



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



# Data journalism and the COVID-19 pandemic: opportunities and challenges

Academics and public health practitioners studying communicable disease dynamics have long advocated for open-access data to better inform risk assessments. During any evolving outbreak, the collection, aggregation, visualisation, and analysis of granular data is paramount to developing appropriate public health interventions.<sup>1</sup> The COVID-19 pandemic has underscored the need for this type of information, especially in relation to context (eg, timing and intensity of interventions) and epidemiology (eg, spatially resolved and age-specific case counts).

Non-traditional disease surveillance tools, including news media reporting, have disseminated event-based information during past disease outbreaks.<sup>2</sup> The current global health crisis has highlighted the additional possibilities that so-called data journalism can offer. While the news media has traditionally reported on events of public health importance, media outlets over the course of the COVID-19 pandemic have also conducted data collation, including detailed summaries of case counts and deaths, data curation, and, in some instances, analysis (table). Although some of these

	Description	Data sources	URL
<i>Financial Times</i> : coronavirus tracker	The <i>Financial Times</i> analyses the scale of the COVID-19 outbreak including the collection and analysis of data on excess mortality (ie, numbers of deaths higher than the historical average) across the globe	WHO, the COVID Tracking Project, Johns Hopkins University, Our World in Data, US Centers for Disease Control and Prevention, and others	<a href="https://www.ft.com/content/a2901ce8-5eb7-4633-b89c-cbd5b386938">https://www.ft.com/content/a2901ce8-5eb7-4633-b89c-cbd5b386938</a>
<i>The Economist</i> : tracking COVID-19 excess deaths across countries globally	Global excess death tracker	Human Mortality Database, World Mortality Dataset, and EuroMOMO	<a href="https://www.economist.com/graphic-detail/coronavirus-excess-deaths-tracker">https://www.economist.com/graphic-detail/coronavirus-excess-deaths-tracker</a>
<i>The New York Times</i>			
The pandemic's hidden toll	Excess deaths during the COVID-19 pandemic	Data are compiled from official national and municipal data for 24 countries	<a href="https://www.nytimes.com/interactive/2020/04/21/world/coronavirus-missing-deaths.html">https://www.nytimes.com/interactive/2020/04/21/world/coronavirus-missing-deaths.html</a>
Tracking the coronavirus at US colleges and universities	COVID-19 tracker at US colleges and universities; with no national tracking system, and statewide data available only sporadically, colleges have been making their own rules for how to tally infections	<i>The New York Times</i> surveyed more than 1900 US colleges and universities for COVID-19 information	<a href="https://www.nytimes.com/interactive/2020/us/covid-college-cases-tracker.html">https://www.nytimes.com/interactive/2020/us/covid-college-cases-tracker.html</a>
What we know about coronavirus cases in K-12 schools so far	Reporting focused on district-level and statewide COVID-19 case totals for public schools in the USA; the numbers presented are minimums because of differences in reporting	State and local health and education agencies or were identified by The Covid Monitor or the National Education Association and independently confirmed by <i>The New York Times</i> ; <i>The New York Times</i> directly surveyed every school district in eight states: Colorado, Florida, Georgia, Illinois, Indiana, North Carolina, Texas, and Utah	<a href="https://www.nytimes.com/interactive/2020/09/21/us/covid-schools.html">https://www.nytimes.com/interactive/2020/09/21/us/covid-schools.html</a>
How full are hospital ICUs near you?	Occupancy levels in US ICUs	US Department of Health and Human Services (hospital capacity data); US Department of Homeland Security (hospital locations); and the Covid-19 Hospitalization Tracking Project, University of Minnesota Carlson School of Management	<a href="https://www.nytimes.com/interactive/2020/us/covid-hospitals-near-you.html">https://www.nytimes.com/interactive/2020/us/covid-hospitals-near-you.html</a>
The Atlantic: the COVID-19 tracking project	Collected, cross-checked, and published COVID-19 data from 56 US states and territories regarding testing, hospitalisation, and patient outcomes, providing ethnic demographic information and data on long-term-care facilities*	COVID-19 data from websites of US state or territory public health authorities; a public data API provides access to all their data at a national and state level	<a href="https://covidtracking.com">https://covidtracking.com</a>
<i>The Hindu</i> : coronavirus India tracker	Collects, aggregates, analyses, and visualises state-level COVID-19 cases, deaths, and testing data from India and globally	Bing and Johns Hopkins University	<a href="https://www.thehindu.com/coronavirus/">https://www.thehindu.com/coronavirus/</a>
Zeit Online: coronavirus in Deutschland und bei Ihnen [coronavirus cases in Germany and globally]	Collects, aggregates, analyses, and visualises COVID-19 cases, deaths, patients in ICUs, and vaccinations	Robert Koch Institute, websites of German counties and states, and Johns Hopkins University	<a href="https://www.zeit.de/wissen/gesundheit/corona-zahlen-deutschland-neuinfektionen-inzidenz-aktuelle-karte">https://www.zeit.de/wissen/gesundheit/corona-zahlen-deutschland-neuinfektionen-inzidenz-aktuelle-karte</a>

ICU=intensive care unit. API=application program interface. \*As of March 7, 2021, The Atlantic is no longer collecting new data.

**Table: Examples of COVID-19 data collection, visualisation, and analysis**

data are available through public health department websites, relevant interpretation and data visualisations by media outlets have provided information on the COVID-19 pandemic to the public in near real time.

Before COVID-19 was declared a Public Health Emergency of International Concern, news reports served as key data sources to further understand disease transmission and spread. Academic institutions and researchers assembled early epidemiological data scattered across various news articles to inform risk assessments, forecasts, and policy decisions.<sup>3</sup> The relative dearth of traditional public health data at the beginning of an epidemic is not a new phenomenon. During the 2014–15 west Africa Ebola outbreak, for example, early epidemiological data were often only available through local and international news media articles.<sup>4</sup> As the COVID-19 pandemic evolved, and in response to epidemiological data gaps, news media outlets began to collect and synthesise data for scenarios involving congregate settings such as schools, large public events, and household transmission. In some cases, media have actively reached out and solicited case counts from their readers—a strategy known as participatory surveillance—effectively recruiting the public back into public health (table).

News media outlets have also been among the first to systematically collect, aggregate, and analyse excess death counts. For example, the data behind the *Financial Times* tracker for COVID-19 excess deaths dates to April, 2020; the tracker is open access, and the code and methodology used to clean, analyse, and present the data are available on GitHub. *The Economist* and *The New York Times* have also provided their own analyses on excess deaths (table).

As the COVID-19 pandemic has shown, there is an urgent need for real-time data that can inform risk assessments to guide public health interventions. While traditional data collection remains the cornerstone of outbreak response, public health programmes and information technology infrastructure are chronically underfunded in many countries and are not always well positioned to collect contextual information in a flexible manner. This is of particular concern during an outbreak when traditional data sources might lag in reporting cases early on. Another key need for epidemic forecasting and risk assessments is data surrounding non-pharmaceutical interventions such

as physical distancing, school closures, and lockdowns.<sup>5</sup> Interventions differ regionally and implementation timelines are not often readily disseminated. While non-traditional disease surveillance systems have begun to fill some of these gaps, more can be done. Partnerships between academic research centres and news media should be considered, given their complementary strengths; indeed, collaboration between these entities might mitigate their respective weaknesses as well. While news media can rapidly aggregate and disseminate information, they might be unable to sustain these efforts following the course of an outbreak. Likewise, research centres might be able to continue collating and analysing data long after an outbreak has ended, but might be unable to collect relevant information in a timely manner early in an outbreak. It is also important to note that news media data and data visualisations, while informative, differ from peer-reviewed literature. Divergent incentives, intended audiences, and analytic methodologies can result in very different outputs and conclusions. Supporting collaborations between news media outlets that can provide an expedient data stream and academic institutions that can support targeted analyses could be an important step towards improving outbreak response timeliness in the future.

In response to the COVID-19 pandemic, several global epidemiological data collection and harmonisation efforts have been initiated to provide guidance on conforming case definitions, data formatting, and data sharing. As these efforts are further developed, data collected by media outlets could be integrated for use by researchers and policy makers, although regulatory issues surrounding data privacy will need to be addressed. Cross-collaborations between academic groups and the media should be encouraged and the role of the media in curating, analysing, and sharing epidemiological information that is otherwise hard to collect should be recognised. While these efforts should be considered complementary to traditional public health endeavours, the rapid dissemination of accurate, real-time information remains paramount in the face of current and future communicable disease outbreaks.

We declare no competing interests.

Copyright © 2021 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY 4.0 license.

\*Angel Desai, Pierre Nouvellet, Sangeeta Bhatia, Anne Cori, Britta Lassmann

andesai@ucdavis.edu

Department of Internal Medicine, Division of Infectious Diseases, University of California Davis, Sacramento, CA 95817, USA (AD); School of Life Sciences, University of Sussex, Brighton, UK (PN); MRC Centre for Global Infectious Disease Analysis, School of Public Health, Faculty of Medicine, Imperial College London, London, UK (SB, AC); International Society for Infectious Diseases, Brookline, MA, USA (BL)

- 1 Desai AN, Kraemer MUG, Bhatia S, et al. Real-time epidemic forecasting: challenges and opportunities. *Health Secur* 2019; **17**: 268–75.
- 2 Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection—harnessing the Web for public health surveillance. *N Engl J Med* 2009; **360**: 2153–57.
- 3 Imai N, Dorigatti I, Cori A, Riley S, Ferguson NM. Report 1: estimating the potential total number of novel Coronavirus cases in Wuhan City, China. Version 2. Imperial College COVID-19 Response Group. Jan 17, 2020. <https://doi.org/10.25561/77149> (accessed July 6, 2021).
- 4 Cori A, Donnelly CA, Dorigatti I, et al. Key data for outbreak evaluation: building on the Ebola experience. *Philos Trans R Soc Lond B Biol Sci* 2017; **372**: 20160371
- 5 Blavatnik School of Government, University of Oxford. COVID-19 government response tracker. <https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker> (accessed July 6, 2021).