People's Data Stewardship Hackathon

Equitech Futures

🌡️ **Track 1: Climate Data Vault**



The QuackTorrentinos

Bernardo Gameiro

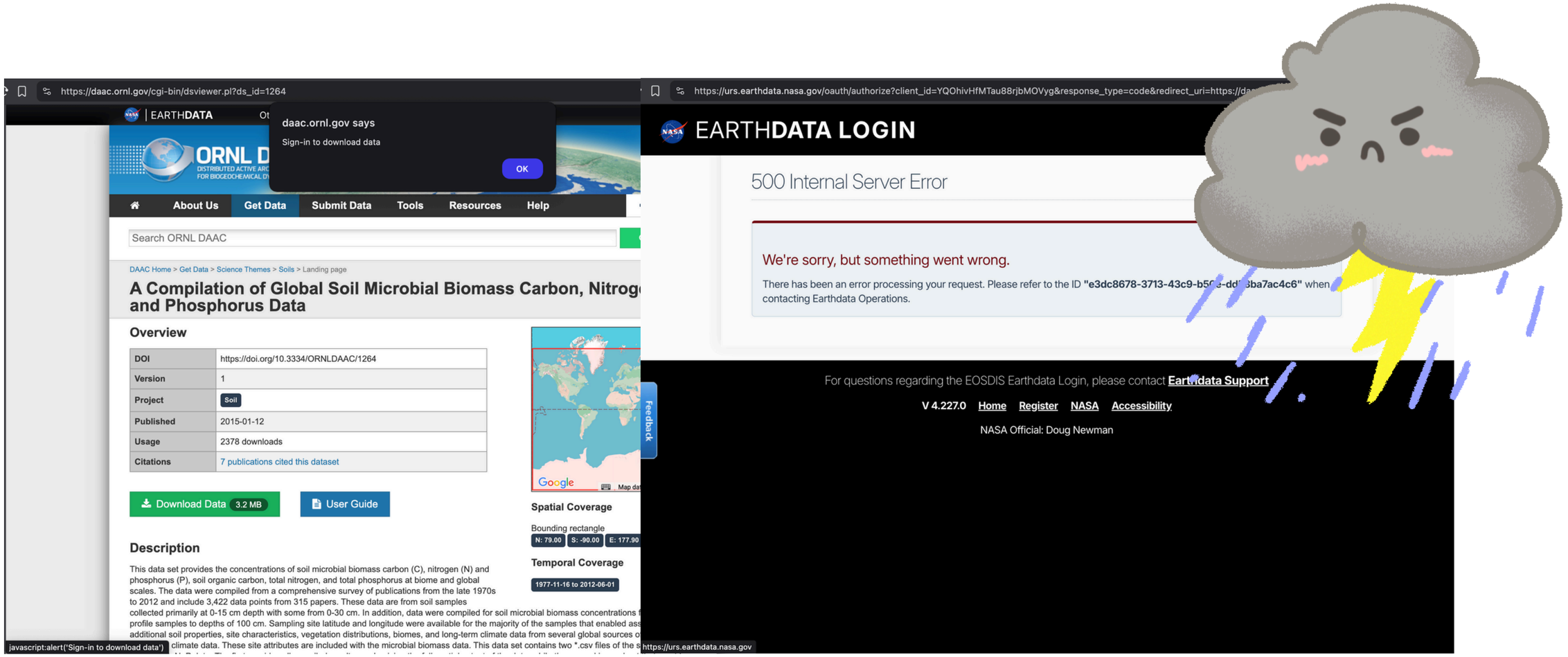
Hanaan Shafi

Marcell Juhász

Praggya Jeyakumar

Loki (team support cat)

27 April, 2025


https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1264
EARTHDATA
daac.ornl.gov says
Sign-in to download data
OK
About Us
Get Data
Submit Data
Tools
Resources
Help
Search ORNL DAAC
DAAC Home > Get Data > Science Themes > Soils > Landing page
A Compilation of Global Soil Microbial Biomass Carbon, Nitrog
and Phosphorus Data
Overview
DOI
https://doi.org/10.3334/ORNLDAAC/1264
Version
1
Project
Soil
Published
2015-01-12
Usage
2378 downloads
Citations
7 publications cited this dataset
Download Data 3.2 MB
User Guide
Spatial Coverage
Bounding rectangle
N: 79.00
S: -90.00
E: 177.90
Temporal Coverage
1977-11-16 to 2012-06-01
Description
This data set provides the concentrations of soil microbial biomass carbon (C), nitrogen (N) and
phosphorus (P), soil organic carbon, total nitrogen, and total phosphorus at biome and global
scales. The data were compiled from a comprehensive survey of publications from the late 1970s
to 2012 and include 3,422 data points from 315 papers. These data are from soil samples
collected primarily at 0-15 cm depth with some from 0-30 cm. In addition, data were compiled for soil microbial biomass concentrations
profile samples to depths of 100 cm. Sampling site latitude and longitude were available for the majority of the samples that enabled ass
additional soil properties, site characteristics, vegetation distributions, biomes, and long-term climate data from several global sources o
climate data. These site attributes are included with the microbial biomass data. This data set contains two *.csv files of the s
Feedback
javascript:alert('Sign-in to download data')
https://urs.earthdata.nasa.gov/oauth/authorize?client_id=YQOhivHfMTau88rjbMOVyg&response_type=code&redirect_uri=https://da
EARTHDATA LOGIN
500 Internal Server Error
We're sorry, but something went wrong.
There has been an error processing your request. Please refer to the ID "e3dc8678-3713-43c9-b5
ba7ac4c6" when
contacting Earthdata Operations.
For questions regarding the EOSDIS Earthdata Login, please contact Earthdata Support
V 4.227.0
Home
Register
NASA
Accessibility
NASA Official: Doug Newman
https://urs.earthdata.nasa.gov

🔥 Problems with Climate Data Storage

Political vulnerability

Centralized control

Infrastructure fragility

Funding instability

Security breaches

Solutions for Climate Data Storage 💧

Democratic accessibility

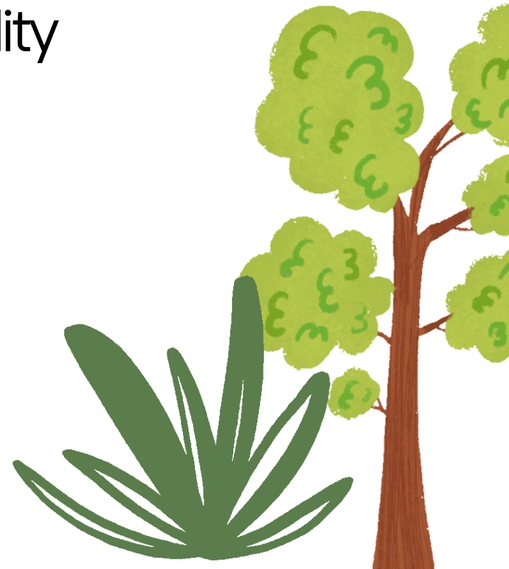
Distributed stewardship

Peer-to-peer architecture

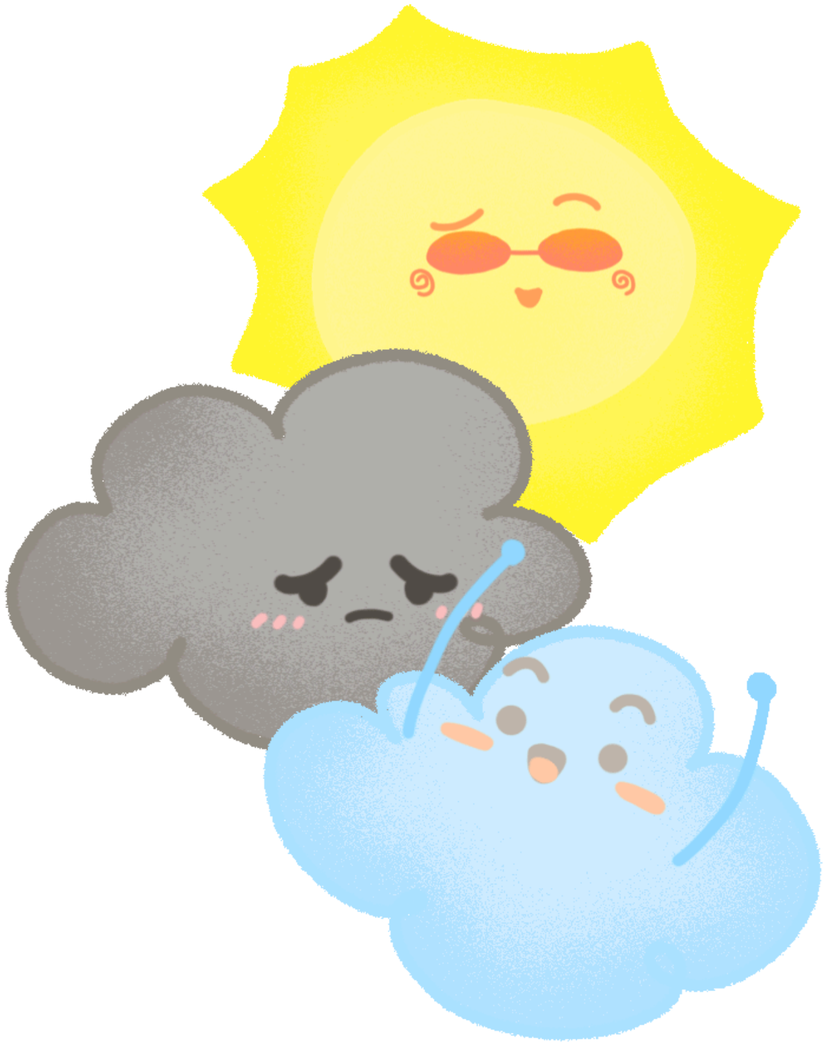
Crowdsourced scalability

Tamper-proof immutability

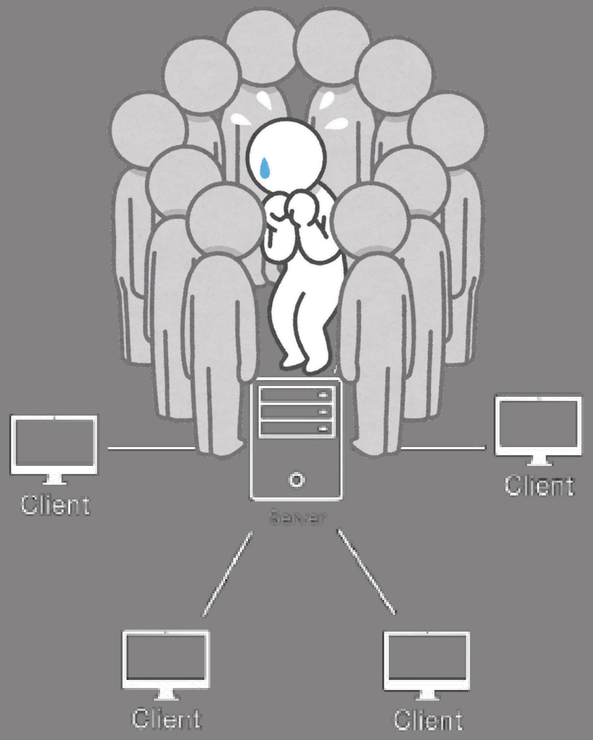




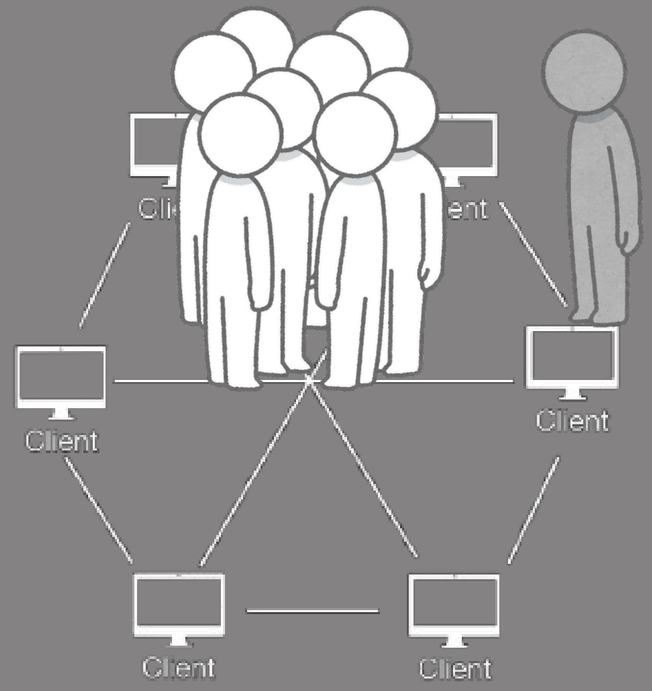
Tools & Technologies



Meet **BitTorrent** : protocol for peer-to-peer file sharing



traditional server



network of peers

Widely adopted

Decentralized

Cost efficient per node

Scalable

Meet DuckDB: fast, portable, and extensible database

```
D SELECT *, torrentfs_get_chunks(first_chunk, last_chunk) AS chunks
FROM torrentfs_info('magnet:?xt=urn:btih:NDSEA6BZTNRBUCOYJ6RNJZK5BM5RCP7G') LIMIT 10;
Torrentfs: Magnet link: magnet:?xt=urn:btih:NDSEA6BZTNRBUCOYJ6RNJZK5BM5RCP7G
Torrentfs: Metadata download status: Concluded
```

file_name varchar	file_size uint32	file_offset uint32	first_chunk uint32	last_chunk uint32	chunks int64[]
torrentBernardo/CSV/Normal/Country=Afghanistan/data_0.csv	558789	4294967295	0	1	[0, 1]
torrentBernardo/CSV/Normal/Country=Afghanistan/data_1.csv	105758	558788	1	1	[1]
torrentBernardo/CSV/Normal/Country=Afghanistan/data_2.csv	214300	664546	1	1	[1]
torrentBernardo/CSV/Normal/Country=Albania/data_0.csv	164931	878846	1	1	[1]
torrentBernardo/CSV/Normal/Country=Albania/data_1.csv	190331	1043777	1	2	[1, 2]
torrentBernardo/CSV/Normal/Country=Albania/data_2.csv	146009	1234108	2	2	[2]
torrentBernardo/CSV/Normal/Country=Algeria/data_0.csv	332217	1380117	2	3	[2, 3]
torrentBernardo/CSV/Normal/Country=Algeria/data_1.csv	156434	1712334	3	3	[3]
torrentBernardo/CSV/Normal/Country=Algeria/data_2.csv	317093	1868768	3	4	[3, 4]
torrentBernardo/CSV/Normal/Country=Angola/data_0.csv	563454	2185861	4	5	[4, 5]

10 rows 6 columns

Widely adopted

Open source

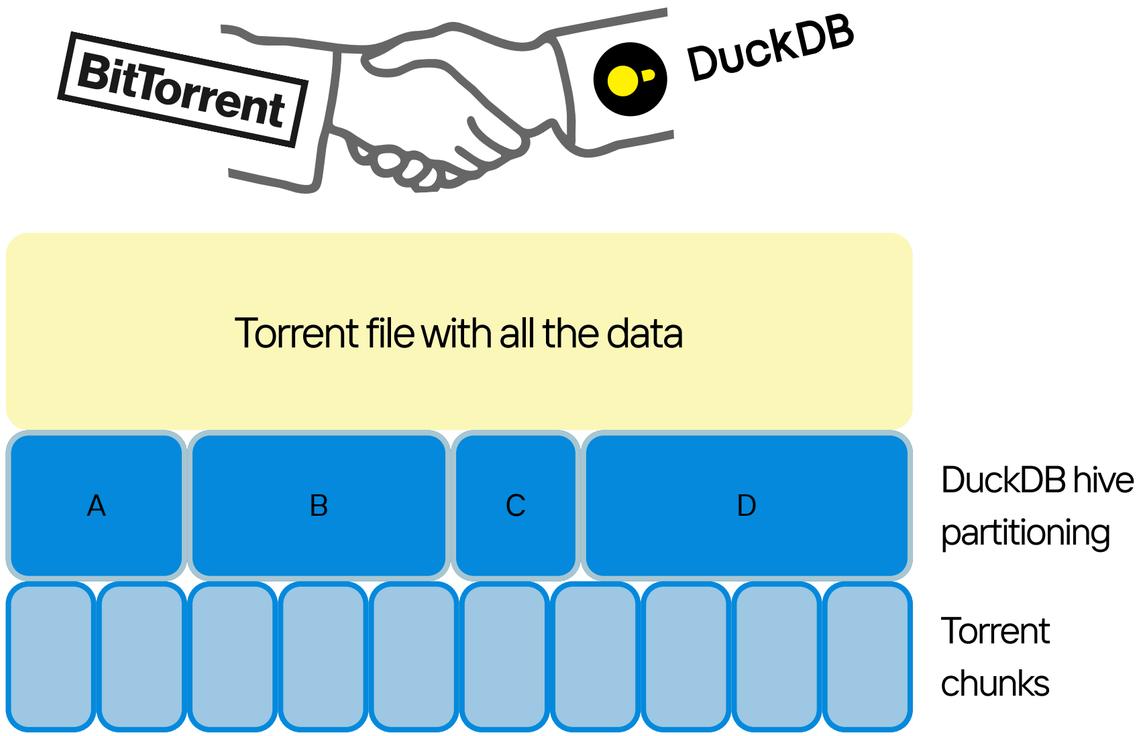
Embedded SQL DB engine

Hive partitioning support

Vast ecosystem of extensions

Solution: **BitTorrent extension for DuckDB**

(AKA query-driven torrent retrieval)



Why This Works?

Chunk Sharing

Every user becomes a mini-server for data they use

Smart Caching

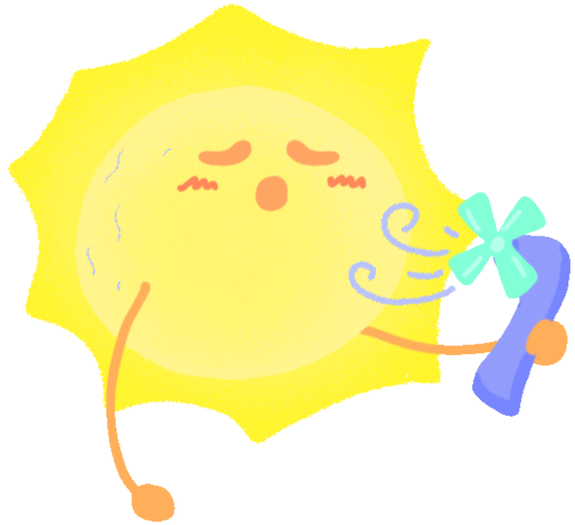
Keeps popular data widely available

Self-Healing

If you go offline, others use copies from alternate nodes

Where are we?

- ✅ Proof of concept in Python
- ✅ Functions in C++
  - Read metadata into table
  - File name, file size
  - First and last chunk
  - List of chunks in file



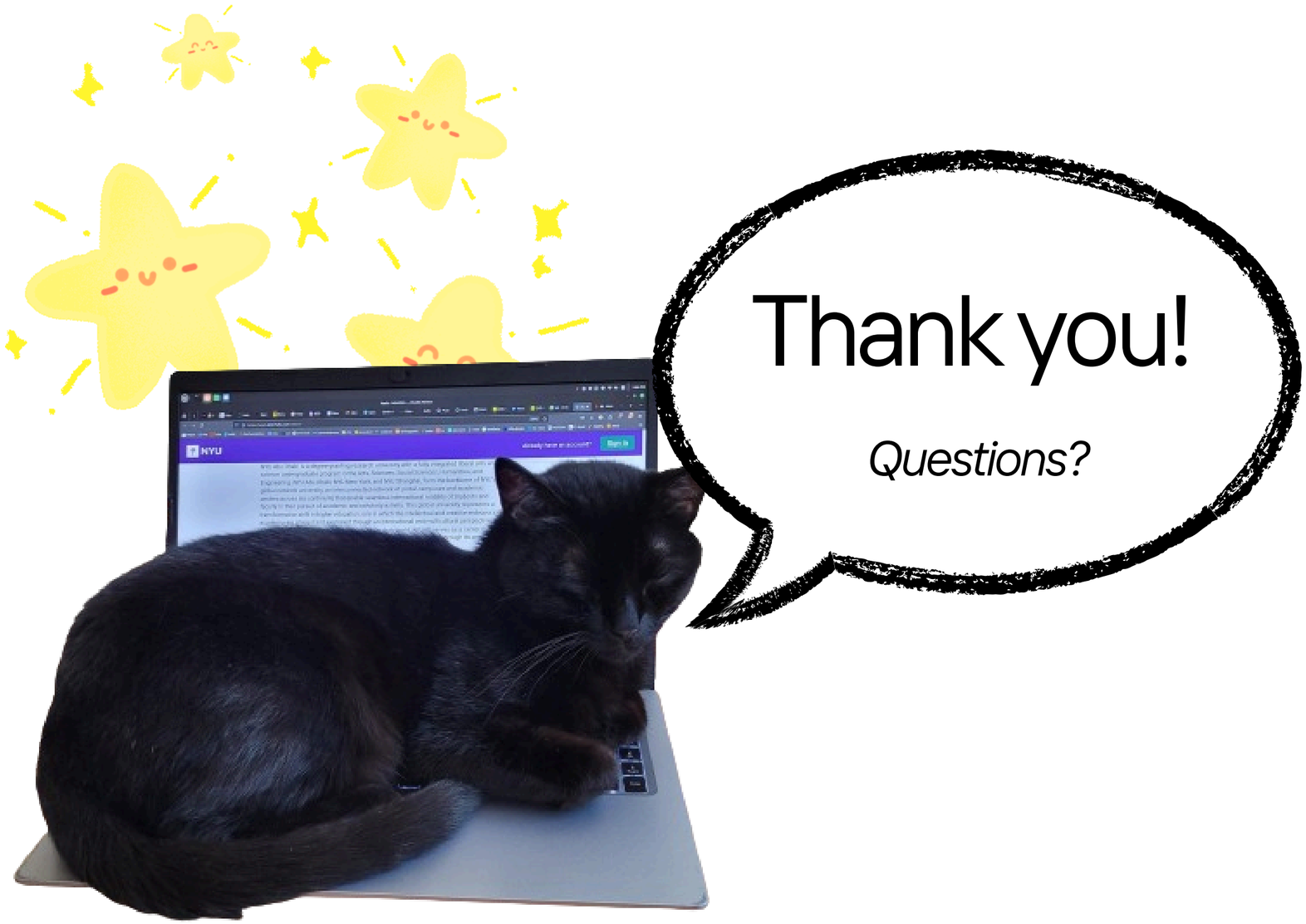
What's missing?

- Downloading said chunks
- Integrating it into the DuckDB virtual file system API
- User interface

Where do we go from here?

- Tech Implementation: **building in C++**
- Usability: **user-friendly interface**
- ✅ Impact: **building community + decentralization**
- ✅ Innovation: **connecting popular tools people already use**
- ✅ Sustainability: **cost-effectiveness + resilience**




Thank you!
Questions?