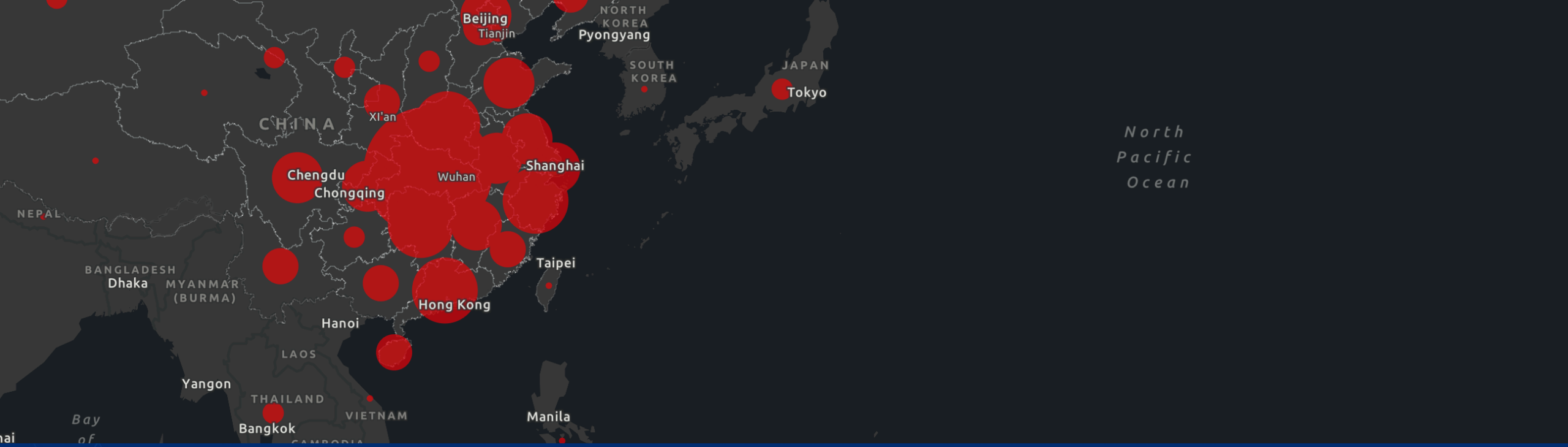
JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

Open Data Operations for COVID-19

October 7, 2020

Aaron Katz

Large Scale Analytics Systems Group
Asymmetric Operations Sector
Johns Hopkins University Applied Physics Laboratory



JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

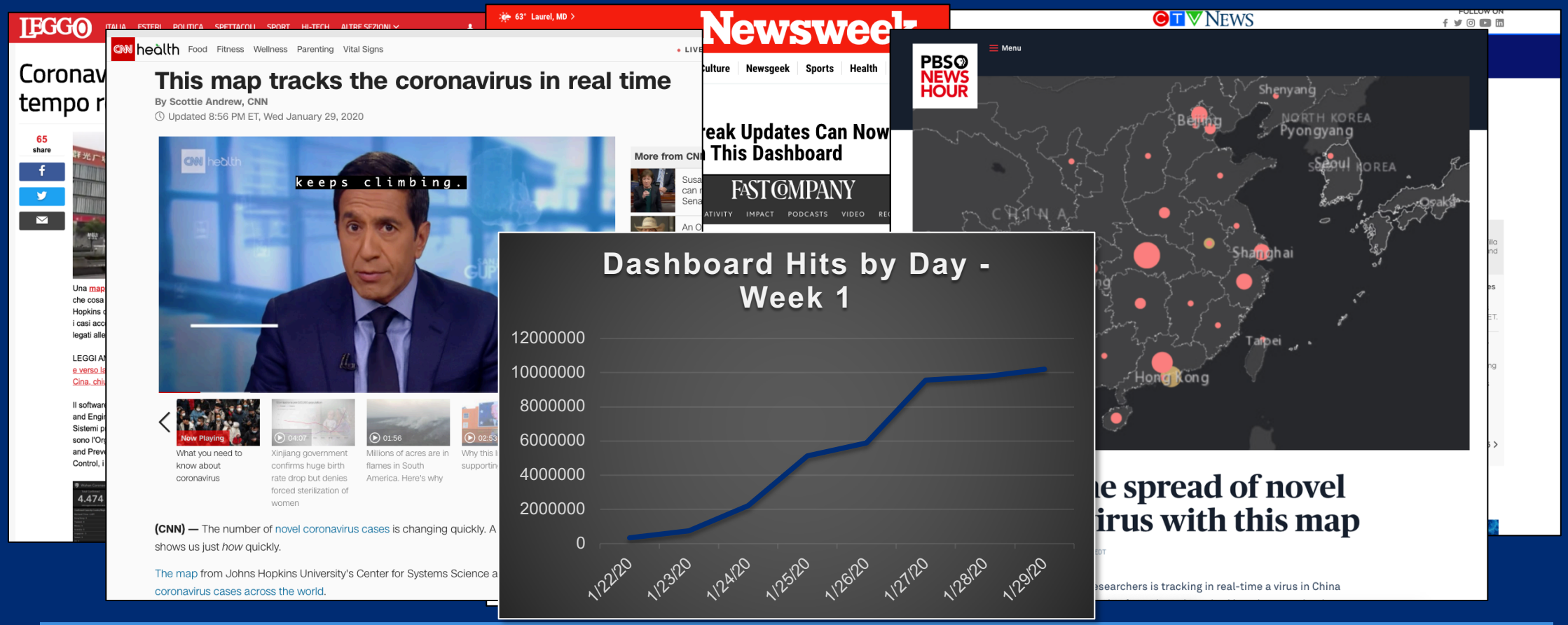
A Brief History

In the Beginning

It started with coffee

And ended with a Tweet heard round the world

Week 1



The Dashboard was rapidly picked up by US and international press

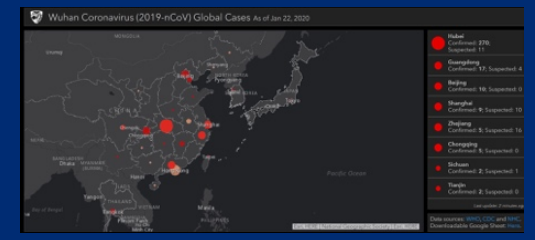
Early Data Operations

“Authoritative” Data Sources

Crowd Sourced Data

Manual Data Curation

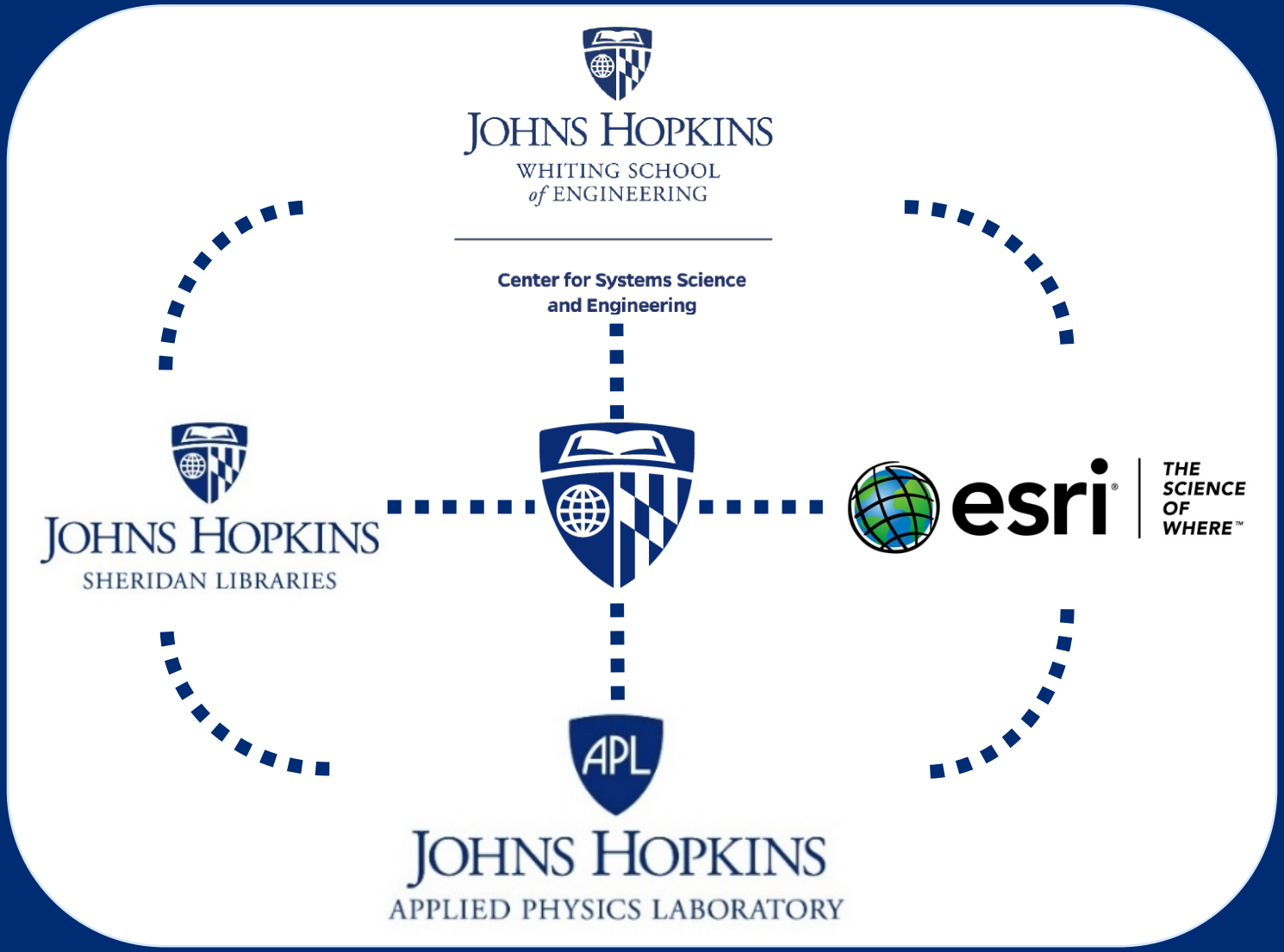
Google Sheets



Manual Dashboard & Data Push

Fully manual process
Not scalable – Limited Sharing

The Early Partnership



JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

A Quick Detour:

PAI Landscape Briefing
Feb 3, 2020

The PAI Landscape Feb 2020

Open Source Surveillance

- **Survive**
  - Help Lauren Garnder and Frank Dong survive the data surge:
    - Help ID trusted data sources
    - Help Triage recommendations and comments
    - Identify opportunities for process automation and prototype solutions
  - Help improve information sharing with key partners
    - Backend augmented to support
  - Other duties as assigned
- **Thrive**
  - Help get surveillance picture closer to real-time while preserving quality
  - Create a configurable infrastructure that allows us to rapidly stand-up capability in any crisis

The PAI Landscape Feb 2020

Disinformation Shapes the Battlespace

Julia Carrie Wong @juliacarriew · 7h

As coronavirus misinformation spreads, the press points the finger a
social media and trolls. But professional media outlets have been a
conduit of disinformation.

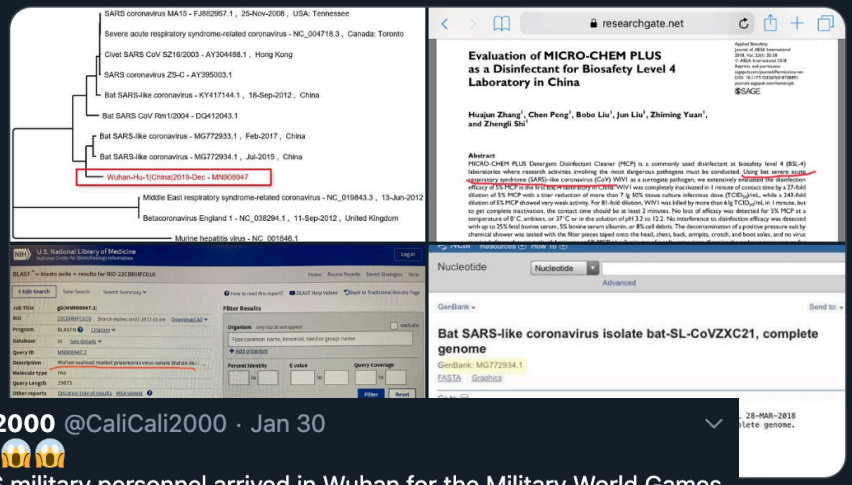
That racist "**bat soup**" canard? Thank British newspapers for sprea
around.

CNN @CNN · Jan 28

There's been a run of **surgical masks** in the US because of the **coronavirus** scare. You don't need them, physicians say. cnn.it/2GzpKyM

Chris Cyrek @SquarejawCity · Jan 27

#Wuhan BSL-4 Research Laboratory was using a bat-derived SARS-like **#coronavirus** to test it's disinfectant procedures in 2018. Genetic sequencing of 2019 Novel CoV suggests the current virus is extremely similar. Proximity of the Research Lab to outbreak origin deeply troubling.



CaliCali2000 @CaliCali2000 · Jan 30

" 300 US military personnel arrived in Wuhan for the Military World Games on October 19. The first **coronavirus** case appeared two weeks later, on November 2. **Coronavirus** incubation period is 14 days."

#China #bioweapon

Coronavirus: the Dark Side

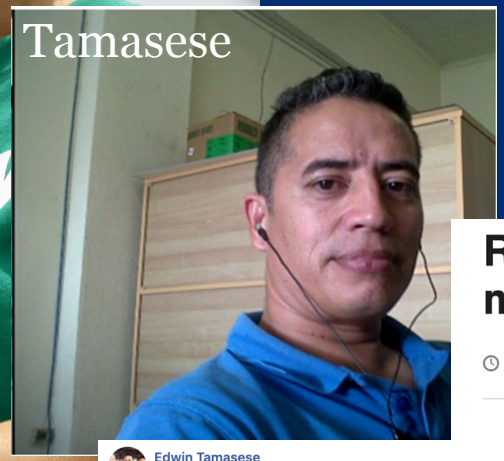
Vioxx killed 500,000 Americans: a toll that could have been reduced by 90% had the FDA issued a ...

unz.com

26 99 180

The PAI Landscape Feb 2020

This has Real Consequences



New York's Orthodox Jewish community is battling measles outbreaks. Vaccine deniers are to blame.

Rockland County and New York City have declared

Tiny Samoa has had nearly 5,000 measles cases. Here's how it got so bad.

This island nation saw 70 measles deaths so far this year, most in children under age 5.

By Julia Belluz | @juliaoftoronto | julia.belluz@voxmedia.com | Dec 18, 2019, 9:20am EST

Russia trolls 'spreading vaccination misinformation' to create discord

24 August 2018 Share

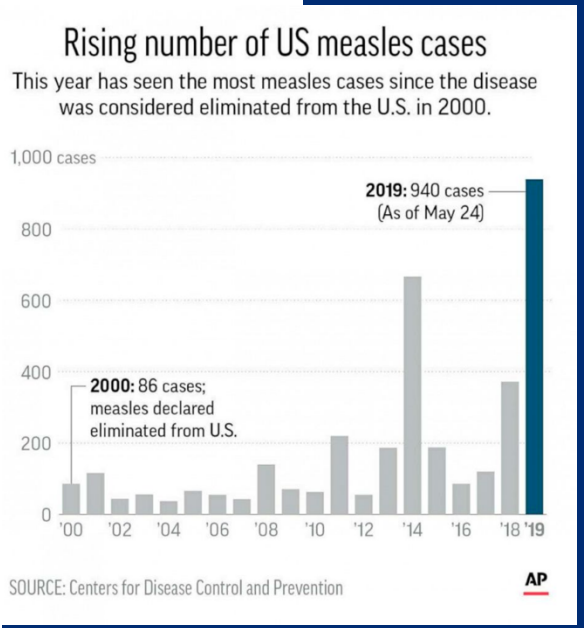
Edwin Tamasese
December 4, 2019 ·

Just before my phone gets taken 1000mg vit C sodium ascorbate taken dissolved in 1/4 cup water every 3 hours. This will save your kids. They should be on their feet in 16 hours. Then start vitamin A for repair of damaged tissue.

Save our kids. Nobody should have died. This is the greatest crime against our people by our own people.

God Bless Samoa. Please stop this madness.

3.3K 5.3K Comments 1.7K Shares



The PAI Landscape Feb 2020

Championing the Truth Can Change Behaviors and Save Lives

STOP FAKE.ORG

Борьба с неправдивой информацией о событиях в Украине

СООБЩИТЬ О ФЕЙКЕ

ПОДДЕРЖАТЬ ПРОЕКТ

Подписаться на рассылку

ГЛАВНАЯ О НАС МНЕНИЯ КОНТЕКСТ ВИДЕО ПОДКАСТ ИНСТРУМЕНТЫ ИССЛЕДОВАНИЯ ТЕСТ ГАЗЕТА

НОВОСТИ

Поиск... Искать

LENTA.RU

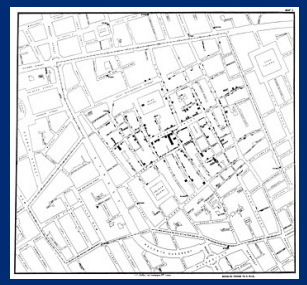
Хотите получать сообщения обо всех важных новостях в рубрике «Бывший СССР»?

На Украине завели дело против Обамы

ЯНВАРЬ 29, 2020 - 21:40 · НОВОСТИ

Фейк: В Украине завели дело пр… Обамы

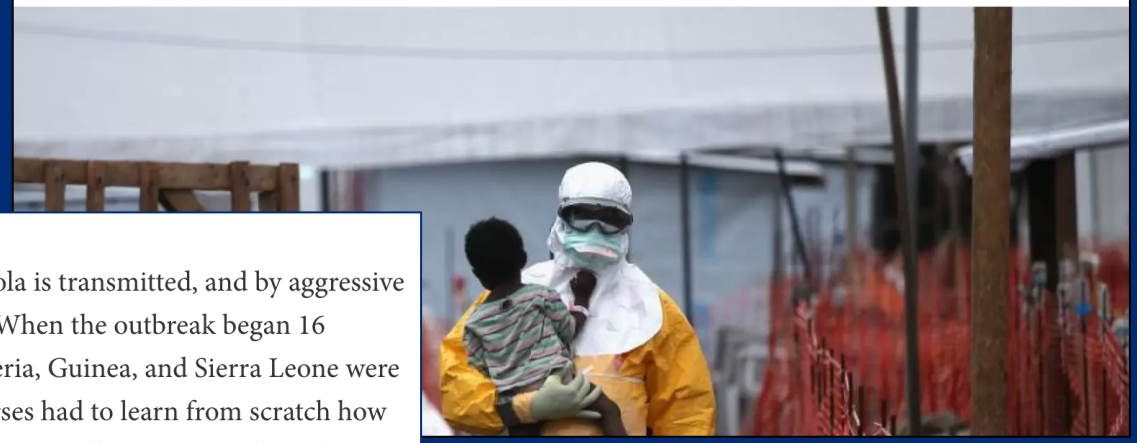
В том, что Национальное антикоррупционное бюро … начало расследование против Обамы, уверены российск… Соответствующая информация была опубликована 30… российскими изданиями Lenta.ru, Украина.ру и д… Первоисточником фейковой информации выступил … Кузьмин, бывший заместитель генпрокурора Украи… президента Янукович, а сегодня неровный депутат от …



THE WEEK

How we beat the Ebola epidemic

The Week ***Staff***



How was it contained?

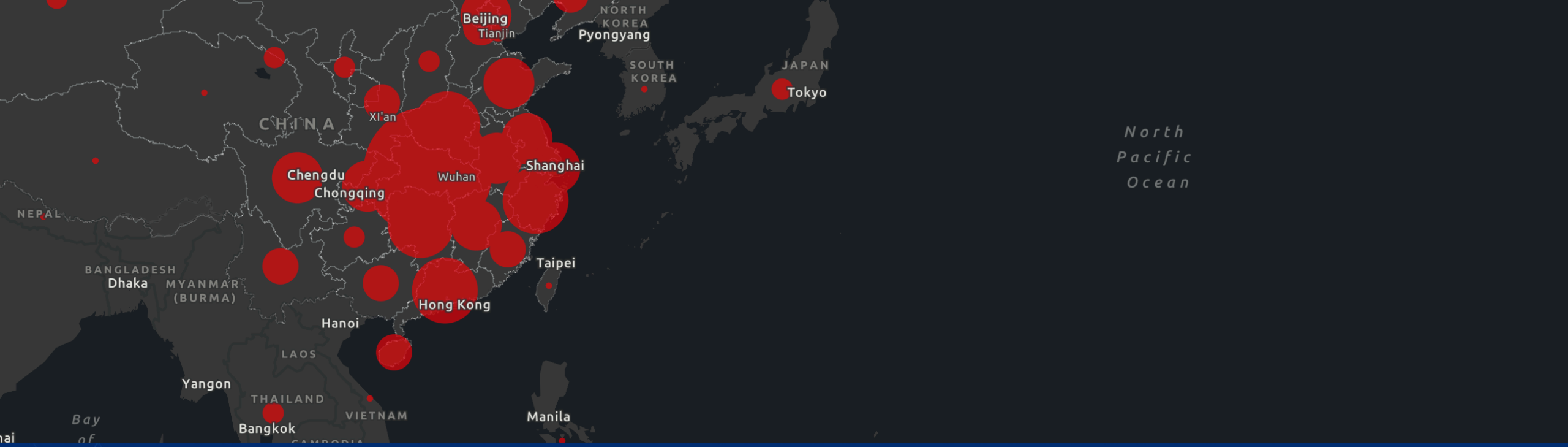
By educating West Africans on how Ebola is transmitted, and by aggressive efforts to isolate and treat the infected. When the outbreak began 16 months ago, health care workers in Liberia, Guinea, and Sierra Leone were initially fighting blind. Doctors and nurses had to learn from scratch how to treat Ebola — the odds of survival dramatically increase with early, effective symptom management — and how to contain it. Helped by more than 10,000 volunteers from around the world, health workers gradually taught people to avoid unnecessary physical contact, to go to a clinic the moment they displayed symptoms, and to forgo the traditional ritual of washing corpses — a practice that accelerated the spread of the disease. "The best way to fight Ebola," says Joseph Boye Cooper, a volunteer worker in Liberia, "is to prevent it."

Int J Environ Res Public Health. 2019 Jun; 16(11): 2020. PMCID: PMC6603859
Published online 2019 Jun 6. doi: 10.3390/ijerph16112020 PMID: 31174301

The Impact of Various Promotional Activities on Ebola Prevention Behaviors and Psychosocial Factors Predicting Ebola Prevention Behaviors in the Gambia Evaluation of Ebola Prevention Promotions

Anna E. Gamma,† Jurgita Slekiene,† and Hans-Joachim Mosler*†

Author information · Article notes · Copyright and License information Disclaimer



JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

Back on Track

OV-1

Global Web Facing Data
on Cases, Deaths, and Recoveries

COVID
Data
Pipeline

Curated Public
Data Products

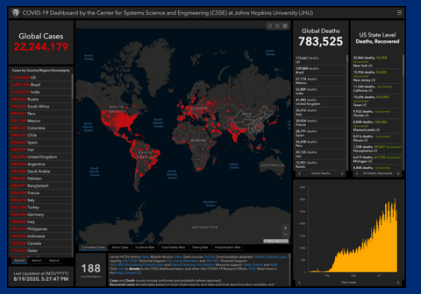
GitHub

Public
REST APIs

GIS & Web
Services

Optimized
JHU
Infrastructure

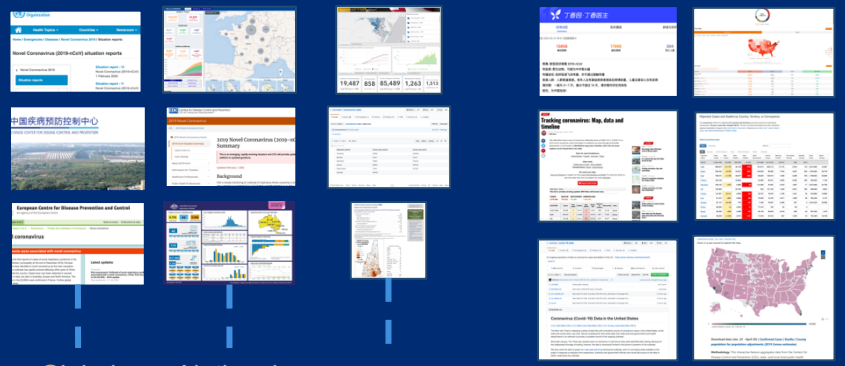
aws



CNN
npr

Public Consumers

COVID Data Pipeline - Initial Architecture



Global Sitreps | National Reporting | State & Local Reporting

"Authoritative" Data Sources

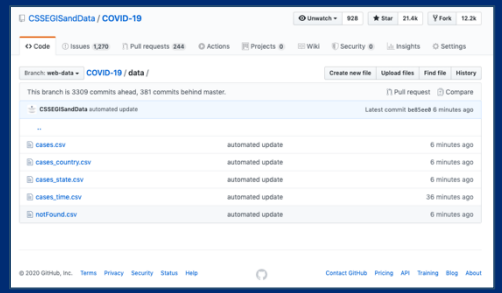
Crowd Sourced Data

~25

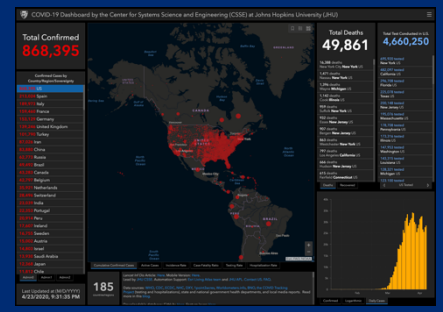
Expanded Data Sourcing

Collection | Curation | Production

Automated Product Generation



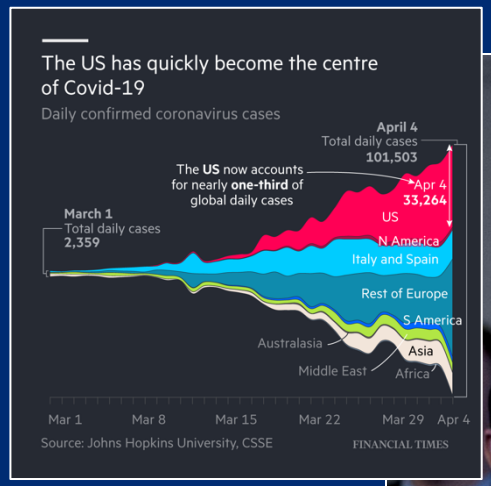
Open Data Sharing

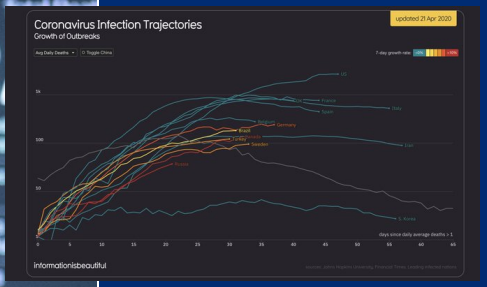


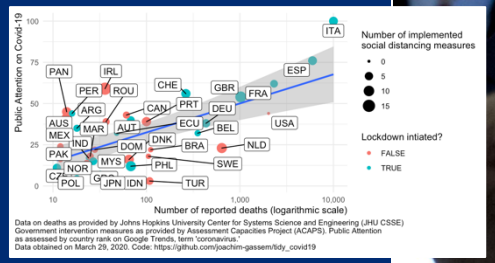
Automated Updates

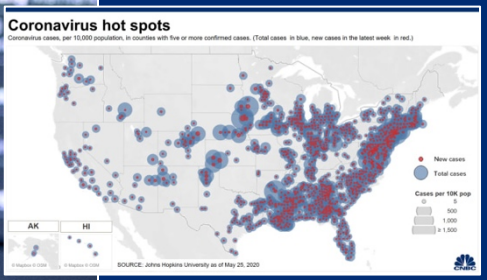
Automated Data Collection – Manual Error Recovery

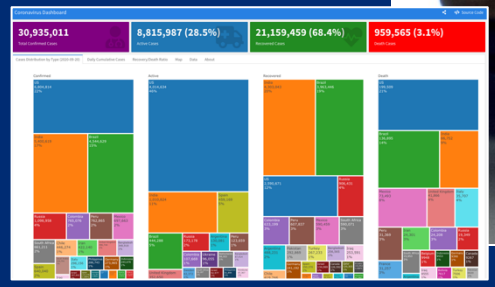
Expanding Use

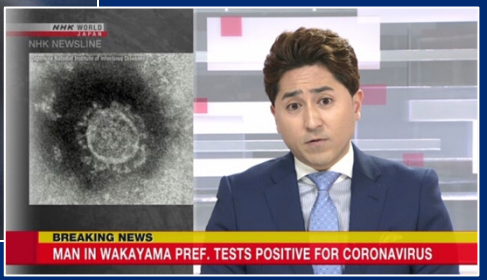




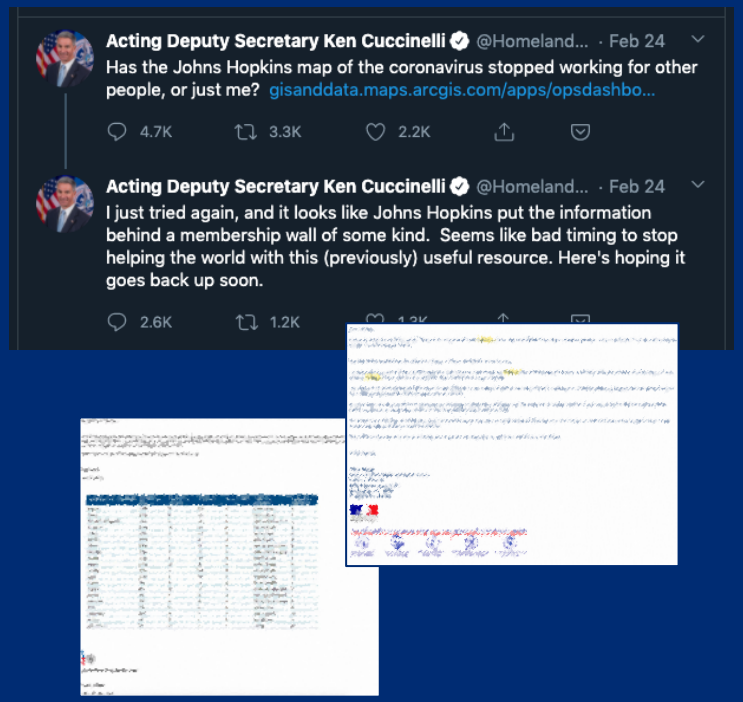


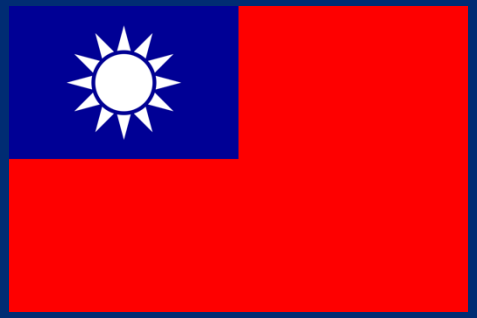






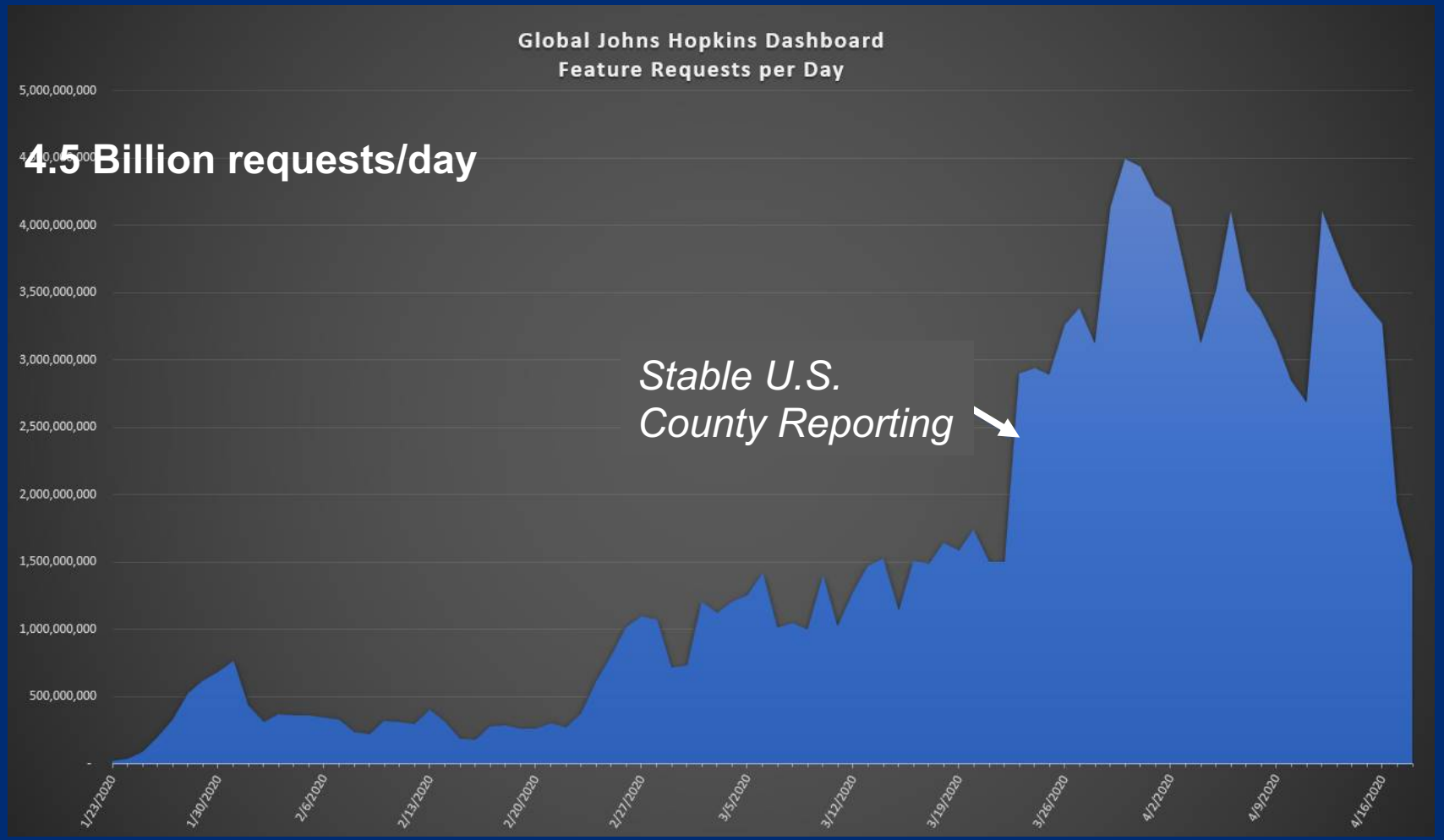
Led to Expanding Demands





Le Monde CNN npr CNBC

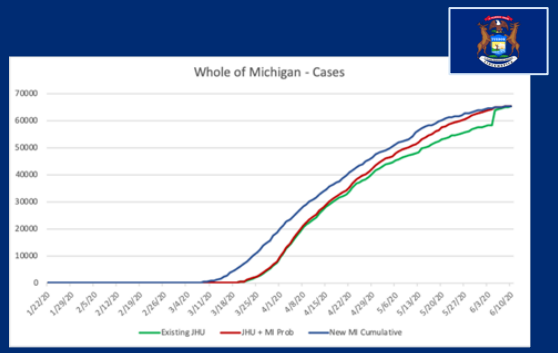
And Massive Growth in Data Requests



Why this is hard

(Hint: Standards Matter)

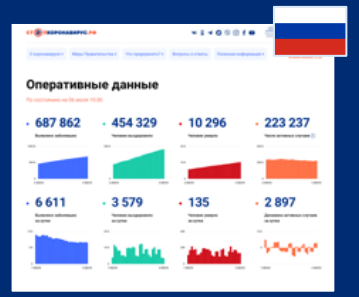
- Data reporting is inconsistent and constantly changing
- The scale of data is such that it requires methods for collection and collation that are both nuanced and complex
- Definitions and collection capabilities are inconsistent across counties, states, and countries

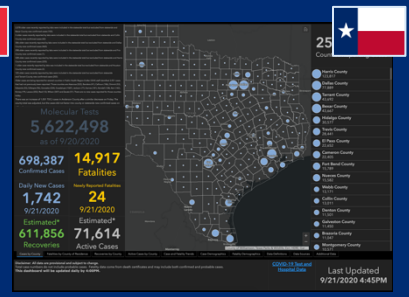


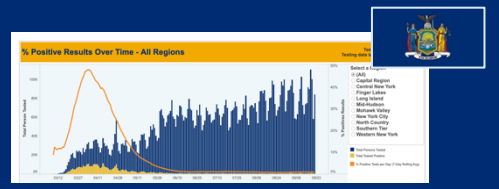
Mar

…

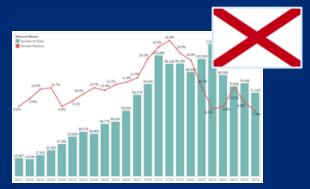
Jul



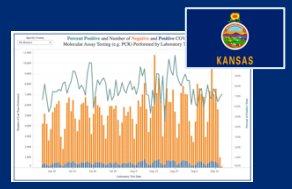




7-Day Rolling PCR



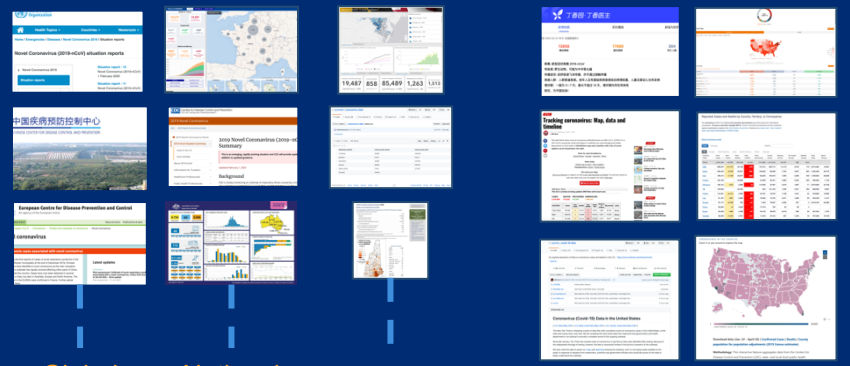
Epi-Week PCR+Antigen



Dsily PCR+Antigen

Simultaneously addressing both source instability globally and operating near a zero error rate requires extensive engineering and effort

Refined Gen 1 Architecture



Global Sitreps | National Reporting | State & Local Reporting

"Authoritative" Data Sources

Crowd Sourced Data

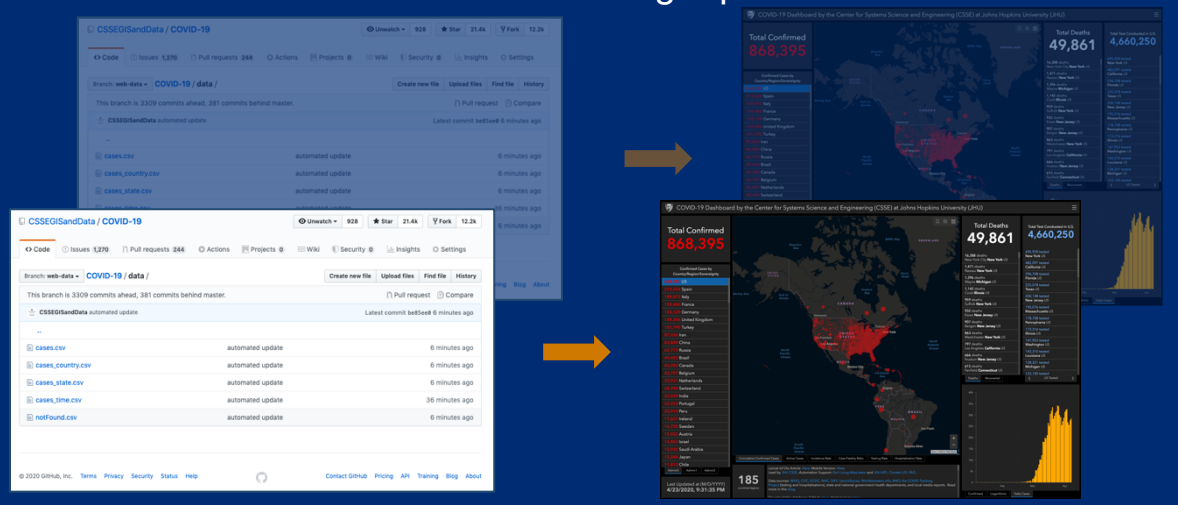
~100

Expanded Data Sourcing

Collection | Curation | Production

Automated Product Generation

Parallel Testing Pipeline



Open Data Sharing

Automated Updates

Automation for Anomaly Detection & Correction

Manual recovery where necessary

Anomaly Detection and Error Correction

Introduction of Anomaly Detection – But Dependency Remains

The Expanding Partnership

Coronavirus Resource Center

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

CENTERS FOR Civic Impact

JHU CSSE COVID-19 Dashboard

JOHNS HOPKINS WHITING SCHOOL of ENGINEERING

Center for Systems Science and Engineering

JOHNS HOPKINS SHERIDAN LIBRARIES

esri THE SCIENCE OF WHERE™

APL

JOHNS HOPKINS APPLIED PHYSICS LABORATORY

How to Build 2 Planes in Flight Simultaneously

Gen 1 Architecture + Ops

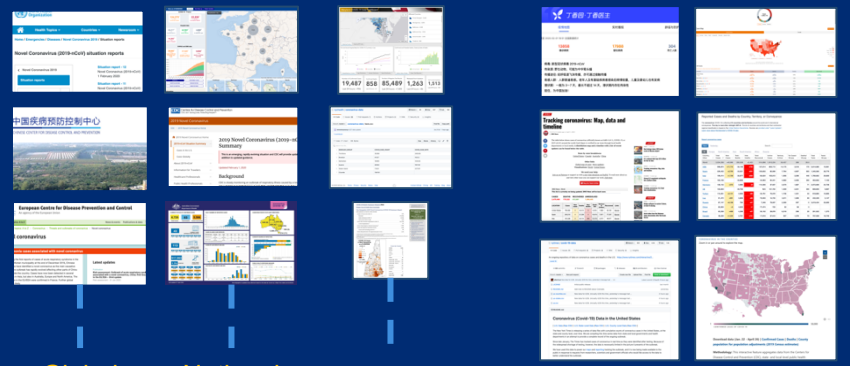
Gen 2 Architecture

Subject Matter Experts + Ops

CSSE COVID-19 Dashboard

Coronavirus Resource Center

Gen 2 Architecture



Global Sitreps | National Reporting | State & Local Reporting

"Authoritative" Data Sources

Crowd Sourced Data

~400

Expanded Data Sourcing

Collection

Autonomous Collection

Raw Data Store

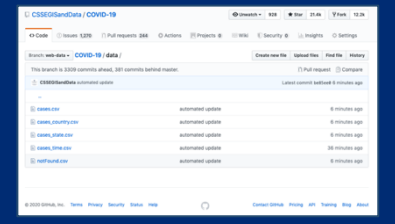
Data Cleaning Process

Clean Data Store

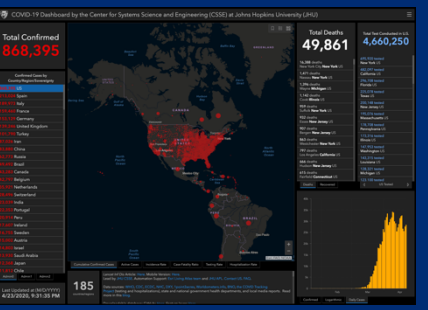
Data Fusion Process *n*

End Product *n*

Anomaly Detection & Recovery



Open Data Sharing



Automated Updates

Today

THE 100 MOST INFLUENTIAL PEOPLE OF 2020

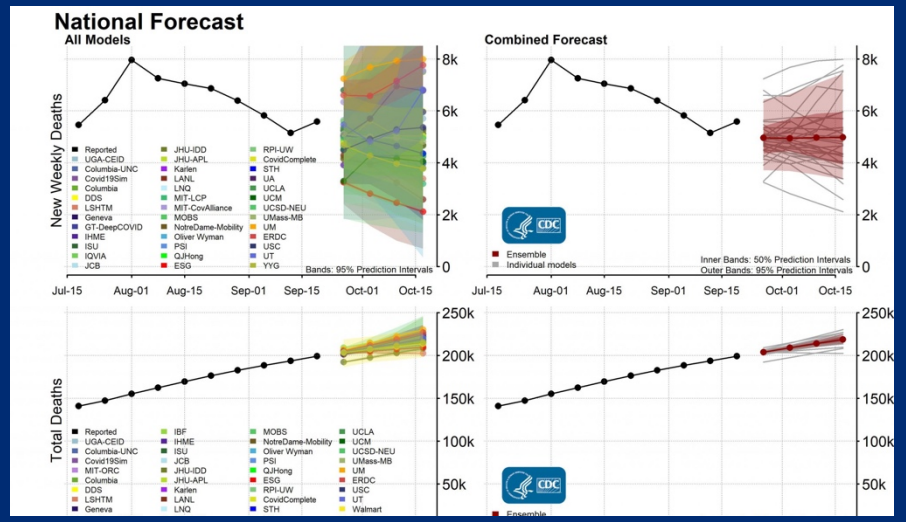
← BACK TO HOME

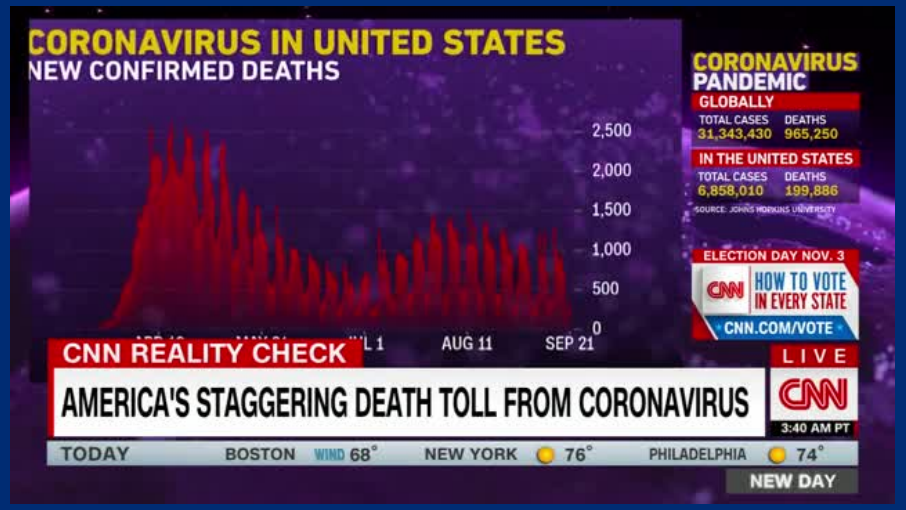
Lauren Gardner

Every morning, the first website I consult is the Johns Hopkins University (JHU) COVID-19 dashboard. It is the go-to resource to track the global pandemic. Public-health authorities use it to guide policies, with the dashboard featured prominently in emergency operation centers around the world. Researchers depend on it for their analysis and modeling. Virtually every news organization uses the JHU data as the basis for their reporting.

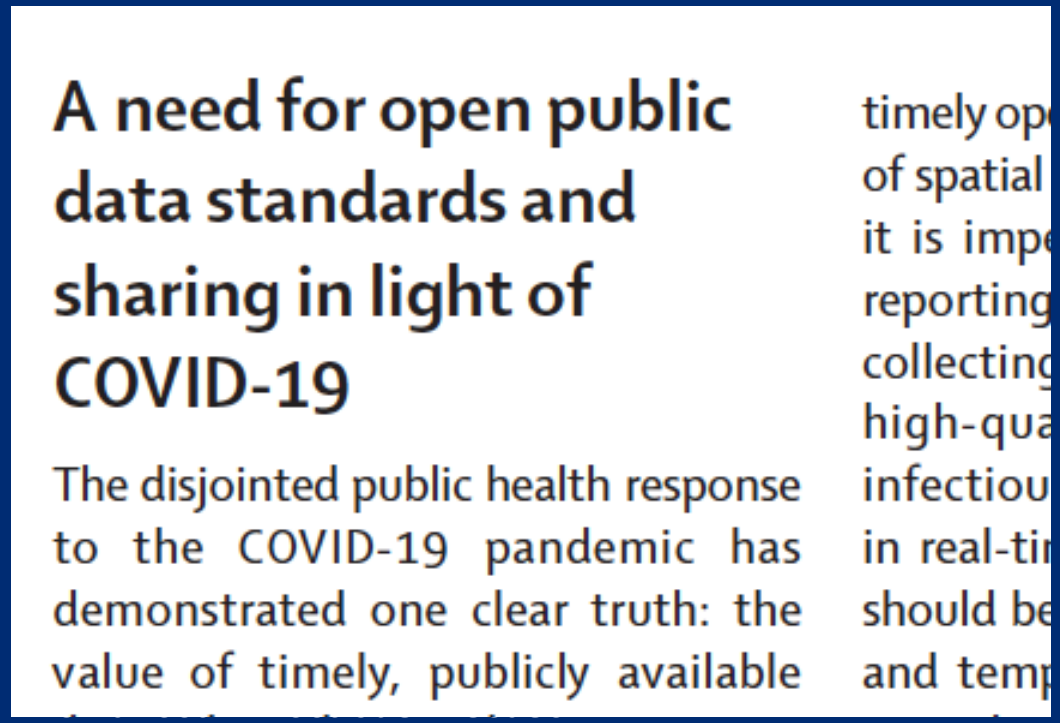
- Dr. Leana Wen

TIME





What's Next


A need for open public data standards and sharing in light of COVID-19

The disjointed public health response to the COVID-19 pandemic has demonstrated one clear truth: the value of timely, publicly available

timely op
of spatial
it is imp
reporting
collectin
high-qua
infectiou
in real-ti
should be
and temp

A need for open public data standards and sharing in light of COVID-19

L. Gardner, J. Ratcliff, E. Dong, A. Katz

August 10, 2020

https://doi.org/10.1016/S1473-3099(20)30635-6

"Moving forward, it is imperative that a **standardized reporting system** for systematically collecting, visualizing, and sharing high-quality data **on emerging infectious and notifiable diseases in real-time** is established. The data should be made available **at a spatial and temporal scale that is granular enough to prove useful** for planning and modelling purposes. Additionally, a critical component of the proposed system is the **democratization of data**; all collected information (observing necessary privacy standards) should be **made publicly available immediately upon release, in machine-readable formats, and based on open data standards**."

Acknowledgements

JOHNS HOPKINS

APPLIED PHYSICS LABORATORY