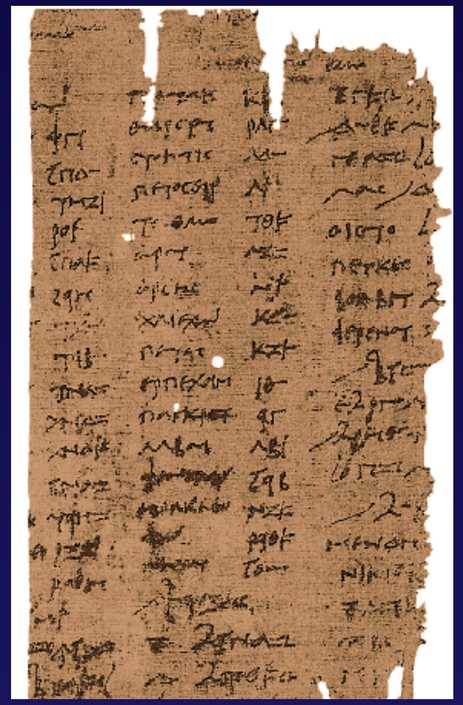
A CLOUD GURU

Streaming Data Collection Concepts

Brock Tubre
INSTRUCTOR



age	village	town
0-4	24	26
5-9	16	26
10-14	7	16
15-19	13	23
20-24	13	18
25-29	14	12
30-34	16	19
35-39	10	6
40-44	12	6
45-49	13	11
50-54	4	6
55-59	11	3
60-64	3	3
65-69	5	2
70-74	6	3
75-79	2	0
80-84	0	1

Early Data Collection

Census of Egypt

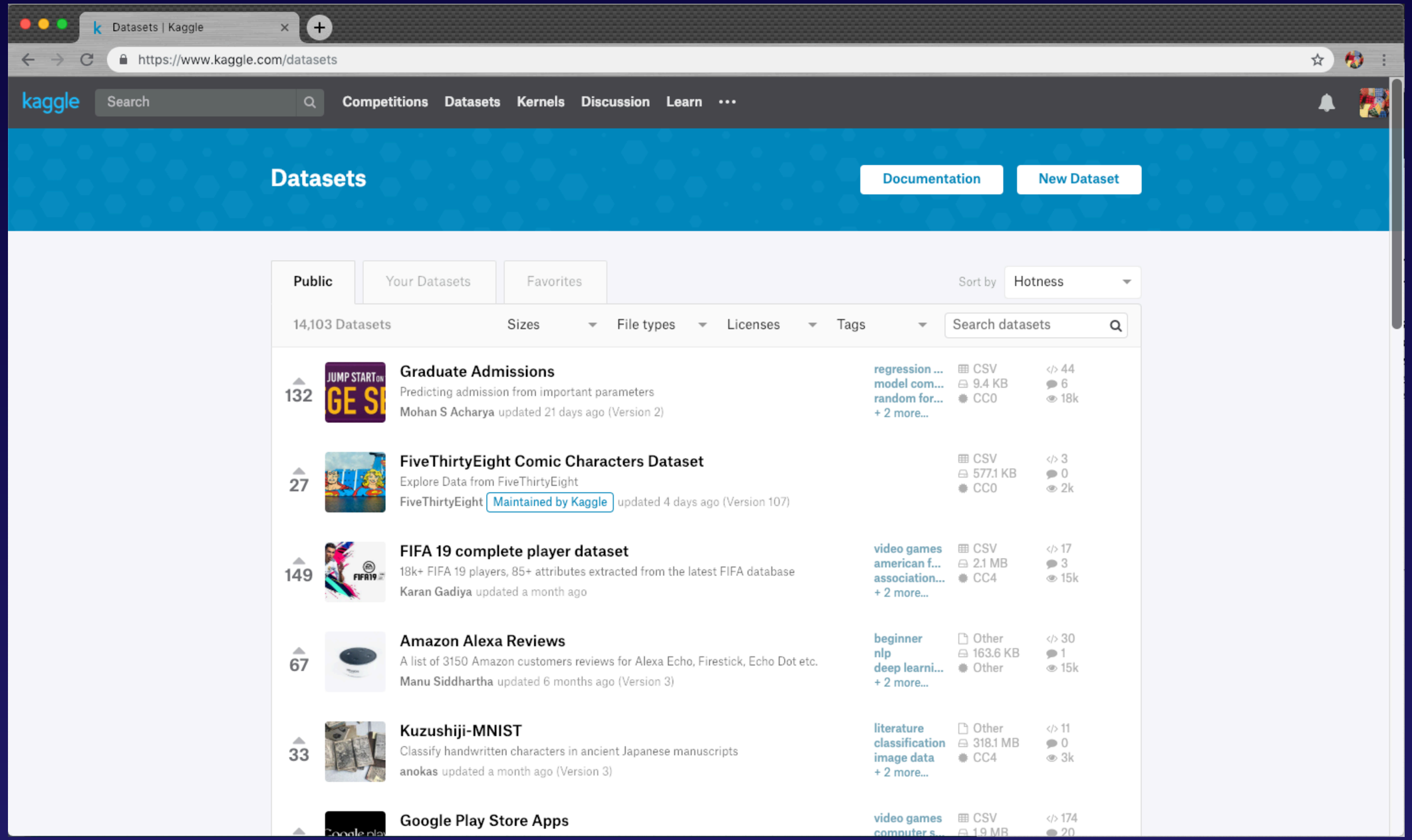
The practice of collecting census data began in Egypt second millennium BC (2000 through 1001 BC), where it was used for tax gathering and to determine fitness for military services.

Where does data come from?

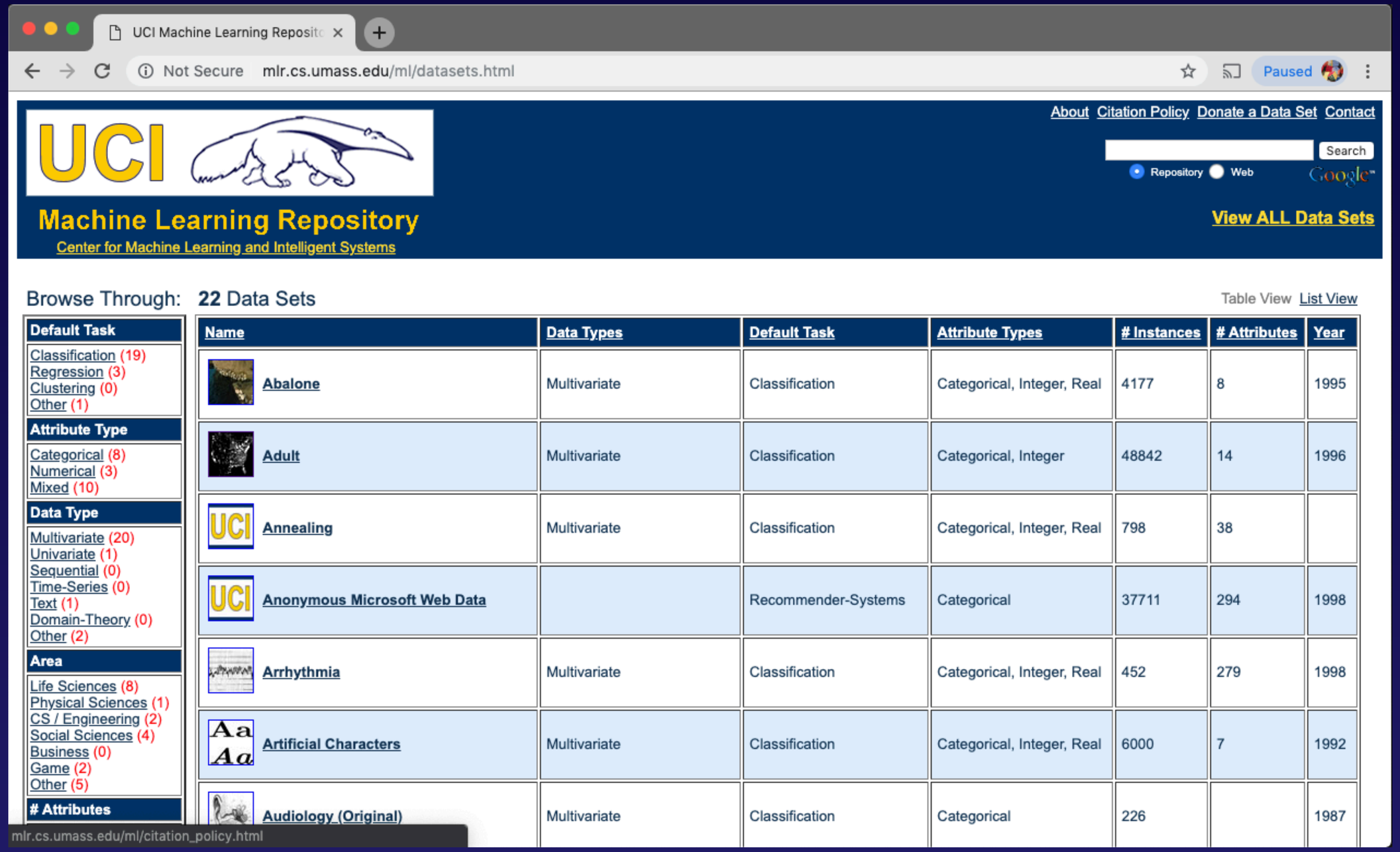
A CLOUD GURU

WHERE DOES DATA COME FROM?

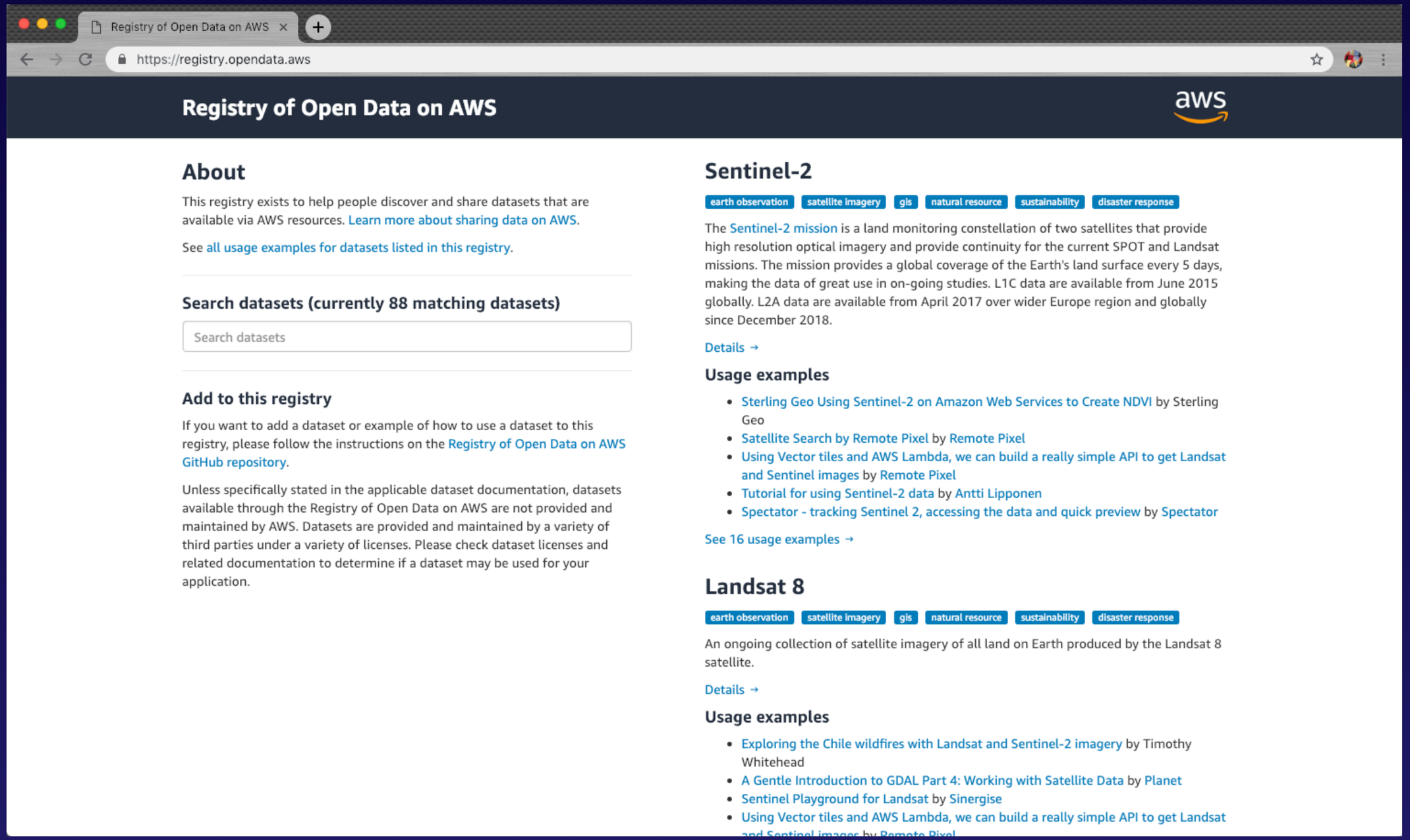
Kaggle


Datasets | Kaggle
https://www.kaggle.com/datasets
kaggle
Search
Competitions
Datasets
Kernels
Discussion
Learn
Datasets
Documentation
New Dataset
Public
Your Datasets
Favorites
Sort by
Hotness
14,103 Datasets
Sizes
File types
Licenses
Tags
Search datasets
132
Graduate Admissions
Predicting admission from important parameters
Mohan S Acharya updated 21 days ago (Version 2)
regression ...
model com...
random for...
+ 2 more...
CSV
9.4 KB
CC0
44
6
18k
27
FiveThirtyEight Comic Characters Dataset
Explore Data from FiveThirtyEight
FiveThirtyEight
Maintained by Kaggle
updated 4 days ago (Version 107)
CSV
577.1 KB
CC0
3
0
2k
149
FIFA 19 complete player dataset
18k+ FIFA 19 players, 85+ attributes extracted from the latest FIFA database
Karan Gadiya updated a month ago
video games
american f...
association...
+ 2 more...
CSV
2.1 MB
CC4
17
3
15k
67
Amazon Alexa Reviews
A list of 3150 Amazon customers reviews for Alexa Echo, Firestick, Echo Dot etc.
Manu Siddhartha updated 6 months ago (Version 3)
beginner
nlp
deep learni...
+ 2 more...
Other
163.6 KB
Other
30
1
15k
33
Kuzushiji-MNIST
Classify handwritten characters in ancient Japanese manuscripts
anokas updated a month ago (Version 3)
literature
classification
image data
+ 2 more...
Other
318.1 MB
CC4
11
0
3k
Google Play Store Apps
video games
CSV
174

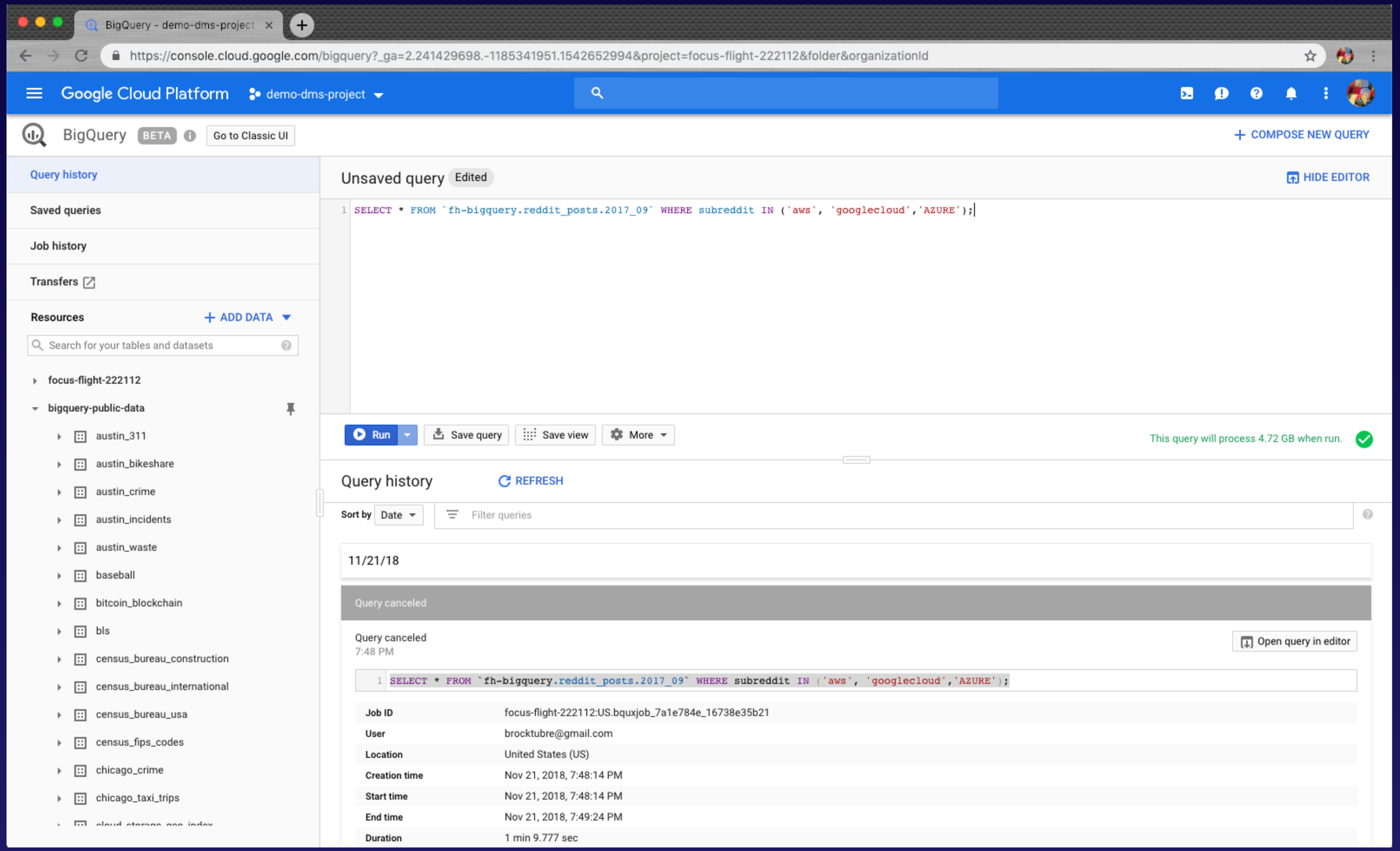
UCI Machine Learning Repository



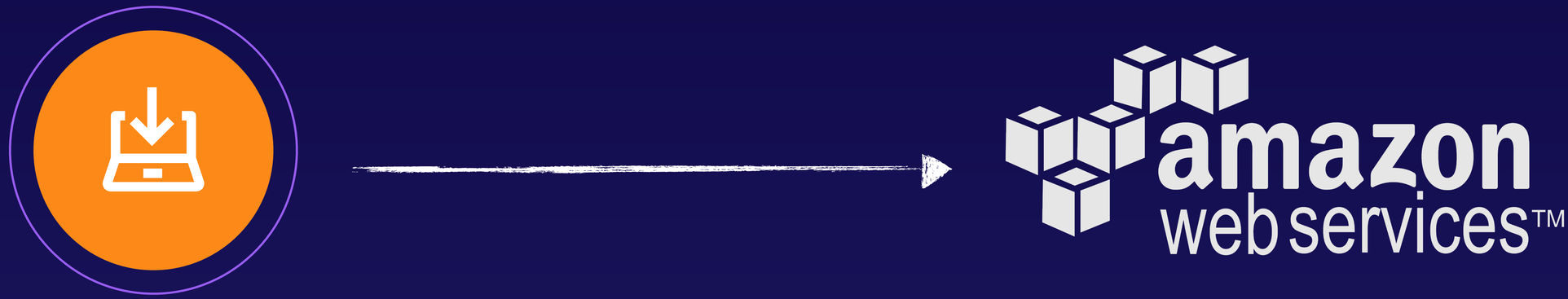
Open Data on AWS



Google's Big Query



Load Data into AWS



What about streaming data?

A CLOUD GURU

WHAT ABOUT STREAMING DATA?

Streaming Data

Streaming Data

?

amazon web services™

Streaming Data

Kinesis

A CLOUD GURU

The Marvel Family

“The Shazam Family”

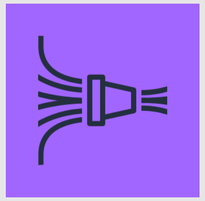
Family of superheroes created in 1942 including characters such as Captain Marvel “Shazam”, Mary Marvel, Captain Marvel Jr., Uncle Marvel.

The Kinesis Family

The Kinesis Family

Kinesis

Kinesis
Data Streams



Kinesis
Data Firehose

Kinesis
Video Streams

Kinesis
Data Analytics