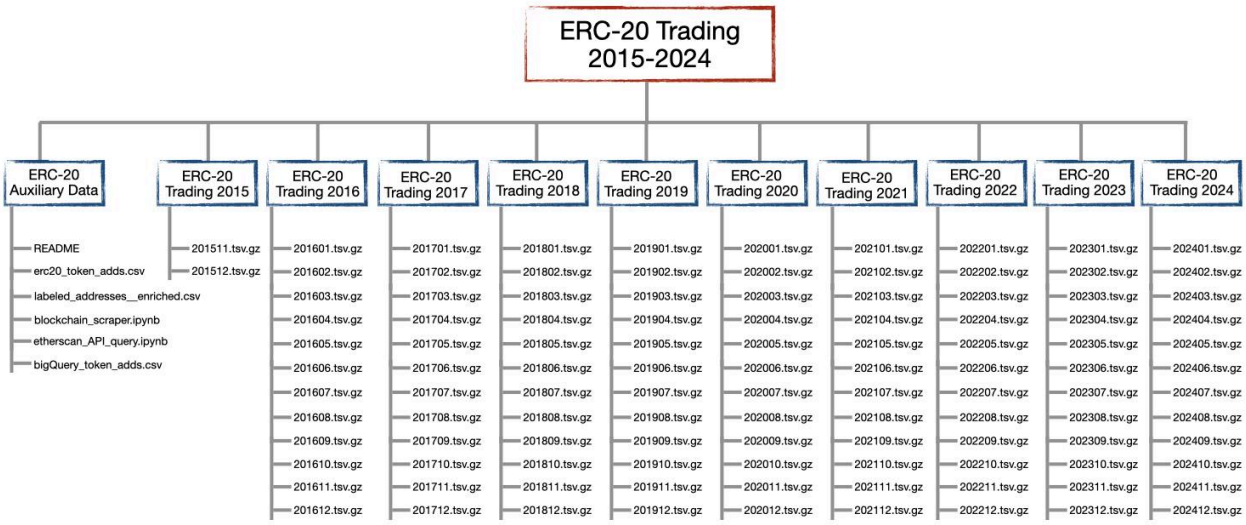
# ERC-20 crypto-tokens trading data over the Ethereum Blockchain

---

## Overview

This dataset contains a comprehensive, longitudinal record of ERC-20 token transactions extracted from the Ethereum blockchain over a 9-year period, between 2015 and 2024. The dataset is intended to support research in network science, financial behavior, DeFi ecosystems, and blockchain analytics.
The following diagram describes the structure of the dataset. The origin is the “ERC-20 Trading 2015-2024” dataverse, located in https://dataverse.harvard.edu/dataverse/erc20Trading.
Each of the blue rectangles stand for different datasets within this tataverse. The first one holding all auxiliary data, including this README file, the list of all ERC-20 crypto tokens, scraped from etherescan.io, a list of exemplary labeled wallet addresses, pertaining to service providers such s crypto-exchanges, and a jupyter notebook containing the scraping code we have utilized for extracting this data, in order to facilitate further, more recent, data extraction.
The rest of the datasets hold ERC-20 trading data for each year, divided by months.




Each month is zipped, and contains daily files, each corresponding to one calendar day of ERC-20 token activity and is named using the format: YYYYMMDD.tsv.gz.

## File Contents

Each TSV file contains the following columns:

1. **Token Address**: indicating the address of the Smart Contract governing the crypto-token.

2. **Token Name**: String identifying the crypto-token.
3. **Token Symbol**: The abbreviated shorthand of the crypto-token, 3 to 4 characters long.
4. **Token Decimals**: The number of decimal places a token supports, determining its smallest transferable unit.
5. **Value**: denoting the transferred token amount.
6. **Sender**: wallet address representing the seller.
7. **Recipient**: wallet address representing the buyer.
8. **Time**: the time at which the transaction took place.
9. **Transaction Hash**: a unique transaction identifier.
10. **Block-id**: the unique integer identifier of a block in the Ethereum blockchain.

## Data Access

Apart from the access provided by the website, Harvard Dataset also supports bulk downloading data using bash code. Relevant code and instructions can be found in the following github repository:
https://github.com/gdcc/dataverse-recipes/tree/main/shell/download

## Citation

If you find this dataset helpful, please cite this dataset:

## Contact

For questions or suggestions, please contact:
Dr. Shahar Somin,
Bar-Ilan University, Faculty of Engineering
shahar.gavrielov-somin@biu.ac.il