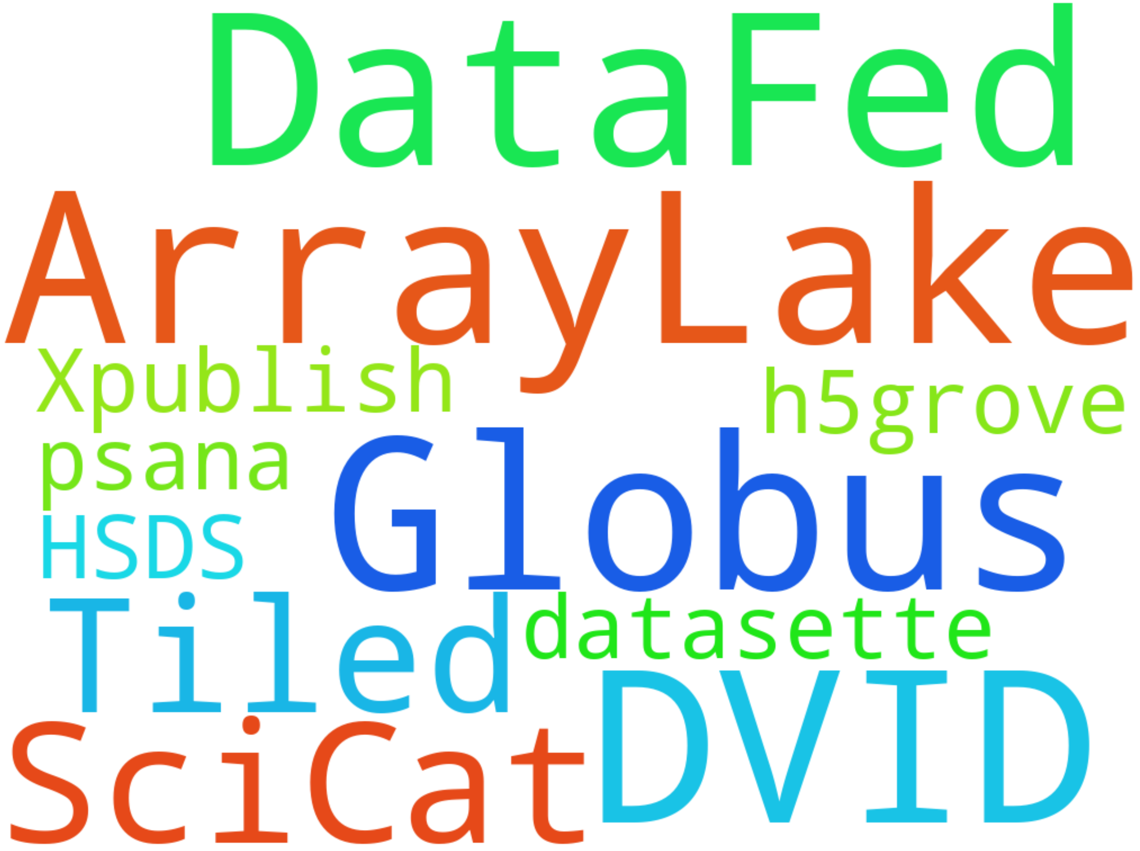
National Synchrotron Light Source II

Tiled in the Context of Data and Metadata Services

Daniel Allan

Data Engineering Group Lead
Data Science and Systems Integration Program, NSLS-II

There are many Data Services

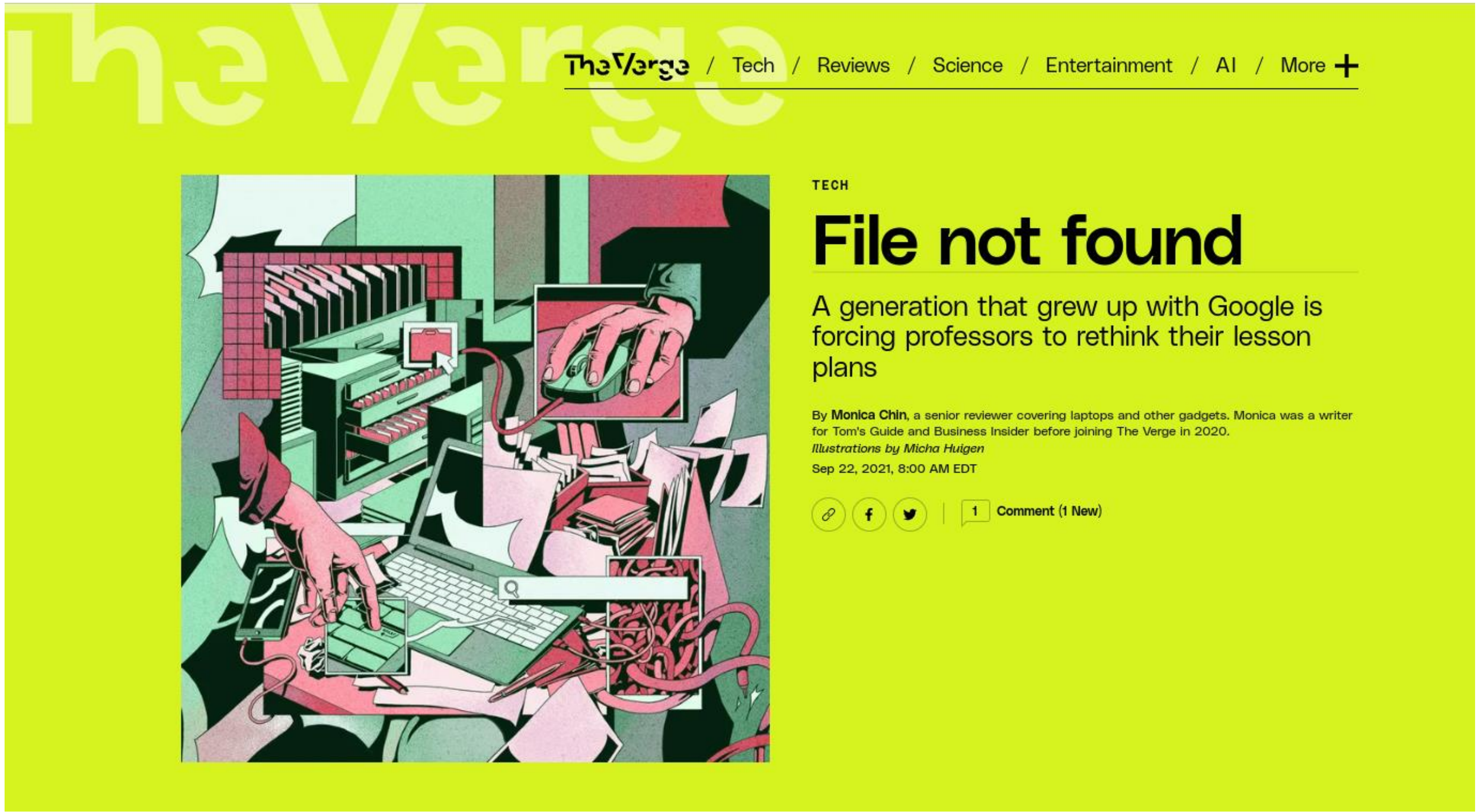


DataFed
ArrayLake
Xpublish
h5grove
psana
Globus
HSDS
Tiled
datasette
SciCat
DVID

Who Ordered a "Data Service"?

- Remove friction from data analysis at the small scale
- Enable new science at the large scale
- Make it easy for busy researchers to be F.A.I.R.
  (Findable, Accessible, Interoperable, Reusable)

User expectations are evolving


The Verge / Tech / Reviews / Science / Entertainment / AI / More +

TECH

File not found

A generation that grew up with Google is forcing professors to rethink their lesson plans

By **Monica Chin**, a senior reviewer covering laptops and other gadgets. Monica was a writer for Tom's Guide and Business Insider before joining The Verge in 2020.
Illustrations by Micha Huigen
Sep 22, 2021, 8:00 AM EDT

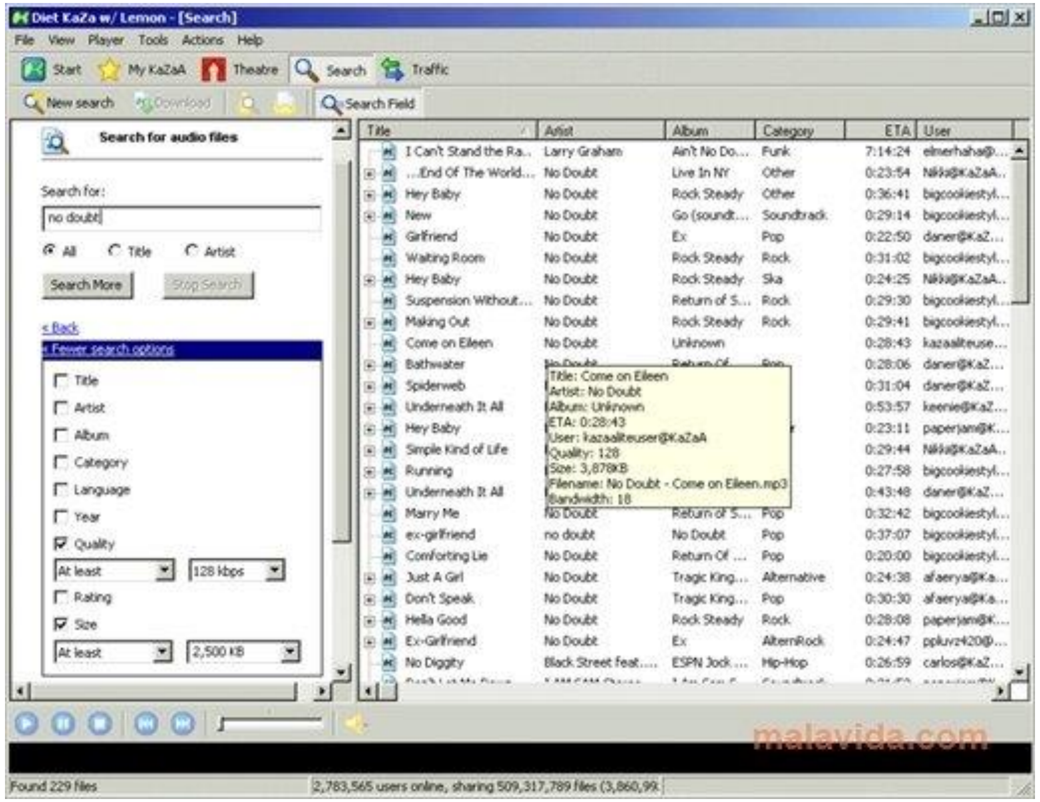
1 Comment (1 New)

https://www.theverge.com/22684730/students-file-folder-directory-structure-education-gen-z

Kids these days are accustomed to higher-level abstractions over their data.

Music Access Has Changed

c. 2000: folders of searchable MP3 files



Diet KaZa w/ Lemon - [Search]
File View Player Tools Actions Help
Start My KaZaA Theatre Search Traffic
New search Download Search Field
Search for audio files
Search More
Stop Search
< Back
« Fewer search options
malavida.com

Search for: no doubt

● All ○ Title ○ Artist

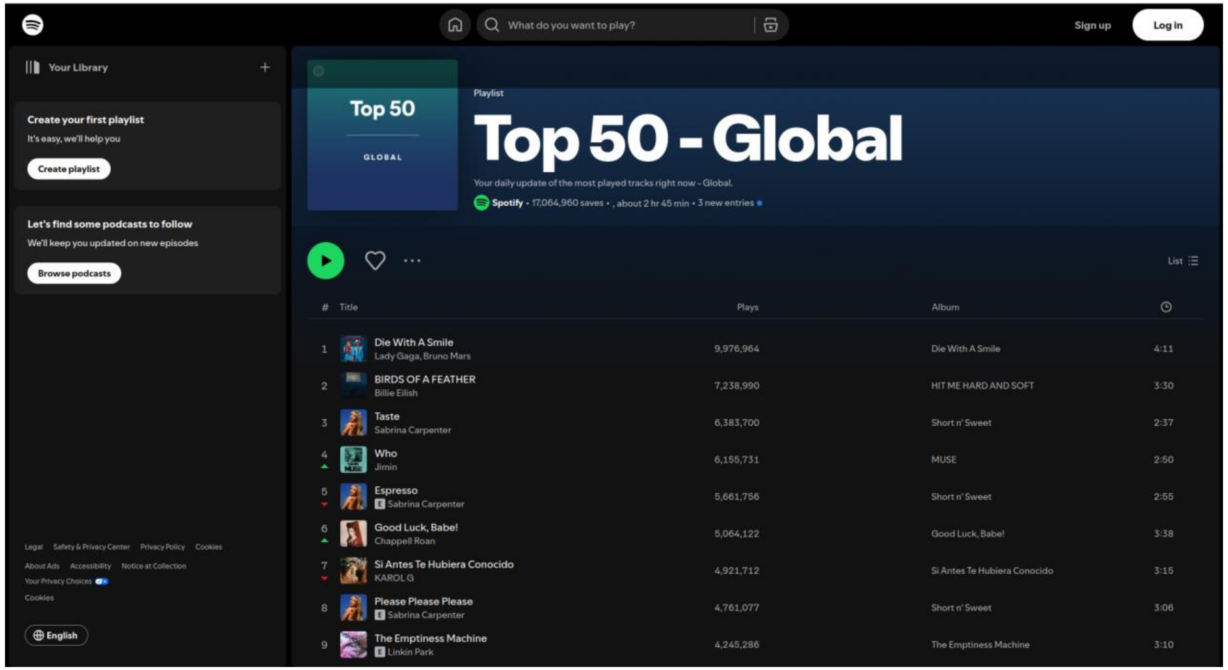
□ Title
□ Artist
□ Album
□ Category
□ Language
□ Year
☑ Quality
At least 128 kbps
□ Rating
☑ Size
At least 2,500 KB

Title	Artist	Album	Category	ETA	User
I Can't Stand the Ra...	Larry Graham	Ain't No Do...	Funk	7:14:24	elmerhaha@...
...End Of The World...	No Doubt	Live In NY	Other	0:23:54	Nikki@KaZaA...
Hey Baby	No Doubt	Rock Steady	Other	0:36:41	bigcookiestyl...
New	No Doubt	Go (soundt...	Soundtrack	0:29:14	bigcookiestyl...
Girlfriend	No Doubt	Ex	Pop	0:22:50	daner@KaZ...
Waiting Room	No Doubt	Rock Steady	Rock	0:31:02	bigcookiestyl...
Hey Baby	No Doubt	Rock Steady	Ska	0:24:25	Nikki@KaZaA...
Suspension Without...	No Doubt	Return of S...	Rock	0:29:30	bigcookiestyl...
Making Out	No Doubt	Rock Steady	Rock	0:29:41	bigcookiestyl...
Come on Eileen	No Doubt	Unknown		0:28:43	kazaaliteuse...
Bathwater	No Doubt	Return Of	Pop	0:28:06	daner@KaZ...
Spiderweb				0:31:04	daner@KaZ...
Underneath It All				0:53:57	keenie@KaZ...
Hey Baby				0:23:11	paperjam@K...
Simple Kind of Life				0:29:44	Nikki@KaZaA...
Running				0:27:58	bigcookiestyl...
Underneath It All				0:43:48	daner@KaZ...
Marry Me	No Doubt	Return of S...	Pop	0:32:42	bigcookiestyl...
ex-girlfriend	no doubt	No Doubt	Pop	0:37:07	bigcookiestyl...
Comforting Lie	No Doubt	Return Of ...	Pop	0:20:00	bigcookiestyl...
Just A Girl	No Doubt	Tragic King...	Alternative	0:24:38	afaerya@Ka...
Don't Speak	No Doubt	Tragic King...	Pop	0:30:30	afaerya@Ka...
Hella Good	No Doubt	Rock Steady	Rock	0:28:08	paperjam@K...
Ex-Girlfriend	No Doubt	Ex	AlternRock	0:24:47	ppluvz420@...
No Diggity	Black Street feat....	ESPN Jock ...	Hip-Hop	0:26:59	carlos@KaZ...

Title: Come on Eileen
Artist: No Doubt
Album: Unknown
ETA: 0:28:43
User: kazaaliteuser@KaZaA
Quality: 128
Size: 3,878KB
Filename: No Doubt - Come on Eileen.mp3
Bandwidth: 18

Found 229 files | 2,783,565 users online, sharing 509,317,789 files (3,860,99

Now: searchable collections of songs



What do you want to play?
Sign up
Log in
Your Library
Create playlist
Browse podcasts
Top 50
GLOBAL
List
Legal Safety & Privacy Center Privacy Policy Cookies
About Ads Accessibility Notice at Collection
Your Privacy Choices
Cookies
English

Create your first playlist

It's easy, we'll help you

Let's find some podcasts to follow

We'll keep you updated on new episodes

Playlist

Top 50 - Global

Your daily update of the most played tracks right now - Global.

Spotify • 17,064,960 saves • , about 2 hr 45 min • 3 new entries

#	Title	Plays	Album	🕒
1	Die With A Smile Lady Gaga, Bruno Mars	9,976,964	Die With A Smile	4:11
2	BIRDS OF A FEATHER Billie Eilish	7,238,990	HIT ME HARD AND SOFT	3:30
3	Taste Sabrina Carpenter	6,383,700	Short n' Sweet	2:37
4 ▲	Who Jimin	6,155,731	MUSE	2:50
5 ▼	Espresso E Sabrina Carpenter	5,661,756	Short n' Sweet	2:55
6 ▲	Good Luck, Babe! Chappell Roan	5,064,122	Good Luck, Babe!	3:38
7 ▼	Si Antes Te Hubiera Conocido KAROL G	4,921,712	Si Antes Te Hubiera Conocido	3:15
8	Please Please Please E Sabrina Carpenter	4,761,077	Short n' Sweet	3:06
9	The Emptiness Machine E Linkin Park	4,245,286	The Emptiness Machine	3:10

Comparison of Data Access Models

Serves	Scientific Data	Music
Files, metadata in filenames	SFTP	Kazaa, Napster
Files with rich metadata	DataFed, SciCat, Rucio ...	iTunes
Datasets with rich metadata	ArrayLake, DVID, h5grove, **Tiled** ...	Apple Music, Spotify

Critical Distinction:

Can the service open and understand the data it is serving?

File Catalogs Enable...

- User interfaces leveraging rich metadata
- Search
- File-based download
- Cross-site replication

```
https://...?filter=...&select=...
```

Structured Data Services Additionally Enable...

- Heterogenous data storage, including blob stores key-value stores, databases, detector memory...
- Partial and "chunk-aware" access: Download a region of interest first, or download chunks in parallel
- Transcoding: HDF5 or TIFF -> PNG

```
https://...?slice=50,100:200&format=image/png

https://...?column=temperature&column=intensity&format=text/csv
```

Important!
You can still get your data out

- *Apple Music* and *Spotify* don't let you access the underlying storage medium.
- Sometimes scientific applications have good reasons to do this.
  - Performance: Go around the service and straight to GPFS.
  - Backward-compatibility with existing workflows
- Structural data services can still simply provide the filepaths (or database/blob URIs) or serve raw assets to clients

A Structured Data Service Knows More

File Catalog Database Sketch:

- JSON Metadata blob
- Filepath

Structured Data Service Database Sketch:

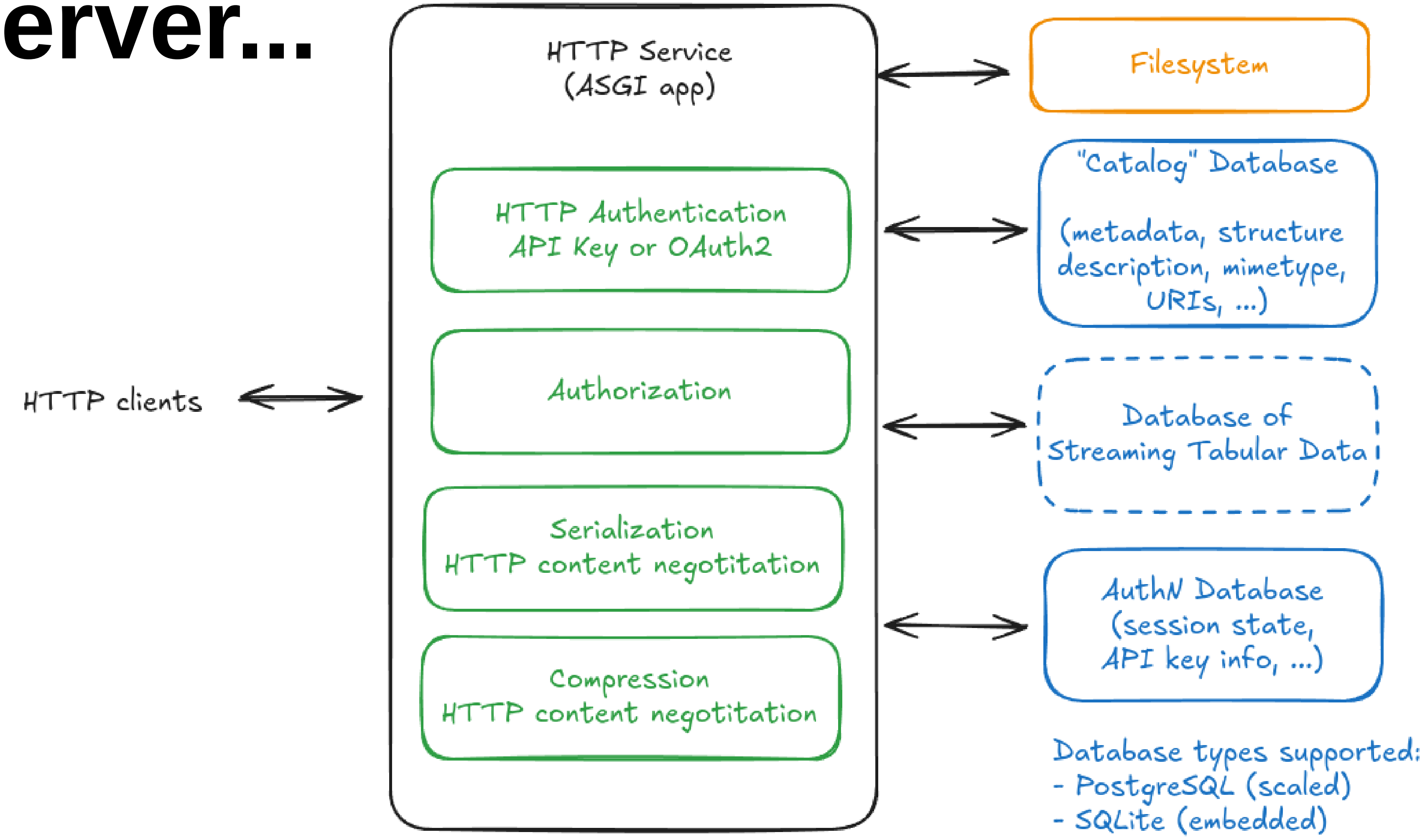
- JSON Metadata blob
- Data Structure Description
  (array data type and shape,
  table column names and data types,…)
- Mimetype (e.g. 'image/tiff')
- Parameters for accessing data within file
  (e.g. HDF5 dataset path)
- URI(s)

Tiled is...

1. A secure HTTP structural data service, for reading and writing
2. A Python client to that service, like generalized h5py
3. A proof-of-concept React app on that service

… with a budding ecosystem of third-party client applications

1. The server...


HTTP Service
(ASGI app)
HTTP Authentication
API Key or OAuth2
Authorization
Serialization
HTTP content negotitation
Compression
HTTP content negotitation
HTTP clients
Filesystem
"Catalog" Database
(metadata, structure description, mimetype, URIs, ...)
Database of Streaming Tabular Data
AuthN Database
(session state, API key info, ...)
Database types supported:
- PostgreSQL (scaled)
- SQLite (embedded)

2. The Python client...

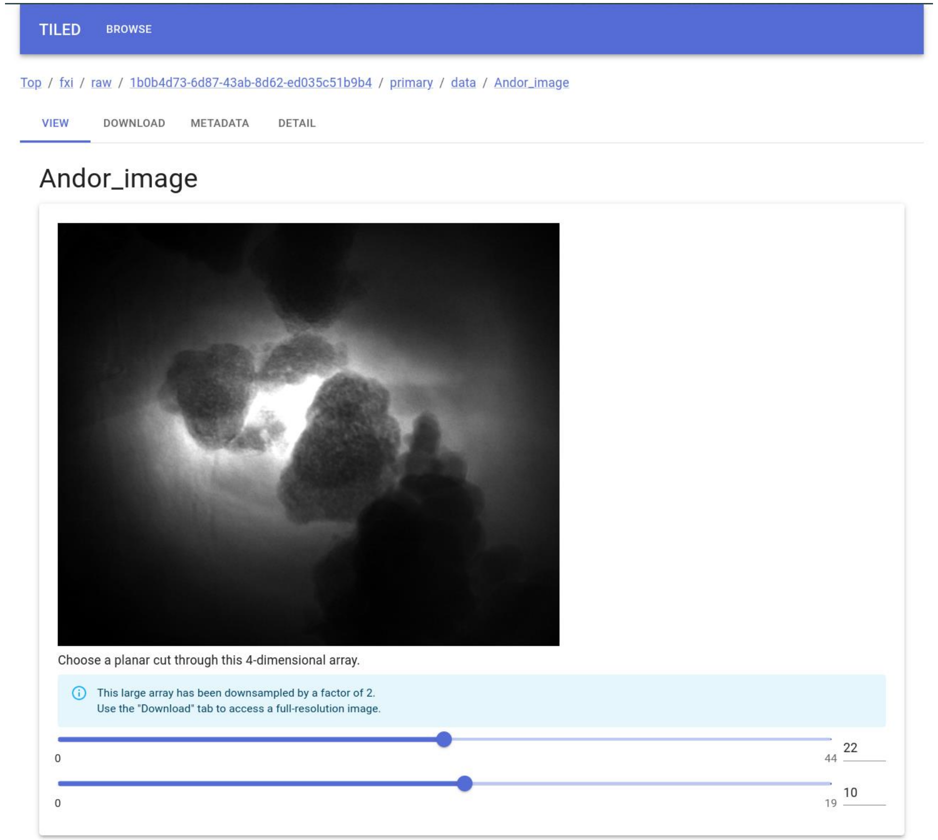
```
>>> from tiled.client import from_uri


>>> client = from_uri("http://localhost:8000")
>>> client
<Container {'some_image', ...} ~500000 entries>


>>> client['some_image'][:]
array([...])


>>> client.write_dataframe({"x": [1, 2, 3], "y": [4, 5, 6]}, key='some_table')
```

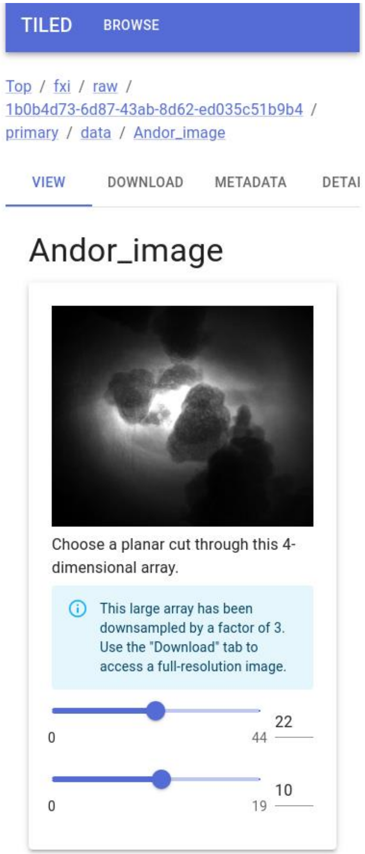
3. The prototype React app...


TILED
BROWSE
Top / fxi / raw / 1b0b4d73-6d87-43ab-8d62-ed035c51b9b4 / primary / data / Andor_image
VIEW
DOWNLOAD
METADATA
DETAIL
0
44
22
0
19
10

Andor_image

Choose a planar cut through this 4-dimensional array.

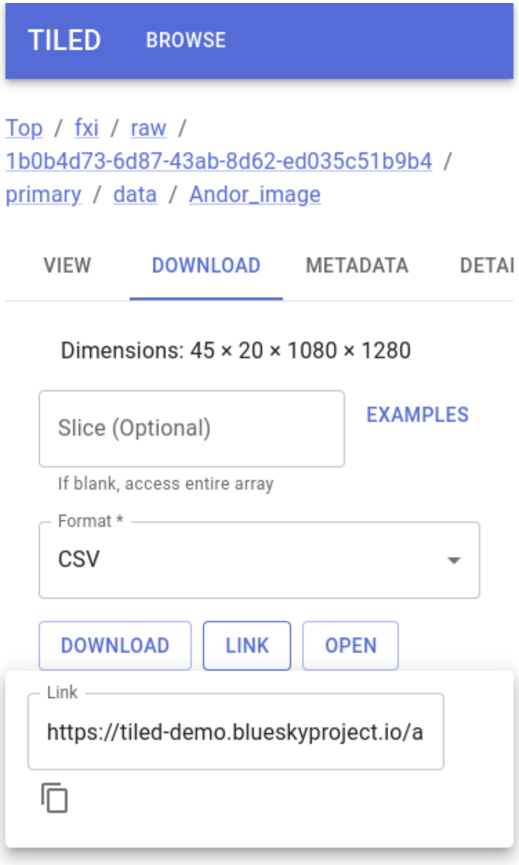
This large array has been downsampled by a factor of 2.
Use the "Download" tab to access a full-resolution image.


TILED
BROWSE
Top / fxi / raw / 1b0b4d73-6d87-43ab-8d62-ed035c51b9b4 / primary / data / Andor_image
VIEW
DOWNLOAD
METADATA
DETAI
0
44
22
0
19
10

Andor_image

Choose a planar cut through this 4-dimensional array.

This large array has been downsampled by a factor of 3. Use the "Download" tab to access a full-resolution image.


TILED
BROWSE
Top / fxi / raw / 1b0b4d73-6d87-43ab-8d62-ed035c51b9b4 / primary / data / Andor_image
VIEW
DOWNLOAD
METADATA
DETAI
EXAMPLES
DOWNLOAD
LINK
OPEN

Dimensions: 45 × 20 × 1080 × 1280

Slice (Optional)

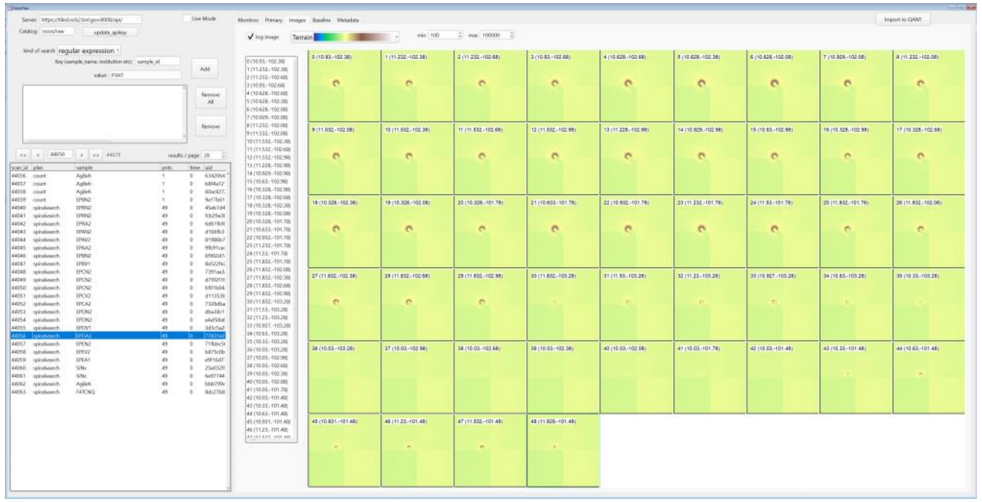
If blank, access entire array

Format *: CSV

Link: https://tiled-demo.blueskyproject.io/a

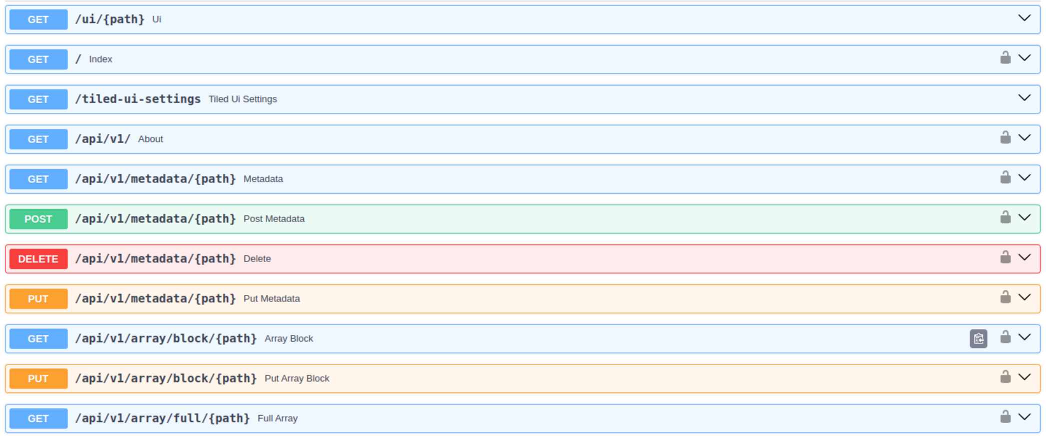
Third-party client applications...

Eliot Gann's custom (scientist-written!) **Igor** program loads data from Tiled over HTTP, integrating search and viz



PyMCA integration is in progress!

And of course it works from `curl...`



GET /ui/{path} Ui

GET / Index

GET /tiled-ui-settings Tiled Ui Settings

GET /api/v1/ About

GET /api/v1/metadata/{path} Metadata

POST /api/v1/metadata/{path} Post Metadata

DELETE /api/v1/metadata/{path} Delete

PUT /api/v1/metadata/{path} Put Metadata

GET /api/v1/array/block/{path} Array Block

PUT /api/v1/array/block/{path} Put Array Block

GET /api/v1/array/full/{path} Full Array

Tiled as an index and an HTTP transport

	Index	HTTP transport
Reading	Tiled provides JSON with URI, e.g. `file://.../data.h5`	Download data from Tiled in any supported format
Writing	Register externally-managed files for Tiled to index and serve	Upload data to Tiled in any supported format for Tiled to store

Vital Statistics for *Tiled*

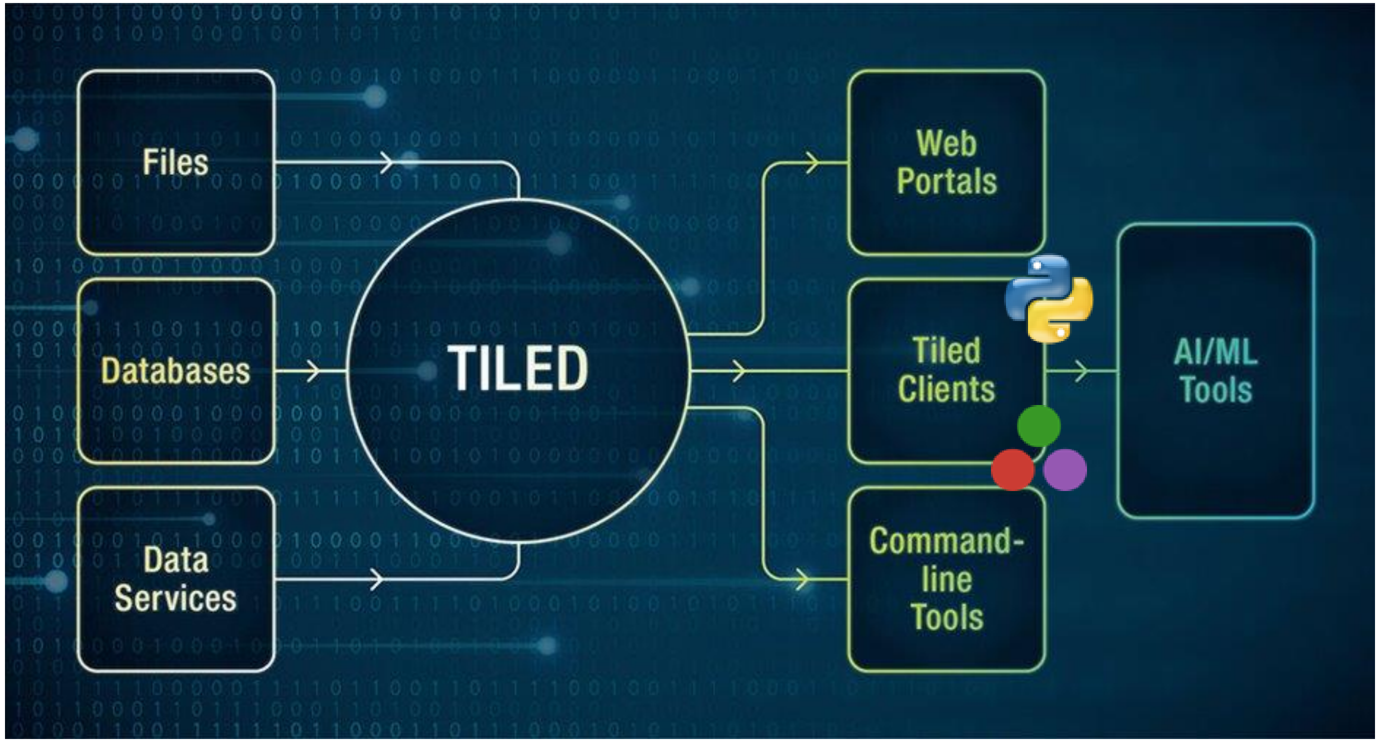
- Developed in the open, **BSD-3** license
- First commit February 2021
- **38** unique code contributors with ~10 unique affiliations
- Under Bluesky **Governance**, may split out on its own someday
- Early usage at: NIST, APS, ALS, BESSY-II, Australian Synchrotron
- **Security** periodically intensively vetted by a third-party cybersecurity penetration testing

Tiled incorporates the input and experience of many people

- **NSLS-II**: Dan Allan, Stuart Campbell, Thomas Caswell, Padraic Shafer, Marcus Hanwell, Juan Marulanda, Kari Barry, Eugene Matviychuk, Seher Karakuzu, Hiran Wijesinghe
- **ALS**: Dylan McReynolds, Joseph Kleinhenz, Wiebke Koepp
- Builds on open-source collaboration with Martin Durant (**Anaconda, Inc.**) with involvement from Garrett Bischof (**NSLS-II**).
- Contributions from the **IRIS-HEP** collaboration to add support for AwkwardArray

Tiled: A Structured API to Data

- **Search** on metadata
- **Slice** into remote datasets
- **Transcode** between formats
- **Download** partial or whole datasets
- Or **find** data storage location(s) for direct access
- Implement web **security** standards and authorization



Links

Demo: https://tiled-demo.blueskyproject.io/

Documentation: https://blueskyproject.io/tiled

Code: https://github.com/bluesky/tiled

Contact: Daniel Allan <dallan@bnl.gov>