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.
The current pandemic highlights the power of data. The data infrastructures we've built have provided excellent platforms to analyze data, yet it brings into focus a gap we have in handling the infodemic challenge we face today. We need mechanisms to enable rapid classification of the trustworthiness of datasets.

Data management goes hand in hand with data. The volume of data generated and captured today is unprecedented, meaning that the ability to effectively manage these data is now a critical function and will only become more important in the future. The world of research data management has matured over the last 2 decades, and data management has gone from being a nascent concept to becoming an accepted norm for those dealing with data in any professional capacity.

Groups like the Research Data Alliance (RDA), Committee on Data (CODATA), and Australian National Data Service (ANDS) have emerged to help codify the increasingly complicated data landscape. These groups have been tremendously valuable in helping converge the thoughts and practices within the community of data practitioners. These efforts have led to an increased understanding of the complicated landscape of data and the codification of best practices required to manage these data for use and reuse.

In December 2019, the world became aware of an impending pandemic, and by February 2020, the world had realized that it is facing a public health crisis the likes of which we have not seen in living history. The COVID-19 pandemic took hold of the world at an imagined speed, and the world was forced to scramble at a previously unprecedented pace to respond to the emerging crisis in order to save millions of lives.

Very soon it became apparent that data are at the heart of the global response to this pandemic. We require data to understand the virus. We need data to understand how it is spreading within the community. We require data to understand the impact it is having on the patients. We need data to know the economic and social impact that our decisions are having.

Unlike any time in the past, where the same data would have been desired but impossible to obtain, we are in the age of data, and we are able to generate the data required at a massive scale. This pandemic highlights the power and value of data; furthermore, it also highlights the tremendous value of data when they are shared and actionable. An example of this global collaboration for data sharing was the speed at which the full genome of COVID-19 was sequenced and published—barely a month after the first patient was admitted into a hospital in Wuhan.

The open sharing of data with minimal friction accelerates our collective ability to respond to global issues. Many key organizations have called on the world to remove the speed barriers to research by having open access to the data. The scientific community responded, and more than 30 leading publishers committed to making all of their COVID-19 and coronavirus-related publications, as well as the available data supporting them, immediately accessible in public repositories.

The data infrastructures we have built over the last 2 decades have provided excellent platforms to collect, gather, and analyze the plethora of data that are being generated. We need to share biological data at a rapid pace, yet we need to maintain confidentiality and anonymity for the safety and security of patients. We need to mix different heterogeneous datasets and make the resulting combined datasets sharable and reusable by others. These data infrastructures provided the capability we need to do this.

Yet our ability to easily generate large quantities of data, our need to bring these data together to consume and inform decision making, be it for research or policy making, brings into focus a gap we have in our data landscape. The OECD report “Why open science is critical to combating COVID-19” highlights a number of areas where our data infrastructures still have challenges that need to be addressed in order to enable the data-driven decision making needed during a crisis. It highlights that we are currently ill equipped to deal with the resulting infodemic challenges that result from the tsunami of data that we are able to generate today. The WHO declared that we know that every outbreak will be accompanied by a kind of tsunami of information, but also within this information you always have misinformation, rumors, etc. ... So it is a new challenge, and the challenge is the [timing] because you need to be faster if you want to fill the void. ...What is at stake during an outbreak is making sure people will do the right thing to control the disease or to mitigate its impact. So it is not only information to make sure people are informed; it is also making sure people are informed to act appropriately.

Against this backdrop, how does the consumer of data know the trustworthiness of the data that they are consuming? How do they know if the data are of sufficient quality for reuse in further research?
or for operational decision making? This poses a significant challenge for the data consumer who needs to make decisions underpinned by data and to understand the trustworthiness of the data that they are basing their decisions upon.

Data are always collected in context. Data are always used in context. When dealing with primary data, the collecting and the use contexts are known. However, when there is secondary use of data, it is very difficult to know the collection context and decide whether the data are of sufficient quality for the task at hand. It is not uncommon for data consumers to spend more than 50% of their effort in understanding the value and trustworthiness of secondary data in context. In most scenarios, reducing this effort would produce significant value in freeing up time and effort to focus on the application of the data. In a crisis situation as we are in now, this can be the difference between lives saved and lives lost.

There is a need for our data infrastructure and platforms to be able assist in this and have mechanisms that enable rapid classification of the trustworthiness of datasets for a specific purpose, be it research or decision making. Such mechanisms are nascent within our data and information infrastructure at present. We require classification frameworks to be established and incorporated into our infrastructures and matured over time.

One framework that the author is aware of is the operational readiness level (ORL) framework (Figure 1). The ORL framework was developed by the All Hazards Consortium and the Disaster Lifecycle Cluster at Earth Science Information Partners (ESIP) as a mechanism for achieving data-driven decision making. The ORL provides a practical model of how a framework could be built to enable the classification of the trustworthiness of data to occur in a rapidly changing environment.

This framework provides a mechanism for rapidly evaluating incoming data and attaching a ranking to the data (the ORL), which informs the data consumer the confidence level they can place on the data for basing their decisions upon.

The framework is being utilized by some of the electric sector services in the United States to inform operational activities. What makes this framework useful is that it enables the context-specific classification of the data and thus enables the data consumer to focus their efforts on seeking the data that are of relevance to the decisions at hand.

In our global data infrastructure landscape, this form of capability is required as an integral component. We need an increasing dialog around how we establish such frameworks to enable context-specific confidence on the fitness of data for specific purposes. We need an increase in the discourse on how we share the context-specific trustworthiness of data. Events such as the COVID-19 pandemic highlight the need, and the urgency, for such capability to be available in order to respond effectively to global (and local) events.

In many respects, this is likely to be the next key frontier where effort will need to be expended to be able to maximize our ability to utilize the value of the data that we are generating and collecting.

Figure 1. ORL Model Decision Tree
Operational readiness level provides a framework for the rapid classification of data so that non-experts can trust and utilize data for confident decision making. Image credit: ESIP Federation Disaster Lifecycle Cluster/All Hazards Consortium, Sensitive Information Sharing Environment (SISE) GIS Committee, design by Kari Hicks, Sr. GIS Analyst, Duke Energy (USA).
REFERENCES

1. Cohen, J. (2020). Chinese researchers reveal draft genome of virus implicated in Wuhan pneumonia outbreak. Science https://www.sciencemag.org/news/2020/01/chinese-researchers-reveal-draft-genome-virus-implicated-wuhan-pneumonia-outbreak.

2. Carr, D. (2020). Publishers make coronavirus (COVID-19) content freely available and reusable, Wellcome https://wellcome.ac.uk/press-release/publishers-make-coronavirus-covid-19-content-freely-available-and-reusable.

3. Organisation for Economic Co-operation and Development (2020). Why open science is critical to combating COVID-19, OECD https://read.oecd-ilibrary.org/view/?ref=129_129916-31pgnl7cb&title=Why-open-science-is-critical-to-combatting-COVID-19.

4. Hua, J., and Shaw, R. (2020). Corona Virus (COVID-19) “Infodemic” and Emerging Issues through a Data Lens: The Case of China. Int. J. Environ. Res. Public Health 17, 2309.

5. Jones, D. (2019). Webinar: Managing disasters through improved data-driven decision-making and ORLs, Australian Research Data Commons, July 16, 2019 https://www.youtube.com/watch?v=HwVuJvw3S8c.

6. All Hazards Consortium. Home. https://www.ahcusa.org/.

7. Earth Science Information Partners. Disasters. https://wiki.esipfed.org/index.php/Disasters.

8. GeoCollaborate. GeoCollaborate Dashboard. https://frwg.geocollaborate.com/dashboard/.

About the Author
Mr. Kheeran Dharmawardena, MBA, BComp, FIML, has 20 years of experience in the delivery of information and communications technology (ICT) services within the higher-education sector, where he has aligned ICT delivery with business strategy. He has played key roles in building national research infrastructure and has a deep understanding of the Australian research landscape. In recent times, he has been instrumental in establishing cross organizational partnerships to develop sophisticated data and analytics infrastructures. He co-chairs the Research Data Alliance interest group on social dynamics of interoperable data, where he leads the discussion on the social dynamics required to facilitate interoperability across data infrastructures.