DraCor API. This can take some time ...
```

```
INFO:root:DraCor API can be reached at 'http://localhost:8088/api/'.
```

```
[8]: True
```



use one single command `[run()]` to fetch images based on a [default Docker compose file](#) and start containers



localhost

ABOUT ▾ CORPORA ▾ TOOLS ▾ HOW TO ▾ MERCH



RAHITZ PRIZE 2022  
TEI





No corpora found

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  

This site runs on  DraCor API  0.90.1-2-g19a3f46-dirty using  eXist-db  6.0.1



DraCor frontend container and API are exposed at <http://localhost:8088>

no corpora are pre-loaded

# Copy Corpus from dracor.org



```
[13]: local_api.copy_corpus(source_corpusname="tat", exclude=["qamal-kaynish"])
```

```
INFO:root:Successfully created corpus tat. All metadata is available. Plays have not been added yet.  
INFO:root:Added contents of corpus tat from https://dracor.org/api/. 2 plays were added.  
INFO:root:Copying tat (as tat) was successful. Plays (that were not excluded) were also copied entirely.
```

```
[13]: True
```

Authors	Title	Year (normalized)	Network Size	Source	ID
Камал, Галиясгар Q2497099	Беренче театр Конедия 1 параваз Q25556355	1908 ✓ 1908	7	Татарская электронная библиотека TEI version	tat000001
Камал, Галиясгар Q2497099	Безнең шаһарнең серләре (Пьеса) Q78598039	1911 ✓ 1911	16	Татарская электронная библиотека TEI version	tat000003
Камал, Галиясгар Q2497099	Кайниш Пьеса 1 параваз 2 манзаравда (картинада) Q78063407	1912 ✓ 1912	7	Татарская электронная библиотека TEI version	tat000002



**Tat DraCor**

Tatar Drama Corpus

**2** Number of plays  
(M: 18, F: 4)

**9,396** Text tokens  
person + personGrp

**512** (Tokens)  
text

**310** (Tokens)  
sp

**stage**

Last update 12.7.2023, 10:38:40

ABOUT CORPORA TOOLS HOW TO MERCH

**Tatar Drama Corpus**

Edited by Daniil Skorinkin and Frank Fischer. Features a handful of plays in Tatar language, provided through Tatar Electronic Library.

Download a comprehensive table with metadata on all plays in the corpus: [JSON](#) [CSV](#)

Authors	Title	Year (normalized)	Network Size	Source	ID
Камал, Галиясгар Q2497099	Беренче театр Конедия 1 параваз Q25556355	1908 ✓ 1908	7	Татарская электронная библиотека TEI version	tat000001
Камал, Галиясгар Q2497099	Безнең шаһарнең серләре (Пьеса) Q78598039	1911 ✓ 1911	16	Татарская электронная библиотека TEI version	tat000003

# Create a corpus from a local folder

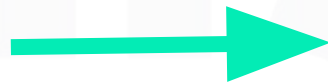


```
[15]: imported_corpus_metadata = {  
    "name" : "files",  
    "title": "My Custom Corpus",  
    "description": "This is a custom corpus that was generated with the 'StableDraCor' client. It contains files from a local directory."  
}  
local_api.add_plays_from_directory(corpusname="files",directory="test_import_dir", corpus_metadata=imported_corpus_metadata)
```

```
INFO:root:Successfully created corpus files. All metadata is available and plays are available.  
INFO:root:Added TEI data from file 'blok-neznakomka.xml' to corpus 'files'.  
INFO:root:Added TEI data from file 'munoz-refugio.xml' to corpus 'files'.  
INFO:root:Imported 2 files from test_import_dir as corpus 'files'.
```

[15]: True

Name	Änderungsdatum	Größe	Art
blok-neznakomka.xml	12.06.2023, 09:31	86 KB	XML Document
broken_file.txt	12.06.2023, 11:17	4 Byte	Reiner Text
munoz-refugio.xml	12.06.2023, 09:32	307 KB	XML Document



**Files DraCor**  
My Custom Corpus

**2** Number of plays

**51** (M: 38, F: 10) **person + personGrp** Number of characters

**27,954** **text** Text tokens

**1,552** **sp** (Tokens)

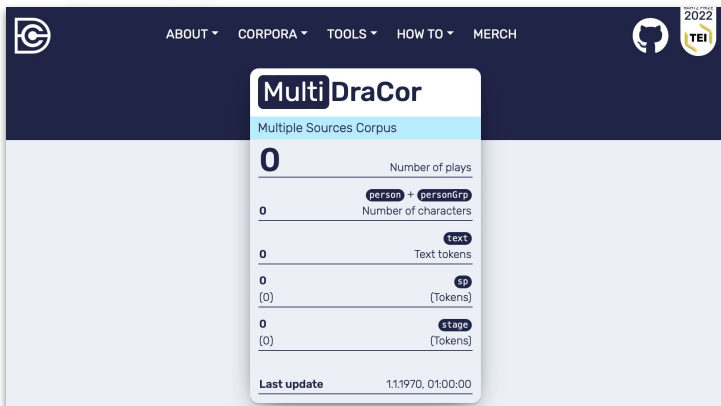
**795** **stage** (Tokens)

**Last update** 12.7.2023, 10:50:23

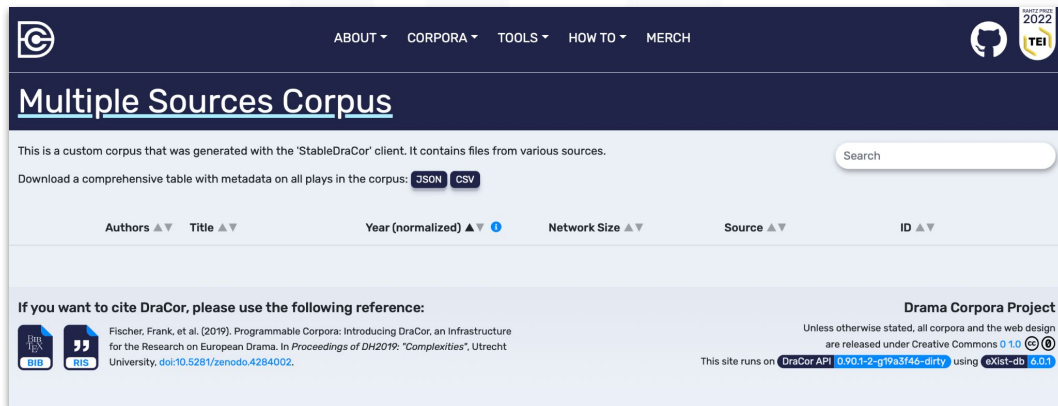
# Custom corpus (from several sources – like *VeBiDraCor*)

## Corpus from several sources

```
[9]: multi_source_corpus_metadata = {  
    "name" : "multi",  
    "title": "Multiple Sources Corpus",  
    "description": "This is a custom corpus that was generated with the 'StableDraCor' client. It contains files from various sources."  
}  
  
[10]: # need to explicitly register this corpus  
local_api.add_corpus(corpus_metadata=multi_source_corpus_metadata, register=True)  
  
INFO:root:Successfully created corpus multi. All metadata is available and plays are available.  
  
[10]: True
```



The screenshot shows the 'Multi DraCor' corpus page. It features a dark blue header with navigation links: ABOUT, CORPORA, TOOLS, HOW TO, and MERCH. A search bar is visible in the top right. The main content area displays the corpus name 'Multi DraCor' and a subtitle 'Multiple Sources Corpus'. Below this, there are several statistics: '0' plays, '0' characters (with a 'person + personGrp' tag), '0' text tokens (with a 'text' tag), '0' tokens (with a 'sp' tag), and '0' tokens (with a 'stage' tag). The 'Last update' is listed as '11.1970. 01:00:00'. A '2022 TEI' badge is present in the top right corner.



The screenshot shows the 'Multiple Sources Corpus' search results page. It features a dark blue header with navigation links: ABOUT, CORPORA, TOOLS, HOW TO, and MERCH. A search bar is visible in the top right. The main content area displays the corpus name 'Multiple Sources Corpus' and a subtitle 'Multiple Sources Corpus'. Below this, there is a description: 'This is a custom corpus that was generated with the 'StableDraCor' client. It contains files from various sources.' A search bar is visible in the top right. Below the description, there are two buttons: 'JSON' and 'CSV'. A table is displayed with columns: Authors, Title, Year (normalized), Network Size, Source, and ID. Below the table, there is a section for citing the corpus: '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.' A 'Drama Corpora Project' logo is visible in the bottom right corner. The footer contains the text: 'Unless otherwise stated, all corpora and the web design are released under Creative Commons 0.10 CC BY'. The site is powered by 'DraCor API 0.90.1-2-g19a3f46-dirty' using 'eXist-db 6.0.1'.

# Adding plays from a Github Repository



```
[11]: local_api.add_files_from_repo(corpusname="multi", repository_name="tatdracor")  
INFO:root:Successfully added all 3 files of repository 'dracor-org/tatdracor' to corpus 'multi'.  
[11]: True
```

Name	Last commit message	Last commit date
..		
qamal-berenche-teatr.xml	Move dates and wikidata ID to standOff	last year
qamal-beznen-shehernen-serlere.xml	Move dates and wikidata ID to standOff	last year
qamal-kaynish.xml	Move dates and wikidata ID to standOff	last year

Add the plays from [Tatar Drama Corpus](#) to the “multi” corpus in the local instance

**Multi DraCor**  
Multiple Sources Corpus

<b>3</b>	Number of plays
<b>30</b> (M: 20, F: 9)	<b>person + personGrp</b> Number of characters
<b>13,037</b>	<b>text</b> Text tokens
<b>701</b> (12,223)	<b>sp</b> (Tokens)
<b>433</b> (1,788)	<b>stage</b> (Tokens)
<b>Last update</b>	12.7.2023, 10:26:09

```
[14]: local_api.add_files_from_repo(corpusname="multi",
                                   repository_name="bashdracor",
                                   exclude=["khudayberdin-aq-bilettar", "karim-tashlama-utty"])
```

```
INFO:root:Successfully added all 1 files of repository 'dracor-org/bashdracor' to corpus 'multi'.
```

```
[14]: True
```

# Multi DraCor

## Multiple Sources Corpus

4

Number of plays

47

person + personGrp

(M: 31, F: 14)

Number of characters

text

25,851

Text tokens

1,557

sp

(24,323)

(Tokens)

722

stage

(2,960)

(Tokens)

Last update

12.7.2023, 10:47:01

## Multiple Sources Corpus

This is a custom corpus that was generated with the 'StableDraCor' client. It contains files from various sources.

Download a comprehensive table with metadata on all plays in the corpus: [JSON](#) [CSV](#)

Authors ▲▼	Title ▲▼	Year (normalized) ▲▼ ⓘ	Network Size ▲▼	Source ▲▼	ID ▲▼
Камал, Галиәсгар 	Беренче театр Комедия 1 пардада 	1908 ✍ 1908	7	Татарская электронная библиотека <a href="#">TEI version</a>	tat000001
Камал, Галиәсгар 	Безнең шәһәрнең серләре (Пьеса) 	1911 ✍ 1911	16	Татарская электронная библиотека <a href="#">TEI version</a>	tat000003
Камал, Галиәсгар 	Кайниш Пьеса 1 пардада 2 манзарада (картинада) 	1912 ✍ 1912	7	Татарская электронная библиотека <a href="#">TEI version</a>	tat000002
Кәрим, Мостай 	Ай тотолган төндә Өс шаршаулы трагедия 	1964 ✍ 1963 🗨 1964 📖 1964	17	kitaptar.bashkort.org <a href="#">TEI version</a>	bash00001

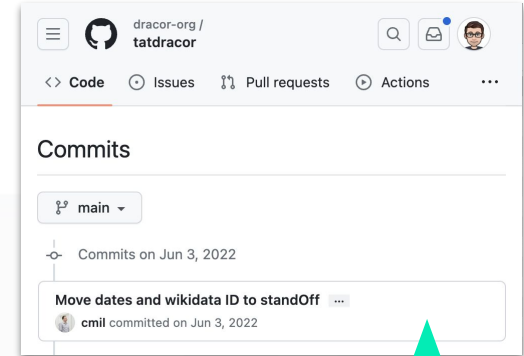
# Manifest documenting system and corpora loaded



```
[12]: local_api.get_manifest()
```

```
[12]: {'version': 'v1',  
      'system': {'id': 'a93e0efb-0088-45fc-a641-bc16d6c50918',  
                'name': 'my-stable-dracor',  
                'description': 'My custom DraCor',  
                'timestamp': '2023-07-12T10:31:23.334919'},  
      'services': {'api': {'container': '1931031ecfda',  
                           'image': 'dracor/dracor-api:v0.90.1-local',  
                           'version': '0.90.1-2-g19a3f46-dirty',  
                           'existdb': '6.0.1'},  
                  'frontend': {'co  
                               'image': 'draco  
                  'metrics': {'con  
                               'image': 'draco  
                  'triplestore': {  
                    'image': 'draco  
      'corpora': {'mult  
                  'timestamp': '2  
                  'sources': {'ta  
                    'commit': '5c  
                    'url': 'https  
                  'timestamp':  
                    'num_of_plays': 3}},  
                  'num_of_plays': 3}}}
```

```
'corpora': {'multi': {'corpusname': 'multi',  
                      'timestamp': '2023-07-12T10:21:56.346942',  
                      'sources': {'tatdracor': {'type': 'repository',  
                                                'commit': '5c71364f39f6533baa3a2e04217fd39e0898c851',  
                                                'url': 'https://github.com/dracor-org/tatdracor',  
                                                'timestamp': '2023-07-12T10:26:04.848341',  
                                                'num_of_plays': 3}},
```





# Create image



```
[16]: local_api.create_docker_image_of_service(service="api", image_tag="my_custom_dracor")
```

```
WARNING:root:The dracor-api container is running. There might be issues with the image, if it is create f
eating the image.
INFO:root:Committed container 1931031ecfda as dracor/stable-dracor:my_custom_dracor. Image identifier sha
7893a3ae15f0d
▪
```

Name	Tag	Status	Created	Size	Actions
<a href="#">dracor/stable-dracor</a> 2e2438012ab7	my_custom_dracor	Unused	less than a mi	456.79 MB	▶ ⋮ 🗑️
<a href="#">dracor/dracor-api</a> 171df59ae0ab	v0.90.1-local	In use	about 1 month	367.3 MB	▶ ⋮ 🗑️
<a href="#">dracor/dracor-metrics</a> 1806ba4d7047	v1.2.0	In use	about 1 month	943.62 MB	▶ ⋮ 🗑️

```
[17]: local_api.list_docker_images()
```

```
[17]: [{'Containers': 'N/A',  
      'CreatedAt': '2023-07-12 10:55:54 +0200 CEST',  
      'CreatedSince': '4 minutes ago',  
      'Digest': '<none>',  
      'ID': '2e2438012ab7',  
      'Repository': 'dracor/stable-dracor',  
      'SharedSize': 'N/A',  
      'Size': '457MB',
```

Extract Docker Object Labels from  
Docker Image

Manifest data converted to labels in  
*org.dracor.stable-dracor.\** namespace

```
[18]: local_api.get_labels_from_docker_image("2e2438012ab7")
```

```
[18]: {'com.docker.compose.config-hash': 'f3e7dbfe4be9df9f284e9ee661814e9b4623d214ea7d49854b4a651aa8e98c7d',  
      'com.docker.compose.container-number': '1',  
      'com.docker.compose.depends_on': 'fuseki:service_started:false,metrics:service_started:false',  
      'com.docker.compose.image': 'sha256:171df59ae0ab650356c45feeabe2a65b63b77c3e9bd7bf362926bfd78e931f8',  
      'com.docker.compose.oneoff': 'False',  
      'com.docker.compose.project': 'my-stable-dracor',  
      'com.docker.compose.project.config_files': '-',  
      'com.docker.compose.project.working_dir': '/Users/ingoboerner/Projekte/dracor/stable-dracor',  
      'com.docker.compose.service': 'api',  
      'com.docker.compose.version': '2.17.3',  
      'org.dracor.stable-dracor.corpora': 'multi,tat,files',  
      'org.dracor.stable-dracor.corpora.files.corpusname': 'files',  
      'org.dracor.stable-dracor.corpora.files.num-of-plays': '2',  
      'org.dracor.stable-dracor.corpora.files.sources': '9edc7a49',  
      'org.dracor.stable-dracor.corpora.files.sources.9edc7a49.timestamp': '2023-07-12T10:50:20.352739',  
      'org.dracor.stable-dracor.corpora.files.sources.9edc7a49.type': 'files',  
      'org.dracor.stable-dracor.corpora.files.timestamp': '2023-07-12T10:50:20.352730',
```

# Manifest <> Image Labels



```
'tat': {'corpusname': 'tat',  
  'timestamp': '2023-07-12T10:38:37.104030',  
  'sources': {'tat': {'type': 'api',  
    'corpusname': 'tat',  
    'url': 'https://dracor.org/api/corpora/tat',  
    'timestamp': '2023-07-12T10:38:37.104047',  
    'exclude': {'type': 'slug', 'ids': ['qamal-kaynish']},  
    'num_of_plays': 2}},  
  'num_of_plays': 2},
```

```
'org.dracor.stable-dracor.corpora.tat.corpusname': 'tat',  
'org.dracor.stable-dracor.corpora.tat.num-of-plays': '2',  
'org.dracor.stable-dracor.corpora.tat.sources': 'tat',  
'org.dracor.stable-dracor.corpora.tat.sources.tat.corpusname': 'tat',  
'org.dracor.stable-dracor.corpora.tat.sources.tat.exclude.ids': 'qamal-kaynish',  
'org.dracor.stable-dracor.corpora.tat.sources.tat.exclude.type': 'slug',  
'org.dracor.stable-dracor.corpora.tat.sources.tat.num-of-plays': '2',  
'org.dracor.stable-dracor.corpora.tat.sources.tat.timestamp': '2023-07-12T10:38:37.104047',  
'org.dracor.stable-dracor.corpora.tat.sources.tat.type': 'api',  
'org.dracor.stable-dracor.corpora.tat.sources.tat.url': 'https://dracor.org/api/corpora/tat',  
'org.dracor.stable-dracor.corpora.tat.timestamp': '2023-07-12T10:38:37.104030',
```

# StableDraCor Client



A screenshot of the GitHub repository page for 'dracor-org / stabledracor'. The page shows the repository name, navigation tabs (Code, Pull requests, Actions, Projects, Wiki, Security, Insights, Settings), and repository statistics (Watch: 0, Fork: 3, Star: 0). The main content area displays the 'master' branch, a file list, and a commit history table. The file list includes 'configurations', 'src/stabledracor', '.gitignore', 'LICENSE.txt', 'README.md', 'pyproject.toml', and 'requirements.txt'. The commit history table shows updates by 'ingoberner' over the last week and 4 days ago. The right sidebar contains an 'About' section describing the package as a Python tool for setting up a local DraCor system using Docker, along with links to the README, license, and activity.

<https://github.com/dracor-org/stabledracor>

A large, faint, light blue stylized code symbol '<||/>' is centered in the background of the slide.

THANK  
YOU

[ingo.boerner@uni-potsdam.de](mailto:ingo.boerner@uni-potsdam.de)  
[trilcke@uni-potsdam.de](mailto:trilcke@uni-potsdam.de)