Subject: Re: Final CNN request for comment re: Worldometer story

Date: Monday, May 18, 2020 at 4:58:09 PM Eastern Daylight Time

From: Jill Rosen

To: Hernandez, Sergio, Douglas Donovan

CC: McLean, Scott, Perez Maestro, Laura

Attribute to a Johns Hopkins University spokesperson: At Johns Hopkins more than two dozen people from multiple divisions – the Center for Systems Science and Engineering, the Applied Physics Lab, the Bloomberg School of Public Health, and the library's data services, and others – have been working around the clock every day making sure the site's automated data feeds are accurate and up to date so that policy makers and the public have a free and open resource for tracking the virus.

Everything below can be attributed to [Lauren Gardner](#):

How did JHU make the initial decision to use Worldometer as a data source for its Covid-19 data? What was its selection criteria and vetting process?
- All sources we use in the dashboard – and there are dozens -- are either primary health authorities or data aggregation websites that provide sources for their data that can be validated. Before incorporating any new source, we validate their data by comparing it against other references.
- It is important to recognize there are differences in reporting across sources because reporting varies across (and even within) countries. For some countries we choose to report data differently than other data aggregation sites, and for these countries we source our data directly from the relevant primary health authority for that country/region. This includes, but is not limited to, data for Spain, France, Germany, Australia, US, China, Hong Kong, Macau, Taiwan, Kosovo, Serbia and the West Bank and Gaza.
- France is one example of the significant challenges in reporting. We source our data directly from the French Health Ministry and relevant regional health ministries. This data includes (confirmed and probable) cases, deaths, and recoveries in overseas regions and "departments" such as Guadeloupe and Mayotte, as well as a breakdown of cases from nursing homes. The departments are included as separate entries so our numbers have to be corrected to avoid double counting. This distinction separates ours from other data aggregation sites. The logic is discussed in depth at [this link](#) on our GitHub repository.

Which specific data points does JHU rely on Worldometer for?
- We try not to use a single source for any of our data. We use a wide collection of data sources to ensure the best-level of consistency and accuracy as possible. We use reporting from public health agencies and sources of aggregation to cross-validate numbers.

We understand that JHU scrapes Worldometer and other sources for new data. Once sources have been scraped and new data identified, is the data automatically published to the ArcGIS dashboard or GitHub repository, or does a human validate it first? (If so, what is that validation process?)
- We scrape data from multiple sources, but before data is published in the dashboard it is processed through a two-stage anomaly detection system we have developed specifically for this dashboard. The first layer alerts us to moderate changes in any updated value, so

we can double check them in real time. A second more conservative layer holds back updates that exceed a certain pre-defined threshold. This second layer of the anomaly detection system requires a human to manually check and approve the values before publication to the dashboard. Any anomalies are then followed up and addressed immediately.

You have referred to a "process of continual cross-checking and refinement." Can you provide more detail on what that process is?

- We regularly compare our dashboard data against reporting by independent sources such as WHO, and validate the data trends for each country and state. An example of this is illustrated in the [Lancet Inf Dis publication](#) detailing the dashboard in its early stages. This same approach has been in place throughout, and presented in multiple (recorded) presentations I have given on the architecture of the dashboard. When inconsistencies between sources are identified we investigate them, and if warranted, we self-correct the github timeseries files, and document any change in the github under our '[data modification records](#)'.

RE: JHU has declined to say what specific data points it relies on Worldometer for, but issues with the counter site's data have caused at least one notable error:

- **On April 8 JHU's global tally of confirmed Covid-19 cases briefly crossed 1.5 million before dropping by more than 30,000.**
- **JHU told CNN at the time, that the error appeared to come from a double-counting of French nursing home cases.**
- **Johns Hopkins's figure appeared to come directly from Worldometer, which did not cite a source for its tally of cases in France.**

    -- The drop in cases was explained in April here, on our Github: https://github.com/CSSEGISandData/COVID-19/issues/2094

From: "Hernandez, Sergio" <Sergio.Hernandez@cnn.com>
Date: Monday, May 18, 2020 at 9:53 AM
To: Jill Rosen <jrosen@jhu.edu>, Douglas Donovan <dougdonovan@jhu.edu>
Cc: "McLean, Scott" <Scott.McLean@turner.com>, "Perez Maestro, Laura" <Laura.Perez.Maestro@cnn.com>
Subject: Re: Final CNN request for comment re: Worldometer story

Hi Jill,

Thanks for getting back to us so quickly,

We understand your position. We could extend our deadline until close of business (5 p.m.) today. That would give the University a full business day. We think that's fair, especially given that most of what we're outlining below should just be a recap/reiteration of CNN's questions from the last couple of weeks.

(Specifically, Scott and Laura's questions from late April and May 10-12, and some of the exchanges we've had on sourcing/vetting when we've noticed errors in the dashboard data.)

Thanks,

—

Sergio Hernandez
Data Editor, **CNN** | desk: (212) ▮▮▮▮▮▮▮ | mobile: (646) ▮▮▮▮▮▮

From: Jill Rosen <jrosen@jhu.edu>
Date: Monday, May 18, 2020 at 9:38 AM
To: "Hernandez, Sergio" <Sergio.Hernandez@cnn.com>, Douglas Donovan <dougdonovan@jhu.edu>
Cc: "McLean, Scott" <Scott.McLean@turner.com>, "Perez Maestro, Laura" <Laura.Perez.Maestro@cnn.com>
Subject: Re: Final CNN request for comment re: Worldometer story

Thanks Sergio. We're going to need more time to be able to look at all of this and respond thoughtfully. Since you had originally mentioned we'd get these questions on Friday, we'd like to have a day or two.

From: "Hernandez, Sergio" <Sergio.Hernandez@cnn.com>
Date: Monday, May 18, 2020 at 9:18 AM
To: Jill Rosen <jrosen@jhu.edu>, Douglas Donovan <dougdonovan@jhu.edu>
Cc: "McLean, Scott" <Scott.McLean@turner.com>, "Perez Maestro, Laura" <Laura.Perez.Maestro@cnn.com>
Subject: Final CNN request for comment re: Worldometer story

Hi Jill,

As you are aware, we are preparing a piece for this week about Covid-19 data. The story focuses on Worldometer and how governments (most recently the Spanish government) and institutions (including Johns Hopkins) have relied on it as a source for some of their statistics.

Below is a list of our questions, followed by a list of statements and quotes about JHU that we currently plan to include in our story. Please let us know if you see anything here that is factually inaccurate or that you'd like to respond to or comment on. If you'd prefer to speak by phone, we'd be happy to record an interview and include those comments in the story.

We're hoping to post this story tonight/early tomorrow. If you could please get back to us with your responses by **2 p.m. ET today**, we would be grateful.

Thanks as always,

Sergio

—

Sergio Hernandez
Data Editor, **CNN** | desk: (212) ▮▮▮▮▮▮▮ | mobile: (646) ▮▮▮▮▮▮

Questions:

- How did JHU make the initial decision to use Worldometer as a data source for its Covid-19 data? What was its selection criteria and vetting process?

- Which specific data points does JHU rely on Worldometer for?

- We understand that JHU scrapes Worldometer and other sources for new data. Once sources have been scraped and new data identified, is the data automatically published to the ArcGIS dashboard or GitHub repository, or does a human validate it first? (If so, what is that validation process?)

- You have referred to a "process of continual cross-checking and refinement." Can you provide more detail on what that process is?

Statements about JHU:

- On April 28, Spanish Prime Minister Pedro Sánchez said: "we have found out about another study, from the Johns Hopkins University, that […] ranks us fifth in the world in total tests carried out."

- Jill Rosen, a spokesperson for the university told CNN that JHU could not identify a report matching Spanish Prime Minister Pedro Sánchez' description.

- JHU has not published an international data on Covid-19 testing.

- The study Sánchez cited does not exist.

- JHU lists Worldometer as one of several data sources for its Coronavirus dashboard.

- JHU has declined to say what specific data points it relies on Worldometer for, but issues with the counter site's data have caused at least one notable error:
    - On April 8 JHU's global tally of confirmed Covid-19 cases briefly crossed 1.5 million before dropping by more than 30,000.
    - JHU told CNN at the time, that the error appeared to come from a double-counting of French nursing home cases.
    - Johns Hopkins's figure appeared to come directly from Worldometer, which did not cite a source for its tally of cases in France.

- Wikipedia editor James Heilman said Wikipedia volunteers have noticed persistent errors with both Worldometer and JHU

- It's not clear why the Spanish government continues to insist the testing data published by Worldometer was put out by JHU

Quotes about JHU:

- Jill Rosen, on why JHU has chosen to rely on Worldometer given concerns about its accuracy or who operates it:
    - *"We use many sources to corroborate the data we publish. It's a process of continual cross-checking and refinement to ensure that the data we are present [sic] is as accurate and timely as*

possible."

- Spanish Health Minister Salvador Illa, on the provenance of testing data cited by Spanish PM Sánchez on April 28:
    - *"It is data given by the John Hopkins University [...] taken from, as a fundamental source of information, the website Worldometer."*

- James Heilman, an clinical assistant professor of emergency medicine at the University of British Columbia and a Wikipedia editor:
    - Noticed persistent errors with Worldometer and also with *"a more reputable name with a long history of accuracy"* (referring to Johns Hopkins)
    - *"We hope they also double check the numbers."*

- Phil Beaver, data scientist at University of Denver, on whether JHU should be using Worldometer:
    - *"I am not sure, that is a great question, I kind of got the impression that Worldometer was relying on [Johns] Hopkins"*

- Eduardo Mathieu, data manager for Our World in Data:
    - *"I think JHU has been under a lot of pressure to update their numbers"*
    - *"Because of this pressure they have been forced to or incentivized to get data from places that they shouldn't have, but in general I would expect JHU to be a fairly reliable source."*

- Spokeswoman for the Embassy of Spain in London:
    - *"Back in April Mr. Sánchez mentioned analysis of statistical data carried out by Johns Hopkins University that are based upon data published by Worldometer."*

- Spokesperson for a UK government office:
    - *"Both Worldometers and John Hopkins provided comprehensive and well respected data. As the situation developed, we transferred from Worldometers to John Hopkins as John Hopkins relies more on official sources"*