Data Drives “Smart”

Glenn.Ricart@us-ignite.org
Otober 23, 2020
U Cincinnati Data Day 2020

usignite

THE UNIVERSITY OF UTAH®

NSF

is a **nonprofit**
accelerating the **smart city movement**
and creating **value** for an entire **ecosystem**
by **guiding communities** into the **connected future**,
creating a path for **private sector growth**,
and **advancing technology research**
that's at the heart of **smart city** development.

How Much Data is About You?

Measuring Volumes of Data

1 Kilobyte (KB) = 1000 characters (bytes) of data
1 Megabyte (MB) = 1,000,000 bytes
1 Gigabyte (GB) = 1,000,000,000 bytes (English Wikipedia text approx. 20 GB)
1 Terabyte (TB) = 1,000,000,000,000 bytes (All of Glenn's photos – lifetime – 50,000)
1 Petabyte (PB) = 1,000,000,000,000,000 bytes (Human brain approx. 2.5 PB)
1 Exabyte (EB) = 1,000,000,000,000,000,000 bytes (All the files backed up by Backblaze)
1 Zettabyte (ZB) = 1,000,000,000,000,000,000,000 bytes

Measuring Speeds to Move Data
1 Kbps = 100 bytes / sec. (dial-up modem)
1 Mbps = 100 KB / sec. (25 Mbps (down) = broadband (FCC))
1 Gbps = 100 MB / sec. (Google Fiber introduced 1 Gbps; fastest networks now 10 Gbps)
1 Tbps = 100 GB / sec. (FABRIC Research Network intended speed per fiber)

Measuring Time
1 ms = 0.001 seconds
1 µs = 0,000001 seconds

History of speeds to homes and small businesses

1970s: Dial-up, 14.4 **K**bps ARPAnet only

1980s: Dial-up, 28 **K**bps AOL & others

1990s: Cable, 10 **M**bps; graphics & browser

2000s: DSL at 1.5 **M**bps; cable at 10 **M**bps

2.5% of homes to 28%

2010s: Fiber at 1 **G**bps

2020s: Fiber at 10 **G**bps

2030s: FABRIC at 1 **T**bps (in research now)

Glenn's Postulate: speeds double every two years

usignite

renci · University of Kentucky · ESnet Energy Sciences Network · Clemson University · Illinois Institute of Technology

University of Virginia · The University of Utah · The University of Chicago · UC San Diego · FIU Florida International University · Georgia Tech

Rutgers · Internet2 · Columbia University in the City of New York · SRI International · TACC · Illinois NCSA · usignite

FABRIC

INTERNET2®

usignite

Information transmission rates of the senses

sensory system	bits per second
eyes	10,000,000
skin	1,000,000
ears	100,000
smell	100,000
taste	1,000

Source: Encyclopedia Brittanica, article on Psysiology, retrieved online 10/21/2020.

usignite

Size of Glenn's Data (personal account)

My personal data being kept at home = 10 TB

Everything Google says they have about me = 80 GB

Your account, your data.

We've finished creating a copy of the Google data you requested on October 19, 2020. However, we were unable to create a copy of all your files. You can still download your files until October 26, 2020, but please be aware that there will be data missing. Inside your export you will find an html file that will have more details about the errors.

Your download will contain data from 57 products.

usignite

We've received and are processing your request to access your personal data.

We will provide your information to you as soon as we can. Usually, this should not take more than a month. In exceptional cases, for example if a request is more complex or if we are processing a high volume of requests, it might take longer, but if so we will notify you that there will be a delay.

Size of Glenn's Data (personal account)

My personal data being kept at home = 10 TB
Everything Google says they have about me = 80 GB
Number of web sites storing stuff about me = 187
Number of remembered passwords = 746
Number of authenticators = 6

Continuous Reporting

My location
My heart rate, steps, floors climbed, distance traveled, Afib, SpO_2
Presence detected by Alexa and Google

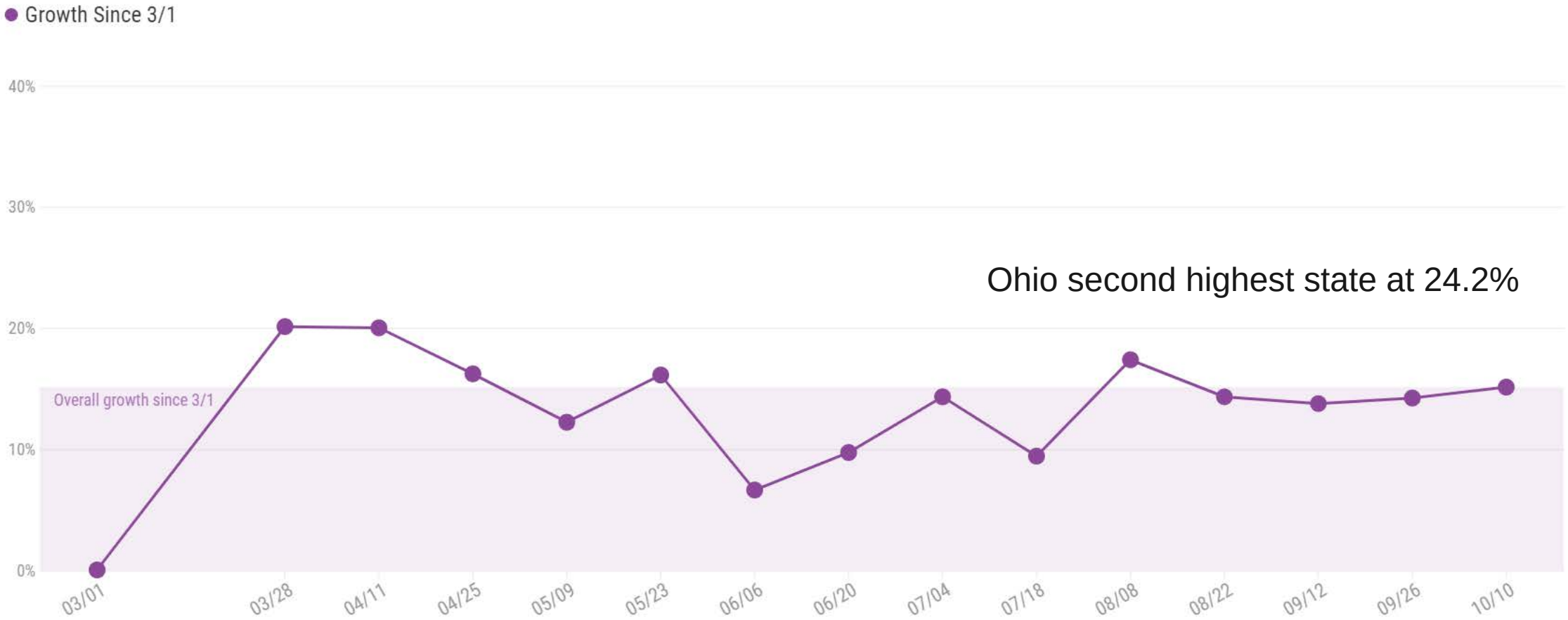
Home Internet Usage (no TV) during COVID-19

6 GB down; 6 GB up; per day = 12 GB/day

National Downstream Peak Growth

Observed Increase in Peak Consumer Usage

Change in Pre-COVID Internet Usage Since Early March



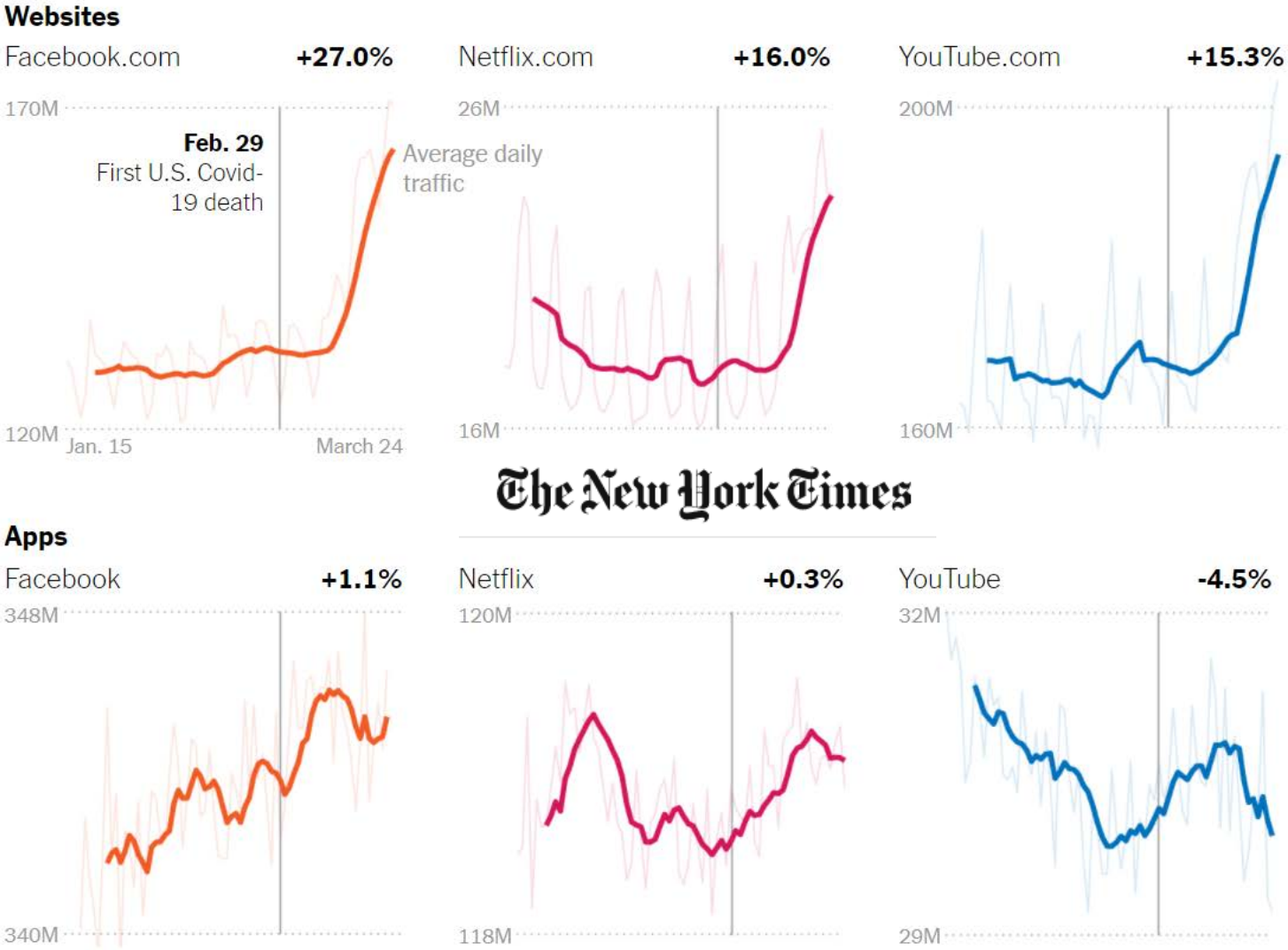
Source: Data from NCTA Member Companies and Others

National Upstream Peak Growth

Observed Increase in Peak Consumer Usage

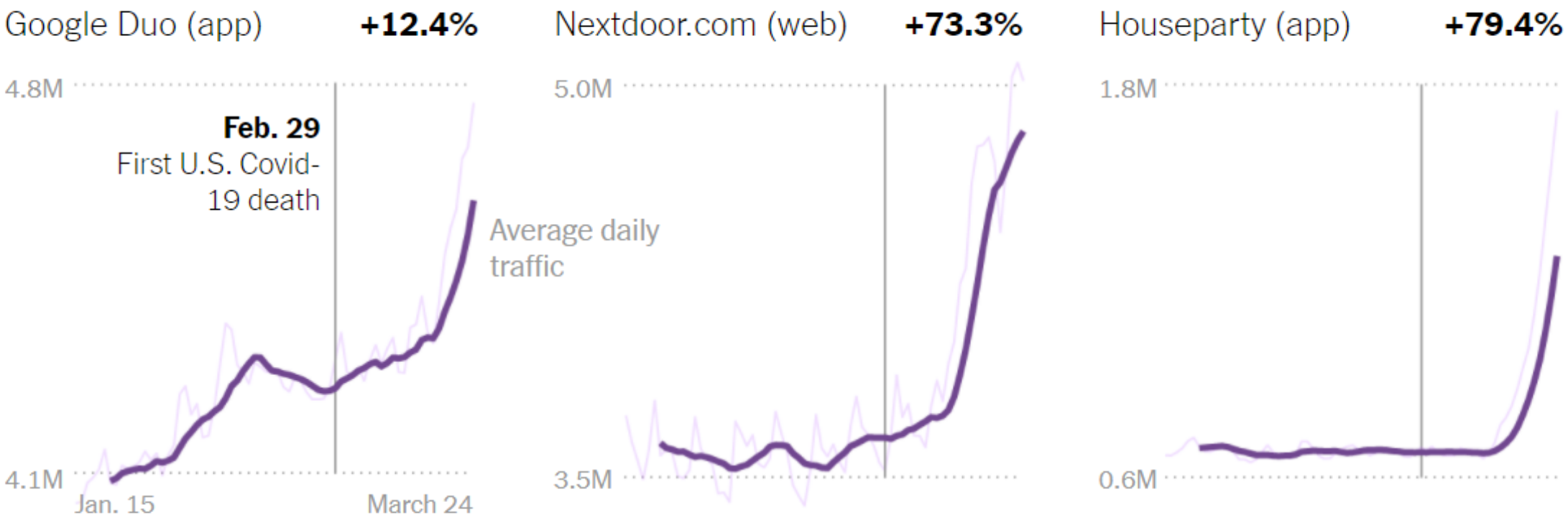
Change in Pre-COVID Internet Usage Since Early March

Source: Data from NCTA Member Companies and Others



Note: Averages are calculated with traffic numbers from each date and the six days preceding it in order to smooth out weekly variations (recreational internet use, for example, often spikes on the weekends). Percent change is from the average on Jan. 21 to the average on March 24. Daily app traffic is measured in sessions — the number of times the app is opened — and one user can have multiple sessions in a day. • Sources: SimilarWeb, Apptopia

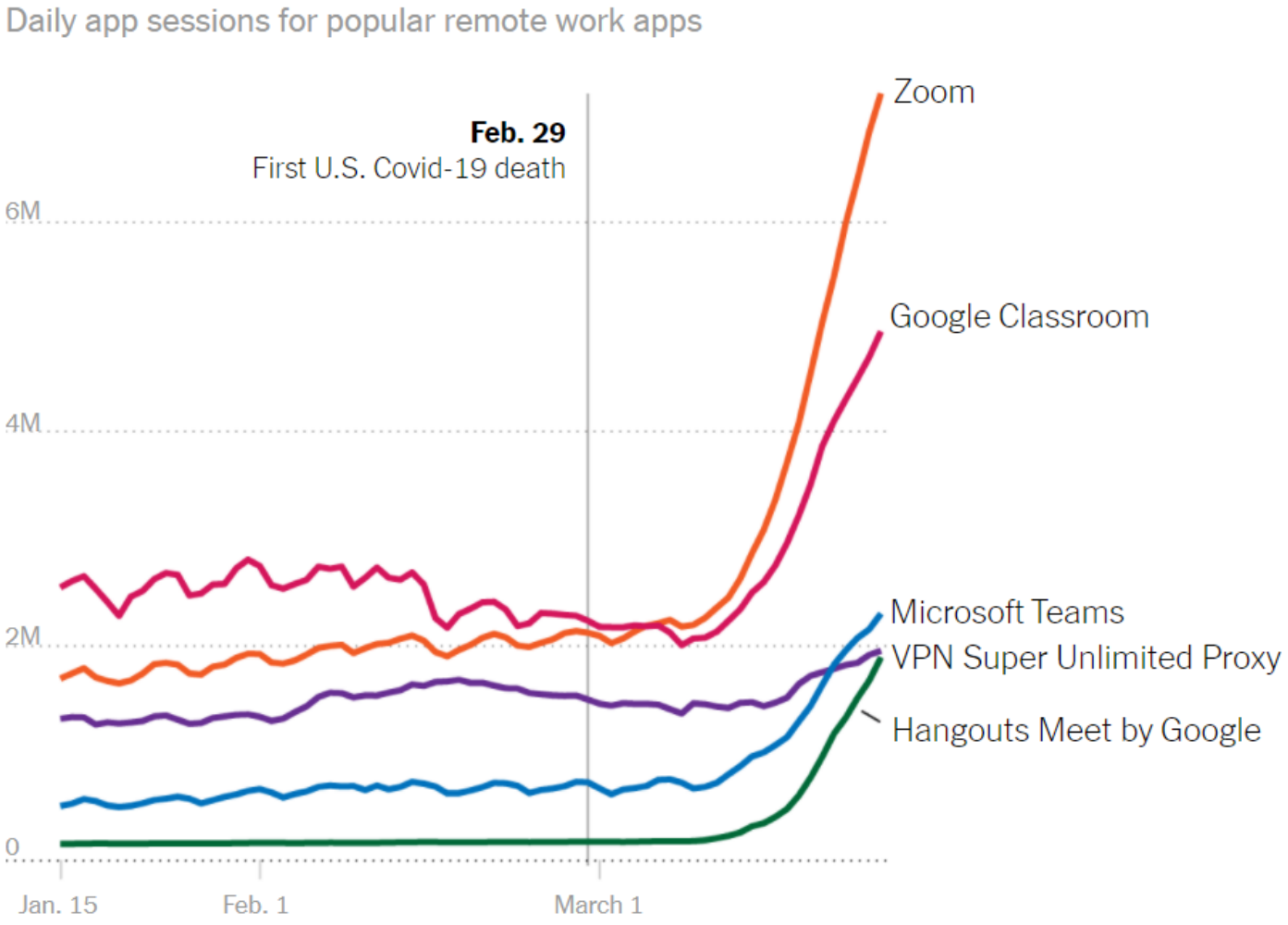
With the rise of social distancing, we are seeking out new ways to connect, mostly through video chat



Percent change is from the average on Jan. 21 to the average on March 24. App daily traffic is measured in "daily sessions." · Sources: SimilarWeb, Apptopia

ignite

We have suddenly become reliant on services that allow us to work and learn from home



App popularity according to iOS App Store rankings on March 16-18. • Source: Apptopia

usignite

Data Drives Smart Things

Smart City

Smart Thermostat

Smart Watch

Smartphone

Smartfood

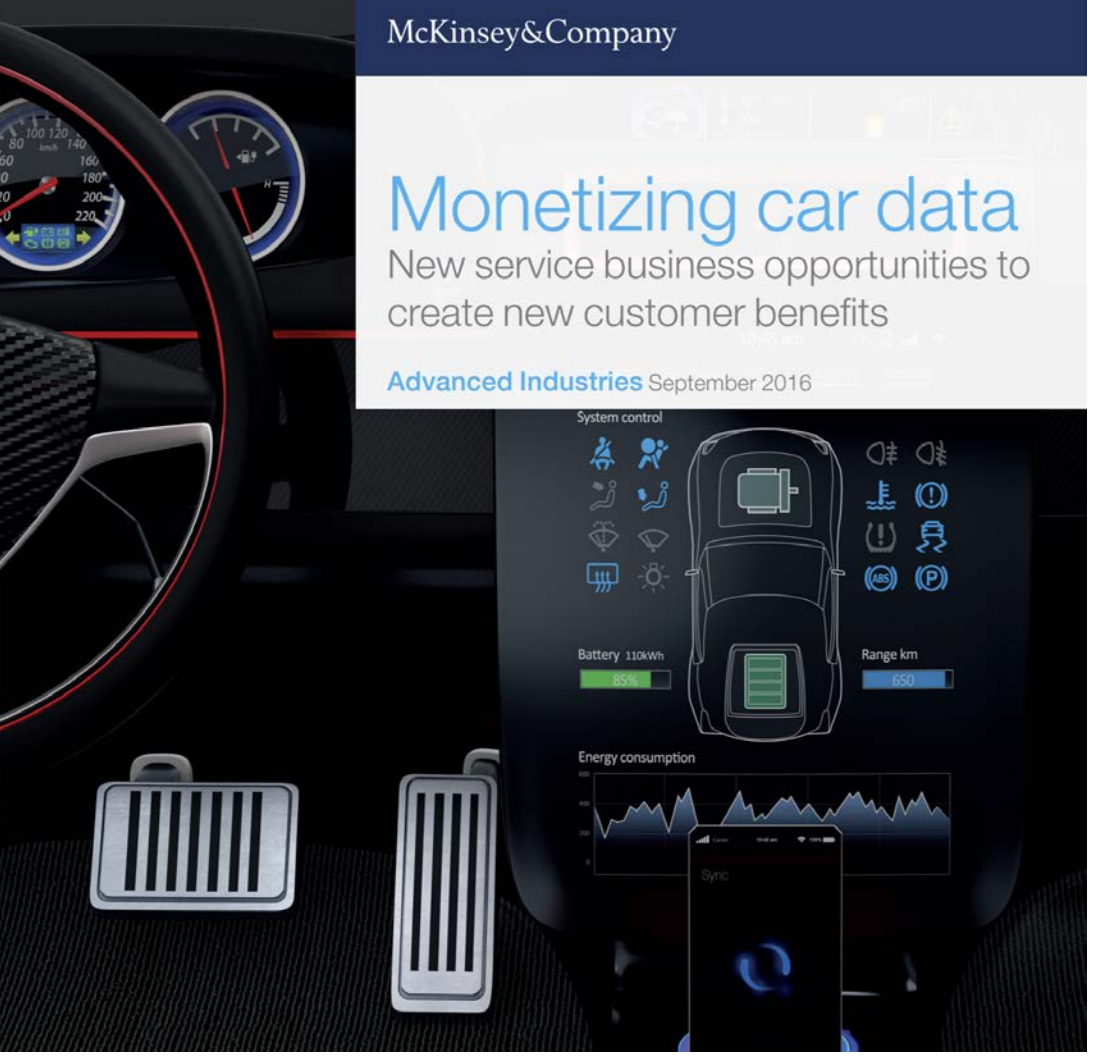
Smart Speaker
(Digital Assistant)

Smart autonomous
vehicle

Smart Home

Smart Kids

usignite



Harvard Journal of Law & Technology
Volume 32, Number 1 Fall 2018

WHO OWNS THE DATA GENERATED BY YOUR SMART CAR?

*Sylvia Zhang**

Collects about 1 TB/h of operation

usignite

TODAY

Edit

13,389
Steps

142	6.49	2,763	218
Floors	Miles	Cals	Zone Mins

Track your mindfulness
3 days left this week

7 hr 24 min
86 Sleep Score

5 of 5 days
You did it!

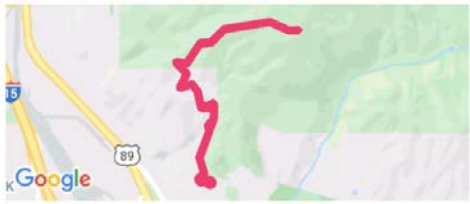
Oct 4 - Oct 10 — 5 hrs 49 mins

Run
Oct 10 at 09:46

5.24 mi — 16'26" pace
120 avg bpm — 143 zone min

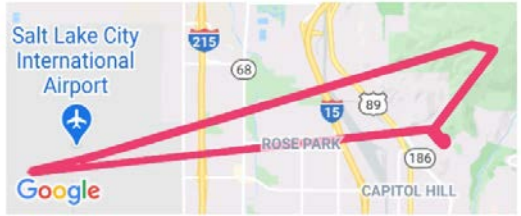
Outdoor Bike
Oct 9 at 18:28
32 min — 170 cals
98 avg bpm — 23 zone min
71% fat burn

Run
Oct 9 at 07:15



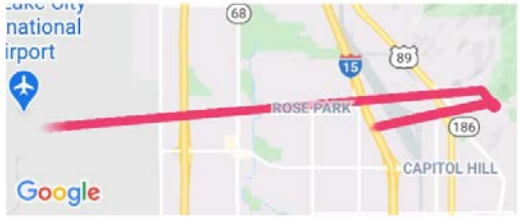
5.24 mi — 16'44" pace
122 avg bpm — 163 zone min

Run
Today at 16:27



5.21 mi — 19'47" pace
119 avg bpm — 175 zone min

Run
Wed at 15:23



4.94 mi — 22'30" pace
112 avg bpm — 151 zone min

usignite

Amazon Discusses Ways That It's Monetizing Alexa

At Amazon's re:MARS conference, the company's VP of Alexa gives a glimpse into the revenue potential of its voice assistant.

ANNIE GAUS · JUN 6, 2019 9:46 AM EDT

Data Drives Smart Cities

Growth of Smart Services

Smart and Connected Community Applications Require Low-Latency, High-Bandwidth

Smart and connected communities

Smart and connected autonomous vehicles
Smart health: e.g., continuous health monitoring
Smart digital assistants
Just-in-time education
Built-in cybersecurity
Smart workspaces

Growth of Smart Services

Smart and Connected Community Applications Require Low-Latency, High-Bandwidth

Bandwidth needed for:

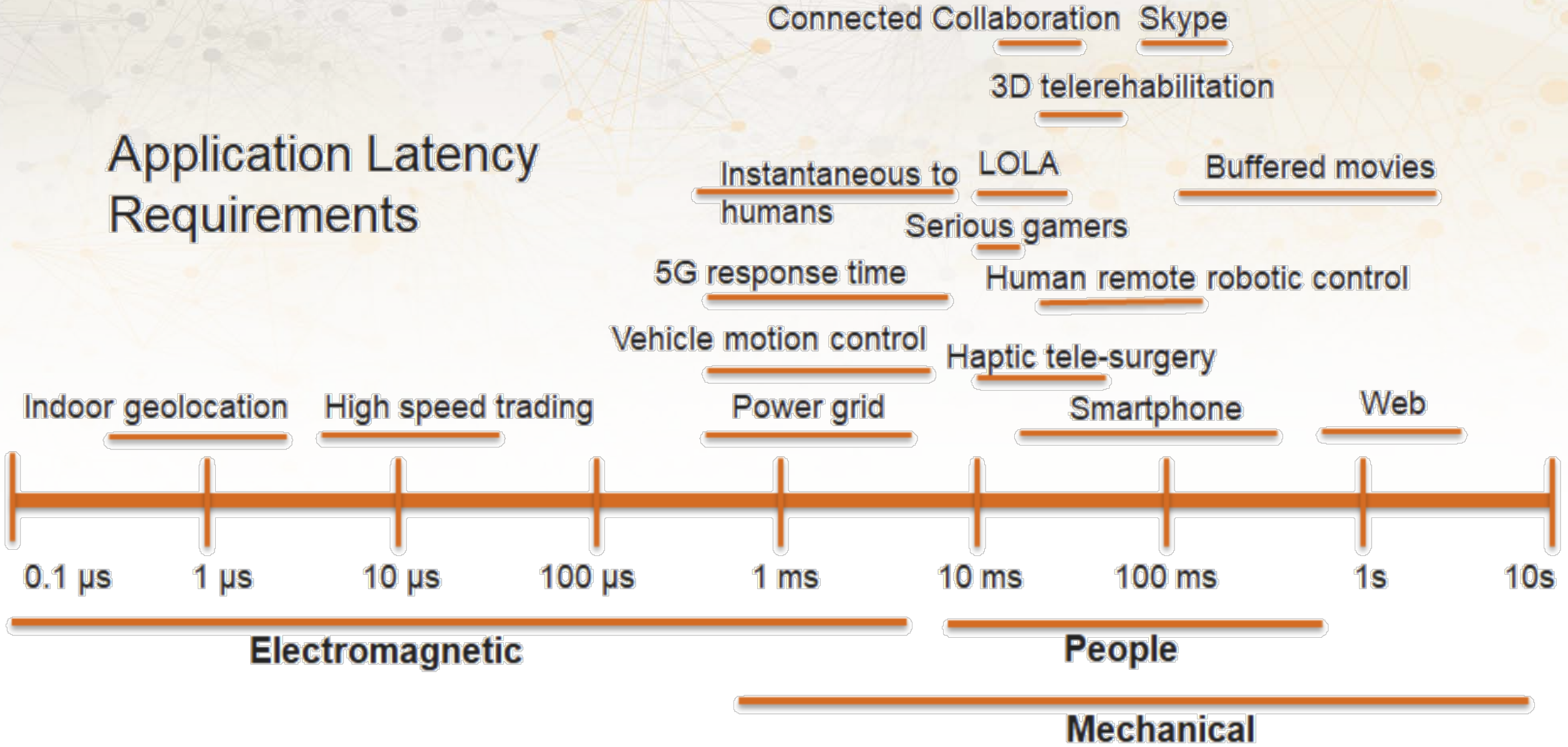
AR/VR, AI, big data

Latency converging on:

- Human response times
- Mechanical / chemical response times
- Electromagnetic response times
- Speed-of-light response times

usignite

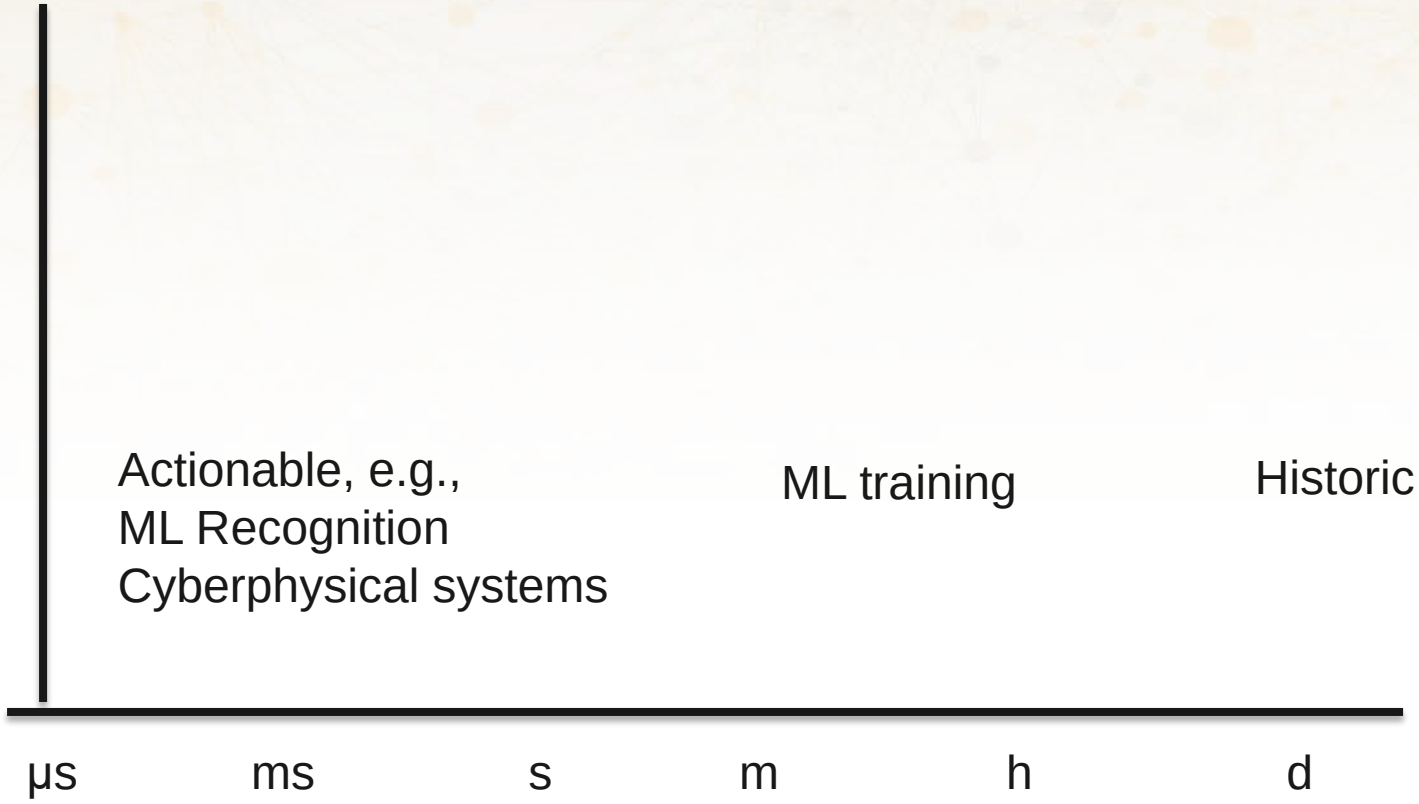
Application Latency Requirements



usignite

Perishable Data from sensors or to actuators

Value in immediately actionable decision-making decreases over time



usignite

Power Microgrid Coordination

Milliseconds count

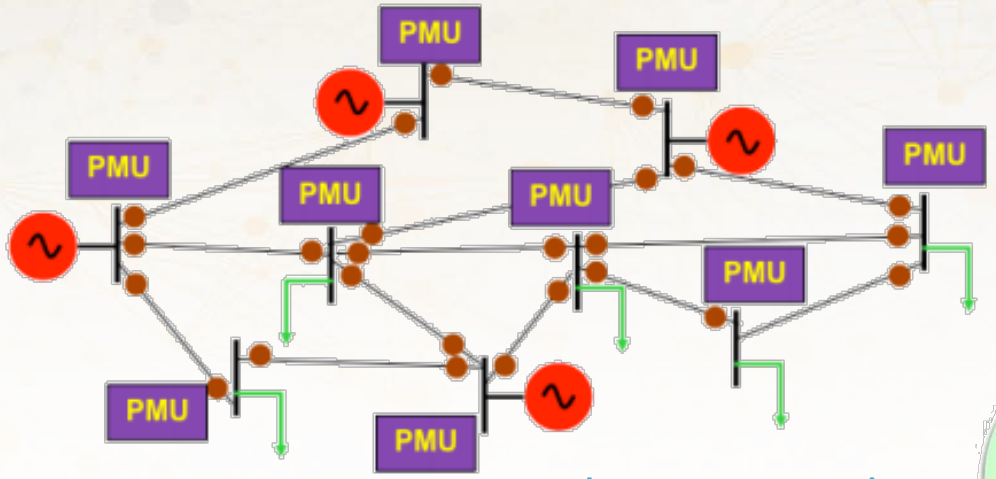


Image: Husham I. Hussein

Time accuracy: 32 µs
React in < 10 ms

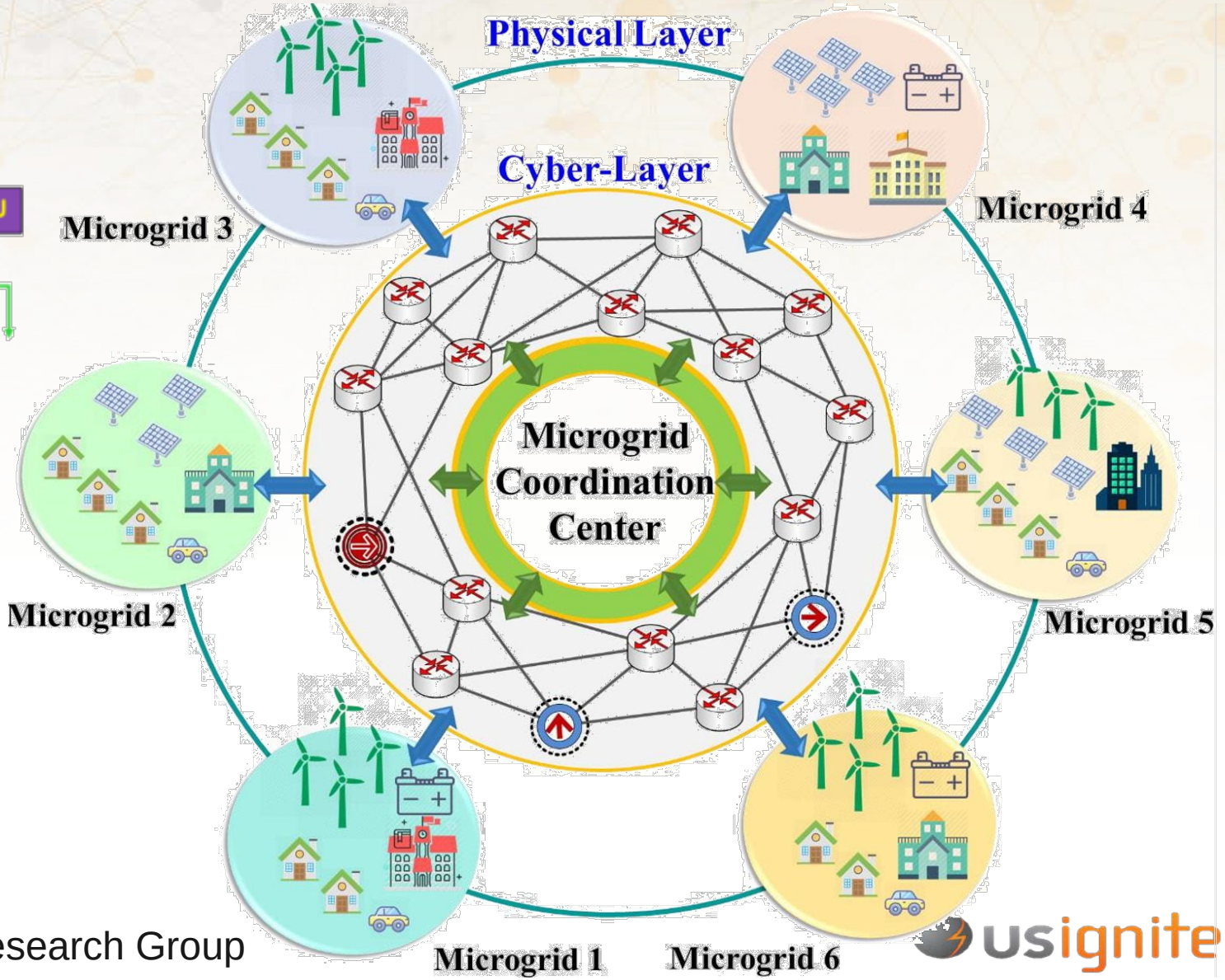


Image: Penn State EMPower Research Group

usignite

Connected Vehicle Management

Milliseconds count

Platooning Vehicles

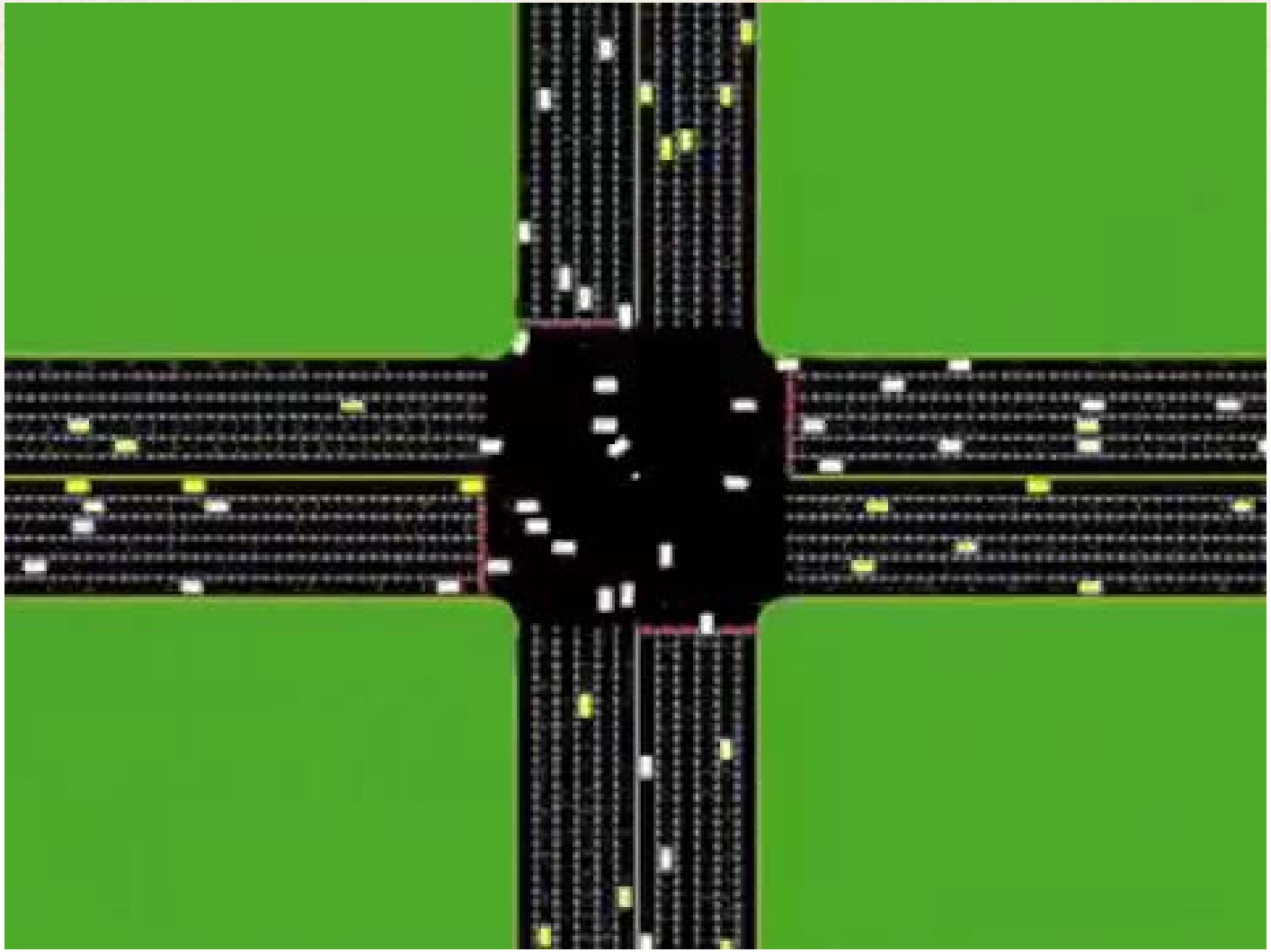
Time accuracy: 1 ms
React in < 10 ms

usignite

Connected Vehicle Management

Milliseconds count

Intersection sharing



Time accuracy: 1 ms
React in < 10 ms

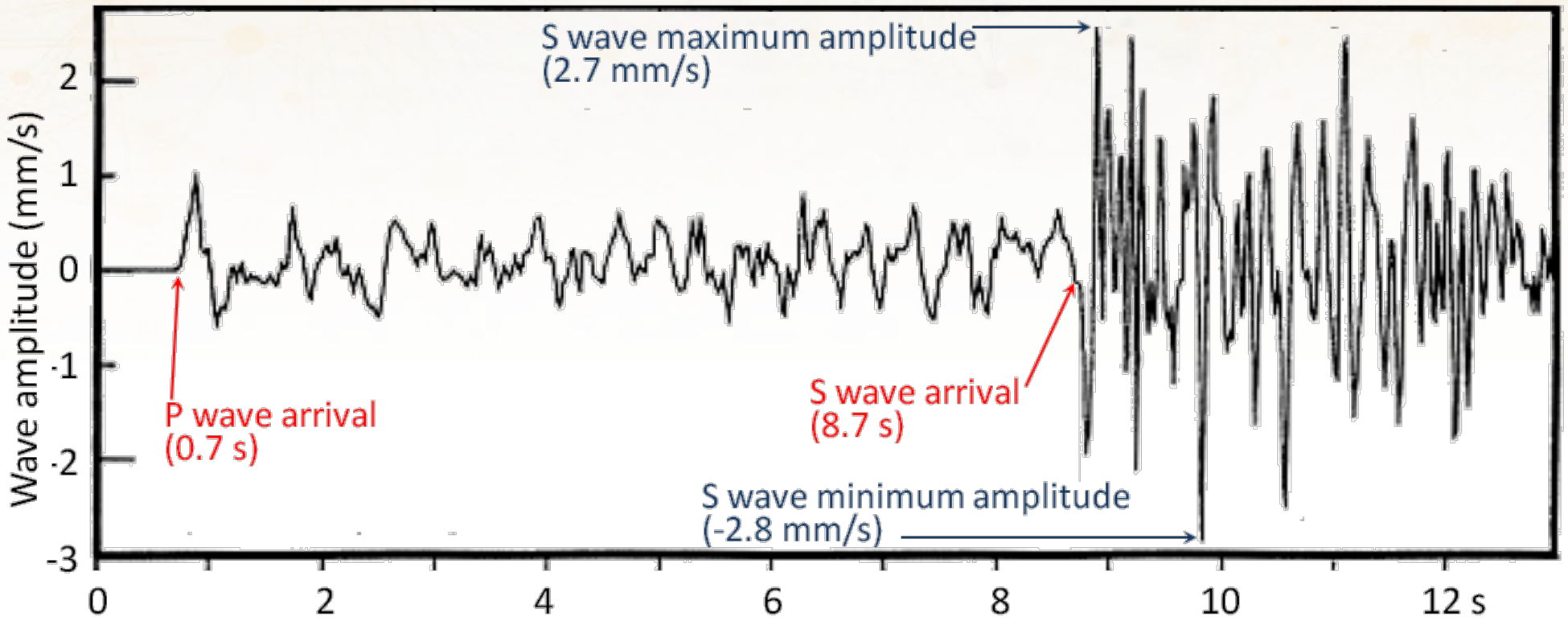
Dresner, Kurt, and Peter Stone. "A multiagent approach to autonomous intersection management." *Journal of artificial intelligence research* 31 (2008): 591-656.

usignite

Earthquake Prediction / Notification

Seconds count

P-wave precedes damaging S-wave



Time accuracy: 0.5 s
React in < 0.5 s

Smart Cities Data Catalog Specification

ATIS-I-0000080 | August 2020

atis usignite

usignite

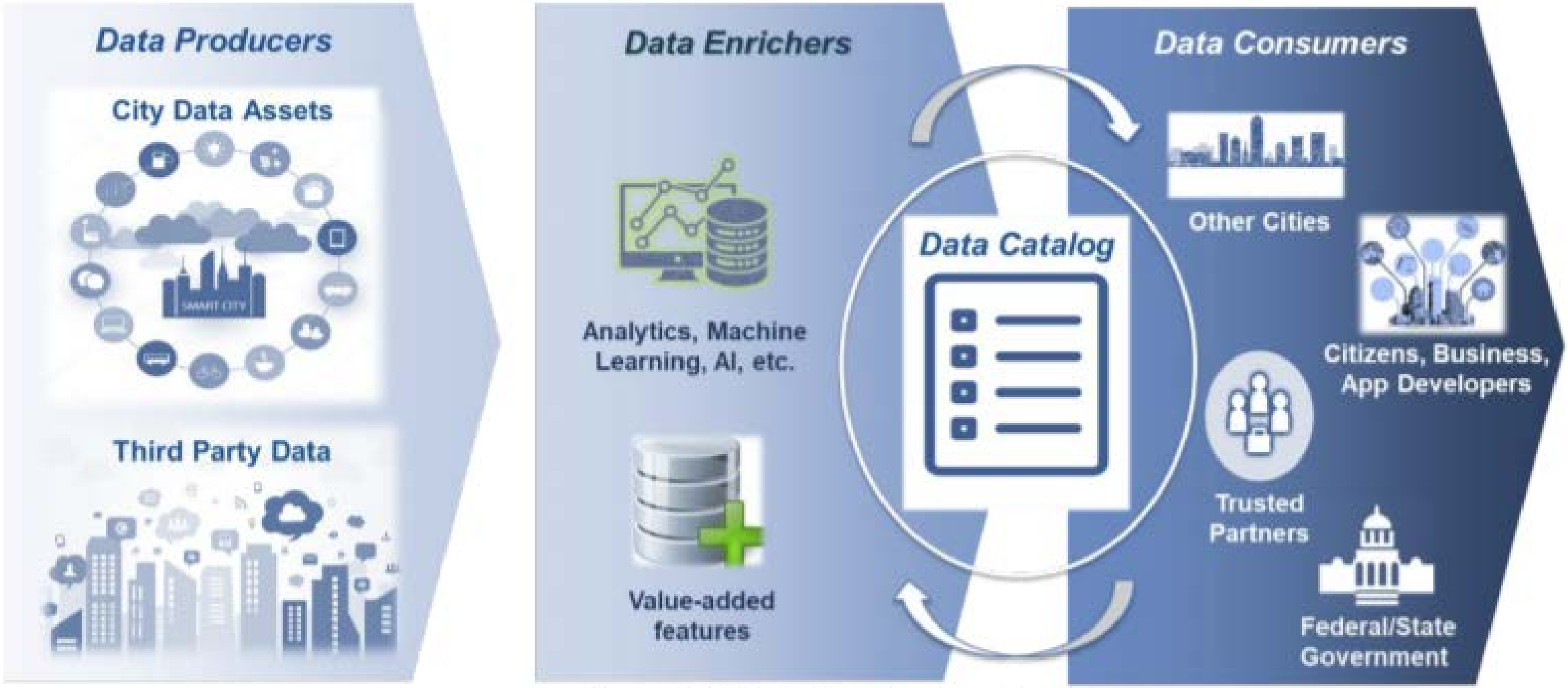


Figure 4.1 Data Catalog Domains

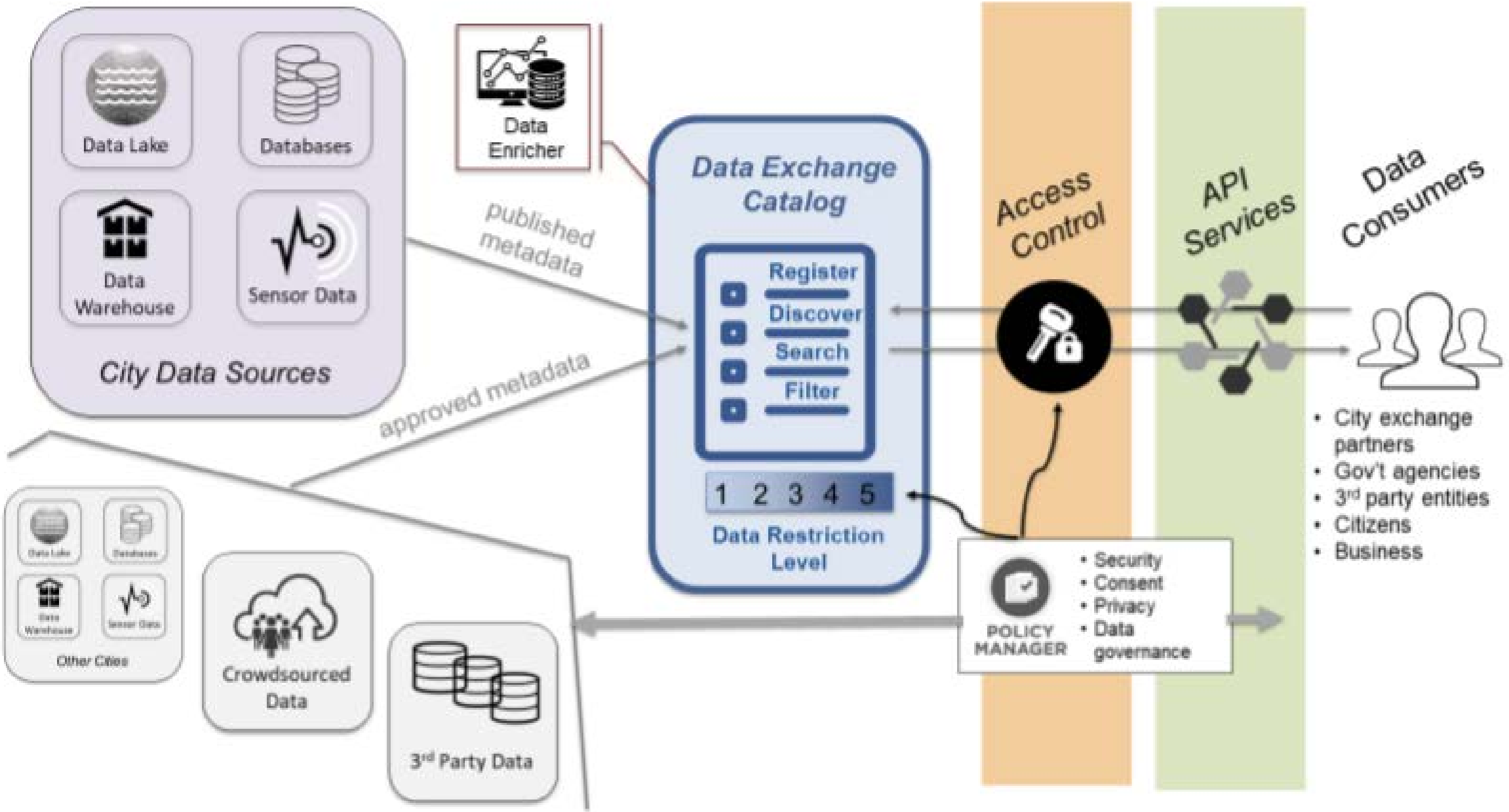
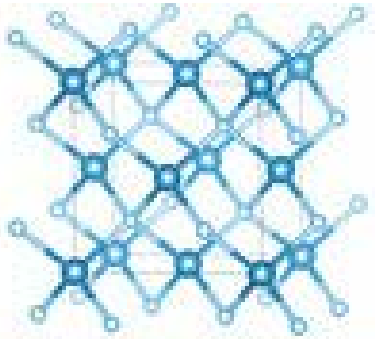


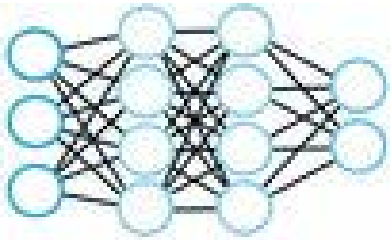
Figure 3.1 Data Catalog Architecture

Data Drives Smart Research and Learning

(See George Turner’s session earlier today)

$$\nabla \cdot \mathbf{D} = \rho$$
$$\nabla \cdot \mathbf{B} = 0$$
$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$
$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}$$





1st paradigm: Empirical science

Experiments

2nd paradigm: Theoretical science

Physical Laws: mechanics, thermodynamics, etc

3rd paradigm: Computational science (simulations)

Density Functional Theory, Molecular dynamics, etc

4th paradigm: (Big) Data-driven Science

Artificial intelligence, Statistical learning, Data mining, Pattern and anomaly detection, etc

1600 1950 2000

usignite

Big Research Data: Square Kilometre Array

South Africa and Western Australia

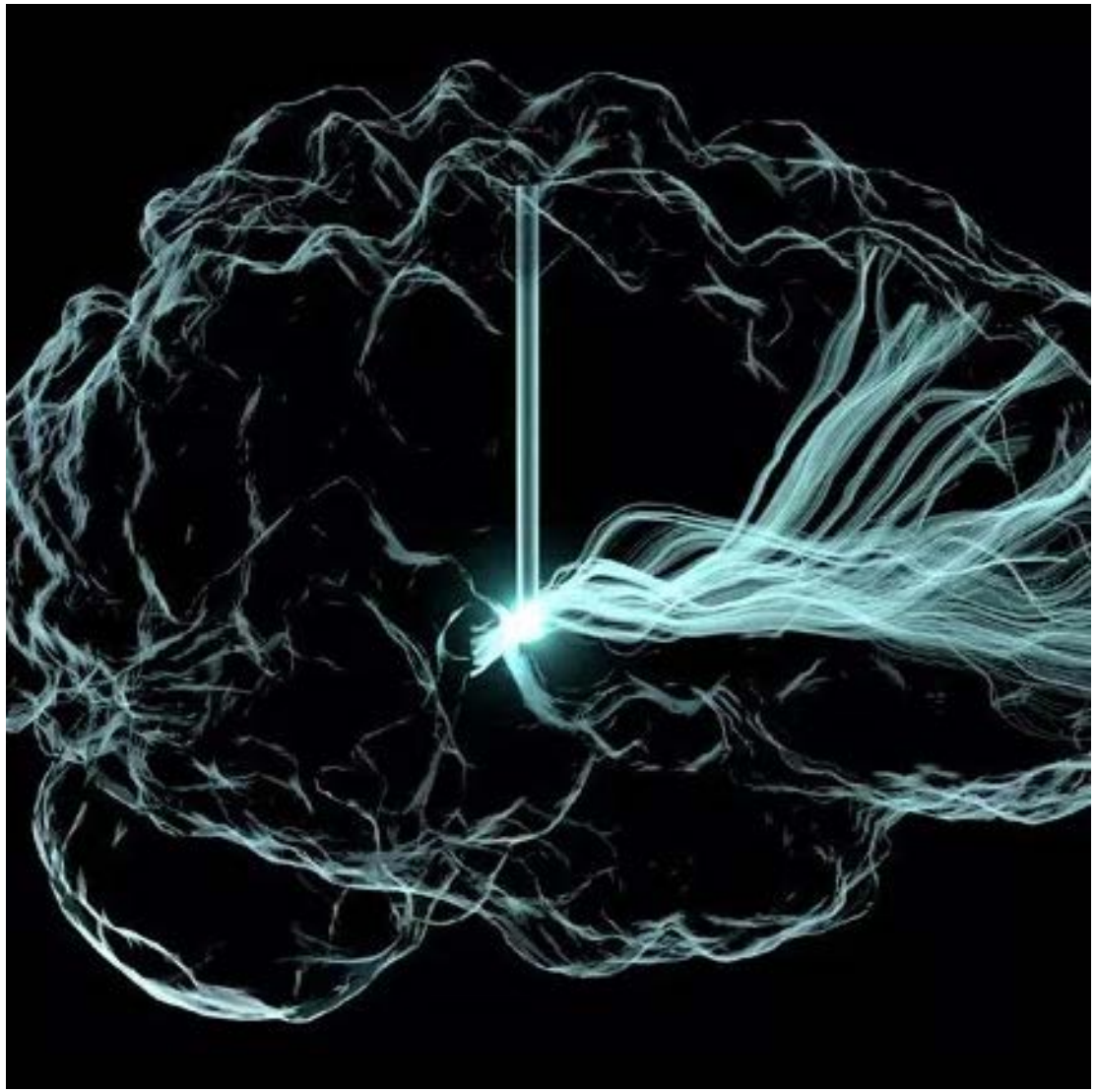
132,000 antennas

generating157 TB/s

= about twice estimated global Internet traffic

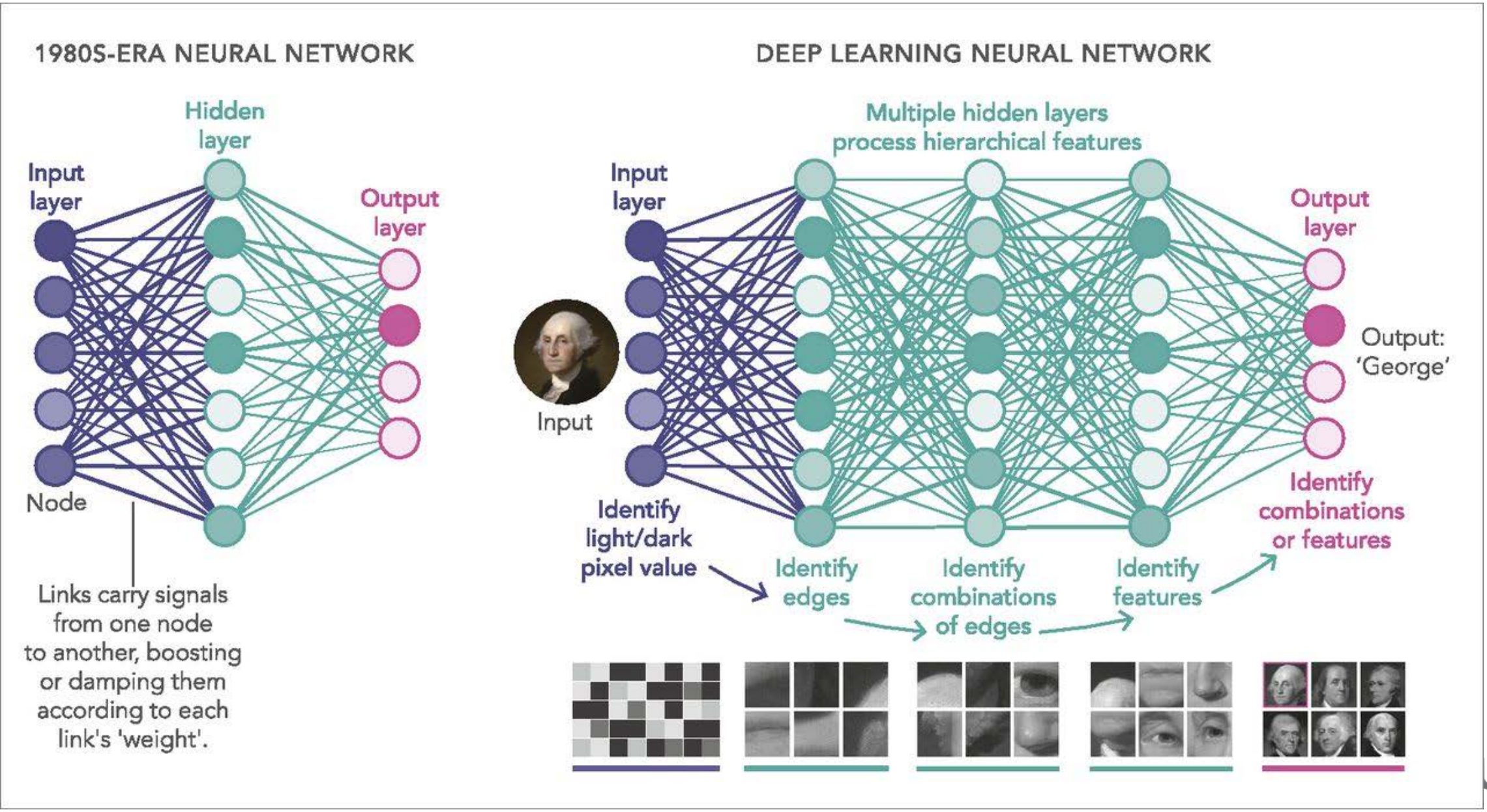
Deep Brain Stimulation Tuning

SCI Institute, U. Utah
US Ignite
National Science Foundation



Data Drives Artificial Intelligence (Machine Learning)

(See Erin McCabe's session earlier today)


1980S-ERA NEURAL NETWORK
Hidden layer
Input layer
Output layer
Node
Links carry signals from one node to another, boosting or damping them according to each link's 'weight'.
DEEP LEARNING NEURAL NETWORK
Multiple hidden layers process hierarchical features
Input layer
Output layer
Input
Output: 'George'
Identify light/dark pixel value
Identify edges
Identify combinations of edges
Identify features
Identify combinations or features

The data-driven world will be **always on**, **always tracking**, **always monitoring**, **always listening** and **always watching** – **because it will be always learning**.

usignite

Things Can Go wRONG

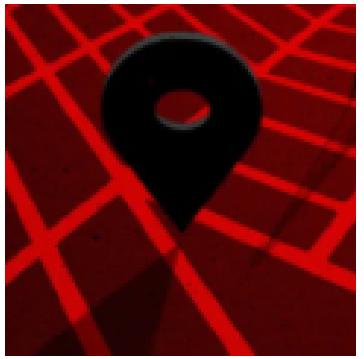
An IP PIN is a six-digit number assigned to you as a taxpayer that helps prevent the misuse of your SSN on fraudulent federal income tax returns.

1 2 3 4 5 6

usignite

Tech

The IRS Is Being Investigated for Using Location Data Without a Warrant



JOSEPH COX

10.06.20

Data's Ability to Carry Meaning

Concepts – Reasoning – Theories – Postulates – Musings - Misleading

Emotions – Feelings – Support – Arguments – Manipulation

Discourse – Debate – Name-calling – Slander

Comprehensive – Tailored – Selected – Taken out of context

Data’s Ability to Describe and Measure

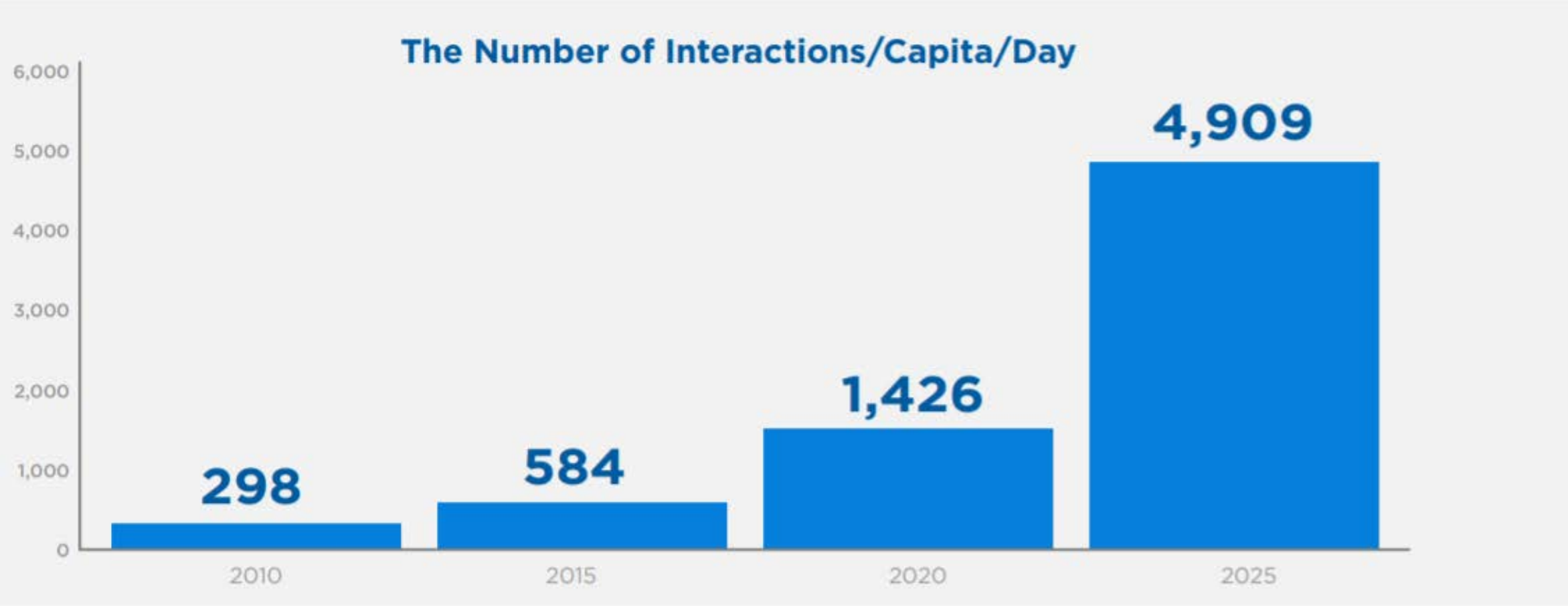
Matches – Closely describes – Proxy for - Correlates with – Related to

Secure – Vulnerable – Forgotten – Lost – Stolen – Kidnapped

Appropriately – Misunderstood – Inappropriately – Intentionally Misused

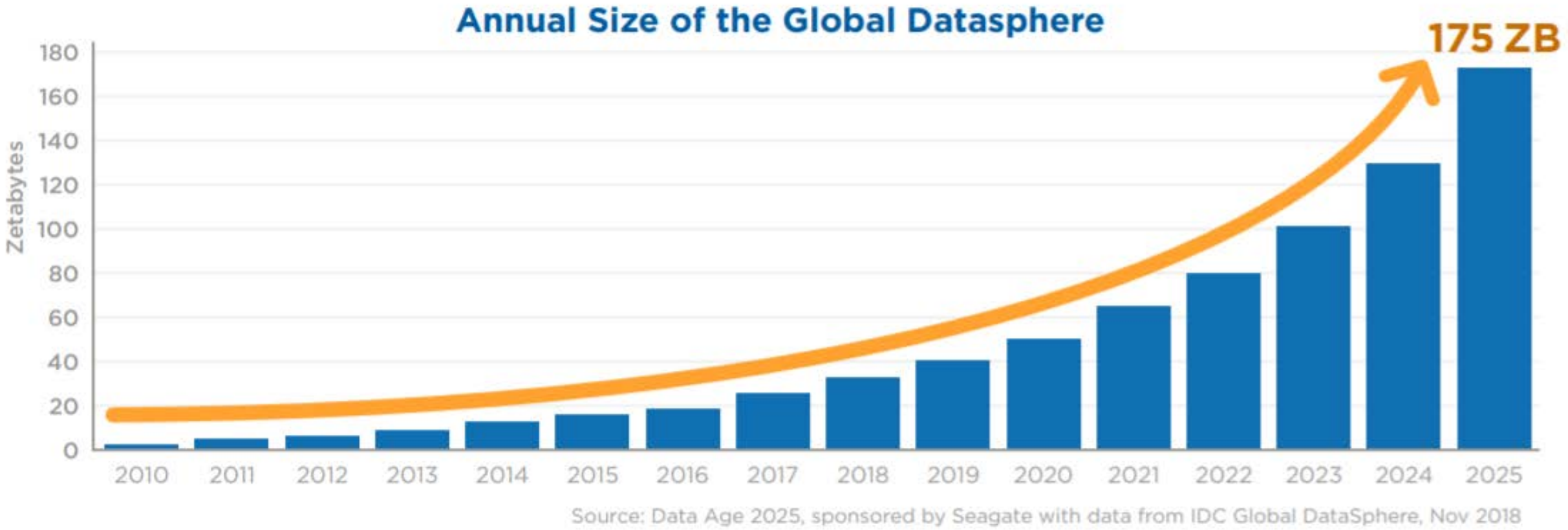
Clean - Inaccurate – Outdated – Corrupted – Wrong

Data Drives Smart Economies

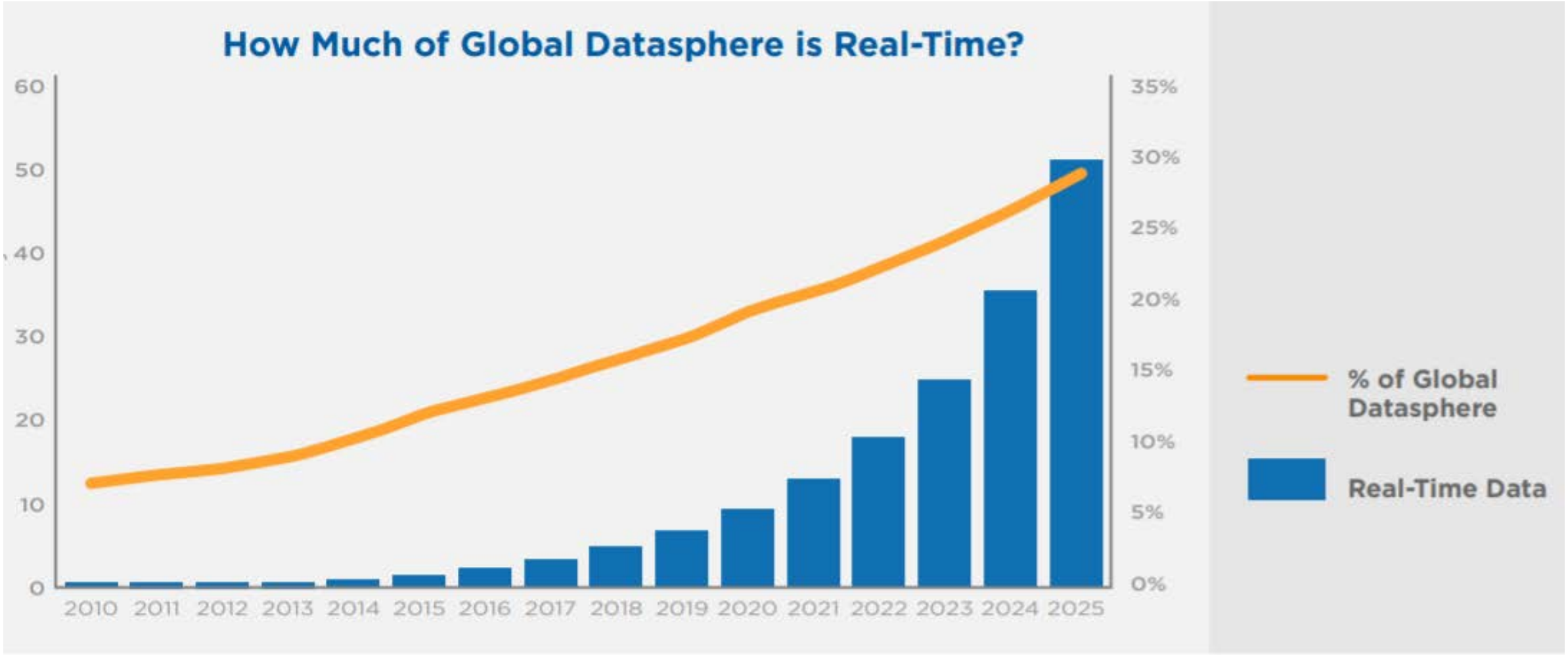


Source: Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018

usignite


Annual Size of the Global Datasphere
175 ZB
Zetabytes
180
160
140
120
100
80
60
40
20
0
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
Source: Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018

usignite


How Much of Global Datasphere is Real-Time?
60
50
40
30
20
10
0
35%
30%
25%
20%
15%
10%
5%
0%
2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025
% of Global Datasphere
Real-Time Data

usignite

American companies alone are estimated to have spent over $19 billion in 2018 acquiring and analyzing consumer data, according to the Interactive Advertising Bureau.

Source: WIRED Magazine, The WIRED Guide to Your Personal Data (and Who is Using It), 2/15/2019.

The data **broker** industry is believed to be **worth** about $200 billion.

Source: LA Times, David Lazarus, Nov. 5, 2019, Shadowy data brokers make the most of their invisibility cloak.

usignite

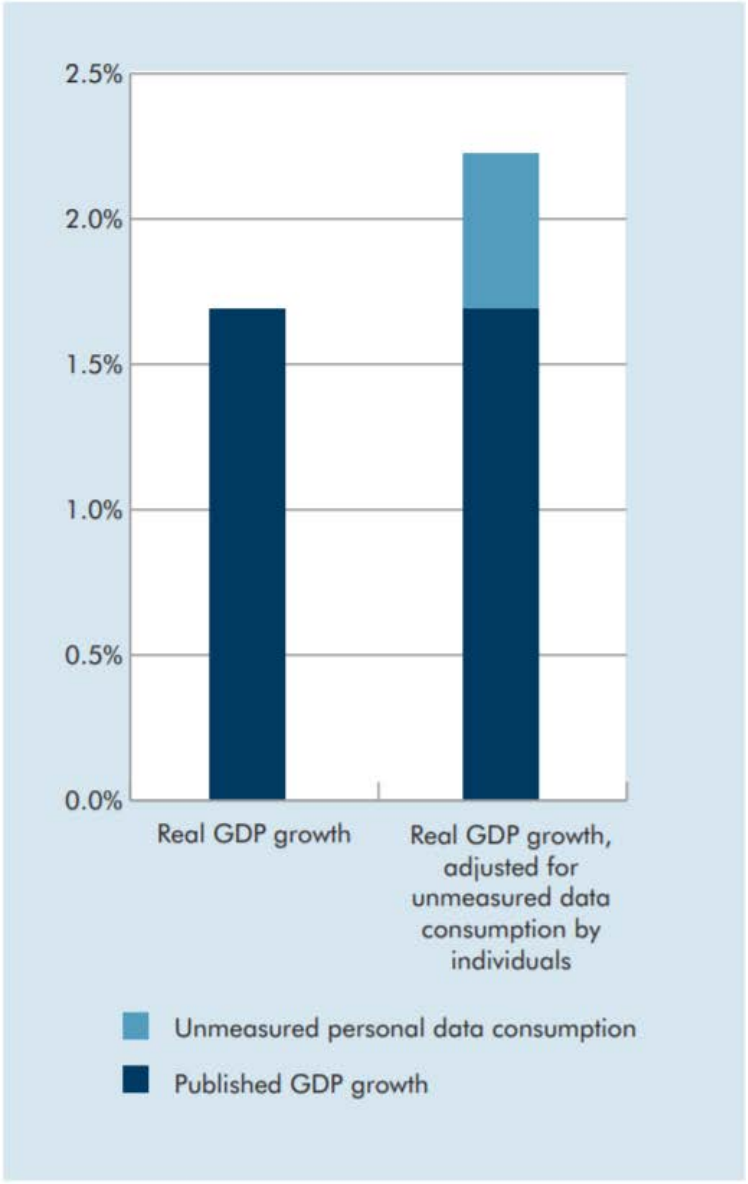
GDDP Per Capita (in Terabytes)

How much data exchange a single person creates for the data economy?

Chart: Data News • Source: GDP data from World Bank, size of data economy data from IDC. • Created with **Datawrapper**

usignite

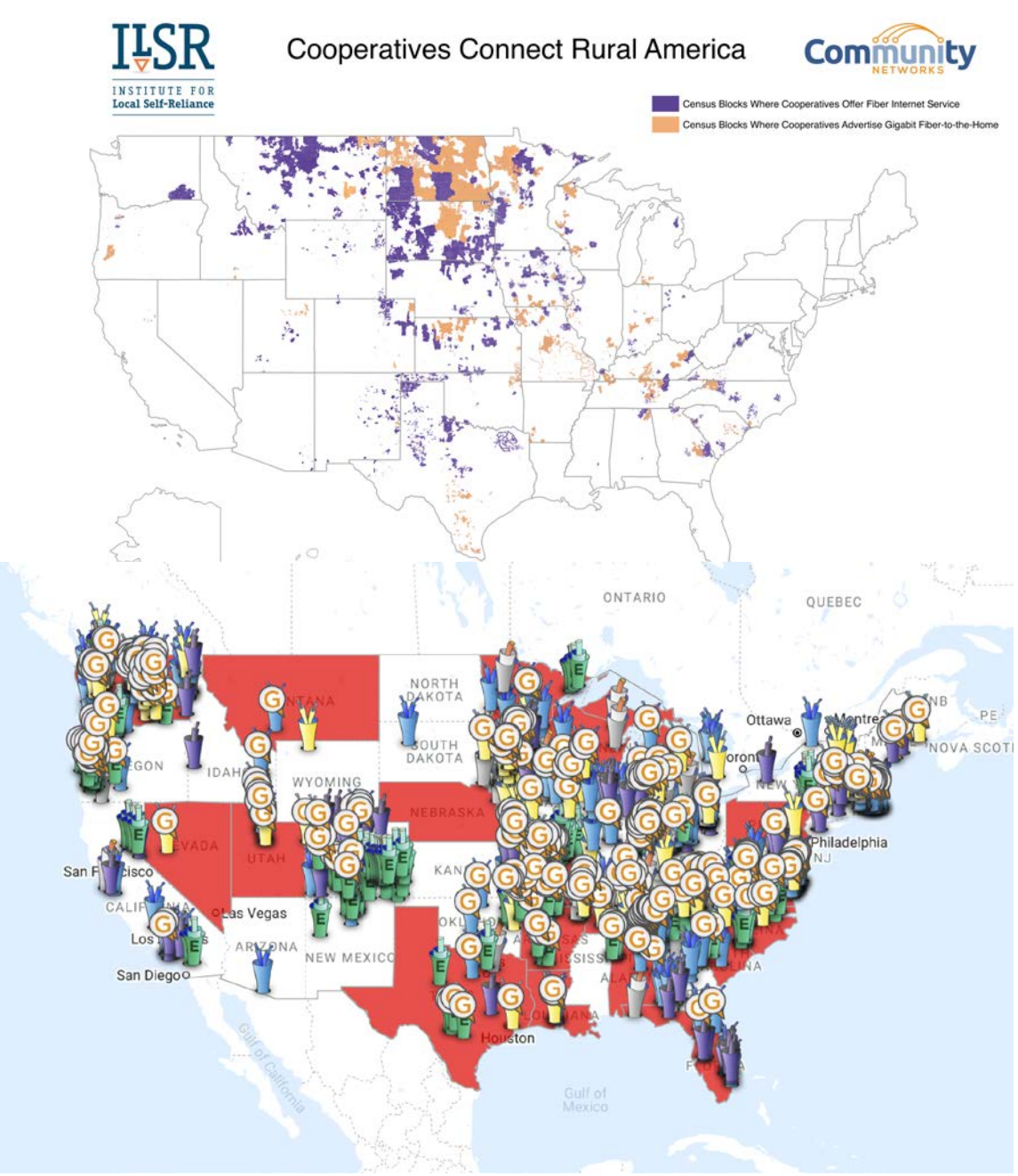
FIGURE 1: REAL U.S. GDP GROWTH, WITH
AND WITHOUT UNMEASURED DATA CONSUMPTION
(FIRST TWO QUARTERS OF 2012, ANNUAL RATES)

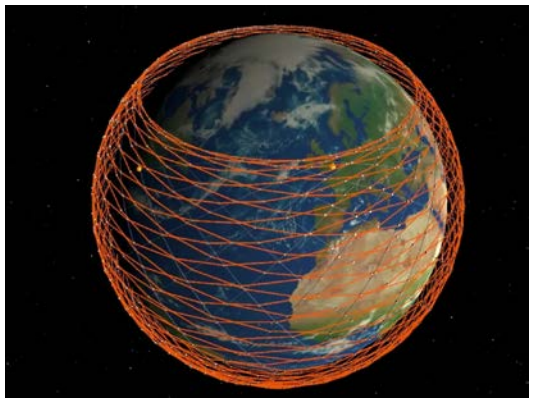


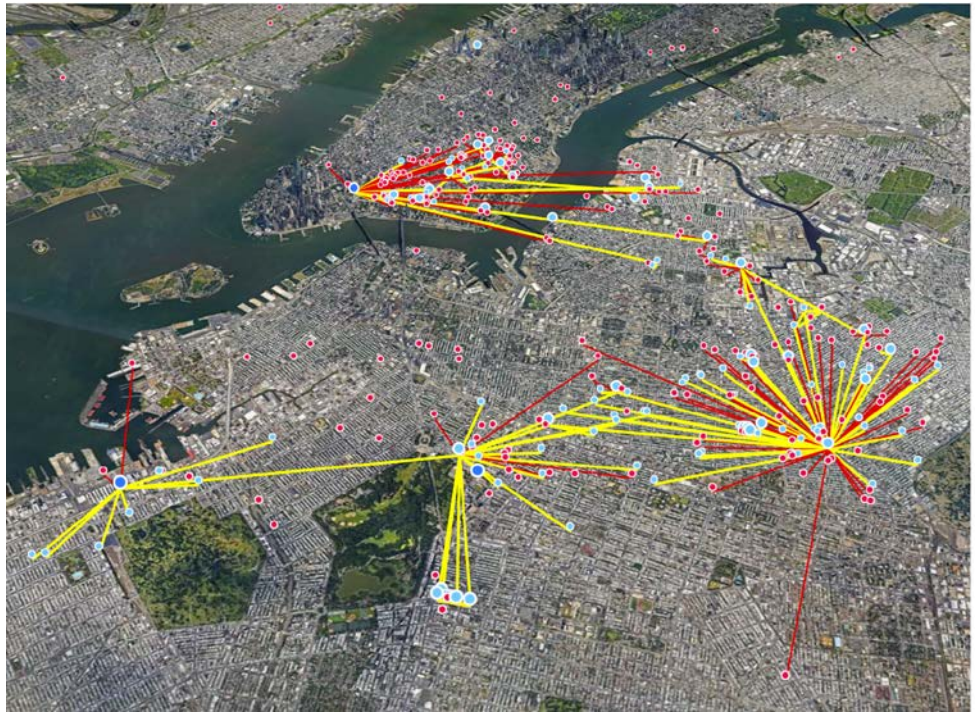
Data: Bureau of Economic Analysis,
Progressive Policy Institute (data as of 8/10/12)

Project Overcome

- **Creative** ways of connecting the unserved and underserved
  - Fiber self-install kits, TV White space, CBRS, etc.
- Five **pilot** communities; one year timeline
- **Community-based organizations** within each community
- Looking for measurable **social impact**
- Working in league with existing organizations in this space


ILSR
INSTITUTE FOR
Local Self-Reliance
Cooperatives Connect Rural America
Community
NETWORKS
Census Blocks Where Cooperatives Offer Fiber Internet Service
Census Blocks Where Cooperatives Advertise Gigabit Fiber-to-the-Home
ONTARIO
QUEBEC
NORTH DAKOTA
SOUTH DAKOTA
WYOMING
NEBRASKA
IDAHO
UTAH
NEVADA
ARIZONA
NEW MEXICO
Las Vegas
San Diego
Houston
Philadelphia
Ottawa
Gulf of California
Gulf of Mexico





usignite

Data Drives a Smart You

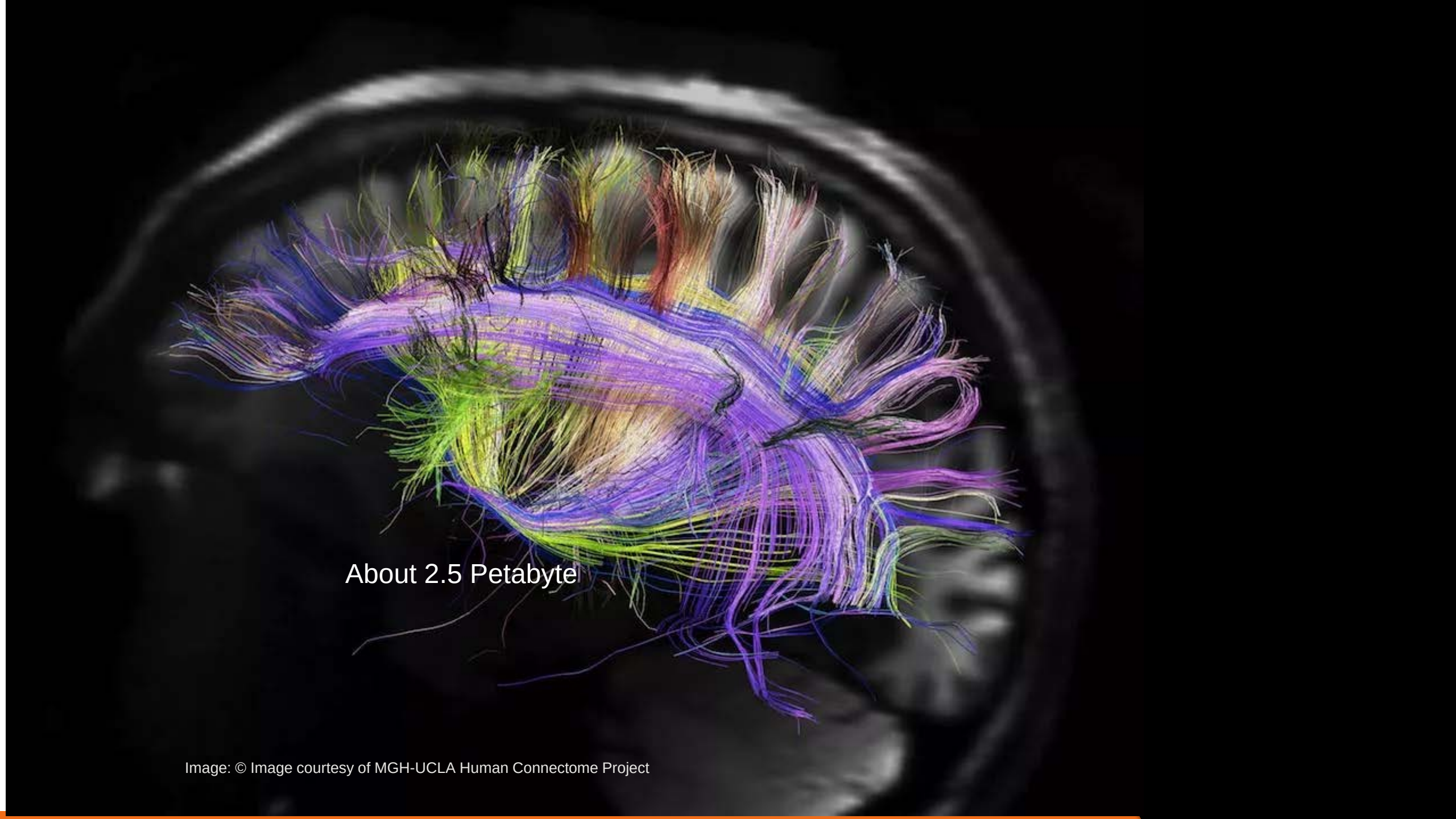


Image: © Image courtesy of MGH-UCLA Human Connectome Project