Tracking COVID-19 in Real-time: Challenges Faced and Lessons Learned

Dr. Lauren Gardner
Associate Professor, Civil and Systems Engineering at Johns Hopkins University
Co-Director, Center for Systems Science and Engineering (CSSE)
Affiliated Faculty, Infectious Disease Dynamics (IDD) Group at JH Bloomberg School of Public Health
Visiting Scientist, CSIRO (Australia)
Global COVID-19 Cases, Deaths and Recoveries

Cases by Country

Deaths and recoveries by country

US deaths, recoveries, testing

Cases by: States/Provinces or U.S. County

Multiple Map Layers

Textbox: Lots of important stuff

Number of new daily cases and deaths
Additional Map Layers

**Active COVID-19 Cases (Global)**

**Case-incidence Rate (Global)**

**Case-fatality ratio (Global)**

**U.S. Testing Data from COVID Tracking Project**
The Genesis of the Dashboard

(January) Modeling the International Spread of COVID-19

A recognized need for high quality, open, real-time data for outbreak modeling and decision making.
“Authoritative” Data Sources
Crowd Sourced Data
Manual Data Curation
Manual Dashboard Push

Fully manual process – did not scale
Limited Data Sharing
Data Sharing, Licensing and Terms of Use

CSSE COVID-19 GitHub

COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University

This is the data repository for the 2019 Novel Coronavirus Visual Dashboard operated by the Johns Hopkins Univer Center for Systems Science and Engineering (JHU CSSE). Also, Supported by ESRI Living Atlas Team and the Johns Hopkins University Applied Physics Lab (JHU APL).

Visual Dashboard (desktop):
https://www.arcgis.com/apps/opsdashboard/index.html#/c8d24f37d0f84e8da37547755cfc2000

Visual Dashboard (mobile):
http://www.arcgis.com/apps/opsdashboard/index.html#/c8d24f37d0f84e8da37547755cfc2000

Lancet Article:
An interactive web-based dashboard to track COVID-19 in real time

Provided by Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE):
https://systems.jhu.edu/

ArcGIS: Feature Service

ncov_cases

Current situation for the novel coronavirus starting from Wuhan, China

Feature Layer by CSSE_GisandData

Created: 25 Jan 2020  Updated: 17 Aug 2020  View Count: 988,516,603

Attribution 4.0 International (CC BY 4.0)
### Confirmed Cases Outside Mainland China

<table>
<thead>
<tr>
<th>Date</th>
<th>WHO</th>
<th>CSSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/22</td>
<td>1000</td>
<td>1200</td>
</tr>
<tr>
<td>1/27</td>
<td>2000</td>
<td>2500</td>
</tr>
<tr>
<td>2/1</td>
<td>4000</td>
<td>5000</td>
</tr>
<tr>
<td>2/16</td>
<td>8000</td>
<td>10000</td>
</tr>
</tbody>
</table>

### Confirmed Cases Inside Mainland China

Hubei changed COVID-19 reporting criteria

<table>
<thead>
<tr>
<th>Date</th>
<th>Confirmed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/22</td>
<td>10000</td>
</tr>
<tr>
<td>1/27</td>
<td>20000</td>
</tr>
<tr>
<td>2/1</td>
<td>40000</td>
</tr>
<tr>
<td>2/16</td>
<td>80000</td>
</tr>
</tbody>
</table>

*Thailand, S.Korea and Japan not shown, as reported before dashboard start date*
Week 1

The Dashboard was rapidly picked up by US and international press.
The Dashboard Team Grows
Stage 2 Architecture (Beginning February ...)

- Global Sitreps
- National Reporting
- State & Local Reporting
- "Authoritative" Data Sources
- Crowd Sourced Data
- ~25

GIS & Web Services
- Optimized JHU Infrastructure
- Collection
- Public Consumers
- Automated + Manual Data Collection – Manual Error Recovery

GitHub

CNN

npr

AWS

Public Consumers
4.5 Billion requests/day

Return to U.S. County Reporting

Dashboards: 2.7+ billion views
Feature Layers: 181+ billion requests
(As of Aug 13)
Led to Expanding Demands

Department of Health and Human Services Secretary’s Operations Center at ASPR
Stage 3 Architecture

Automated Open Data Product Generation

Parallel Testing Pipeline

Open Data Sharing

Automated Updates

Anomaly Detection and Error Correction

Automated process scales – but introduces complexity

Collection → Curation/Fusion → Production

Expanded Data Sourcing

Crowd Sourced Data

(400+) “Authoritative” Data Sources

Automated Open Data Product Generation

Global Sitreps

National Reporting

State & Local Reporting

Global

Sitreps

National

Reporting

State &

Local

Reporting

Automated Updates

Manual recovery where necessary

Anomaly Detection and Error Correction

Automated process scales – but introduces complexity
JHU CSSE COVID-19 Dashboard Data Uses Today

Driving Public Health Policy: CDC COVID-19 Forecasting

Helping inform the decision making of people everyday.

National Forecast

Mainstream Media

Supporting modeling and response efforts

An interactive web-based dashboard to track COVID-19 in real time

E Dong, H Du, L Gardner

The Lancet infectious diseases 20 (5), 533-534
JHU Coronavirus Resource Center – U.S. Map
Global Confirmed
20,950,402

Global Deaths
760,213

U.S. Confirmed
5,254,878

U.S. Deaths
167,253

NEW CASES OF COVID-19 IN US STATES

Have states flattened the curve?
See if new cases are rising or declining as states reopen.

BY REGION
See the latest data in your region
Explore stats and trends specific to your country or U.S. state

TESTING TRENDS TOOL
Track trends in COVID-19 cases and tests

STATE TIMELINE
Impact of opening and closing decisions by state
A look at how social distancing measures may have influenced trends in COVID-19 cases and deaths
Why this is hard (Hint: Standards Matter)

- Instability and inconsistencies in reporting (retrospective reporting, data structure and mechanisms, machine readability)
  - Our solution: Anomaly detection system
- Discrepancies in reported numbers among authoritative sources (TX, UK, France)
  - Our solution: Fusion Logic
- Variability in frequency and time-of-day reporting across locations
  - Our Solution: 0400 GMT was the least bad chalk line
- Ambiguity in parameter definitions (probable vs. confirmed; reported vs onset time; diagnostic vs antibody, recoveries, etc.)
  - Our Solution: Report consistent across locations, and according to latest CDC guidelines.

Simultaneously addressing both source instability globally and operating near a zero error rate requires extensive engineering and effort.
Open Data Principles and the need for a better system

“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.”

Such an established system will both help us make sound and timely decisions when faced with the next potential outbreak, and can help us establish the public trust necessary to combat health related disinformation.
Acknowledgements: The Dashboard Team

(ESRI)
Making a Difference Award

Johns Hopkins University
Center for Systems Science and Engineering

Dr. Lauren Gardner
Engineering Professor and Co-Director

Ensheng Dong
Graduate Student
Acknowledgements: Funding and Resource Support
Thank You.