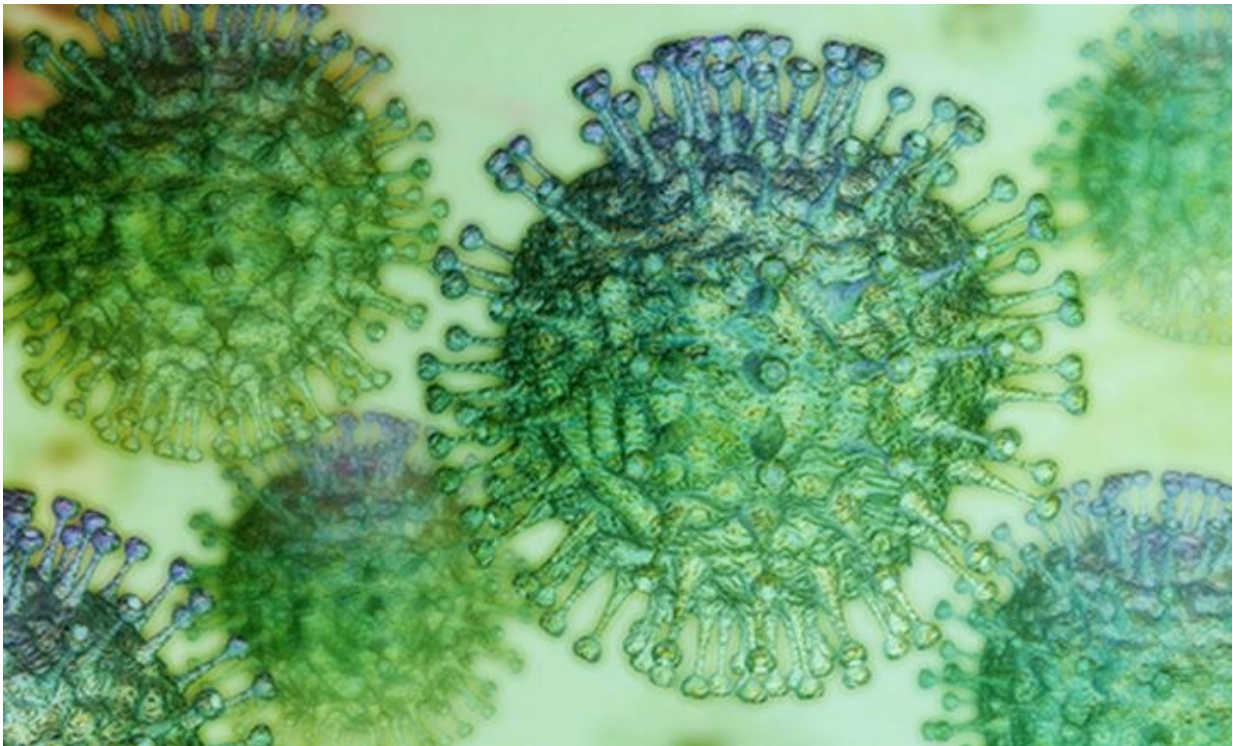


Building better coronavirus databases with automatic quality checks

April 17 2020



Coronavirus. Credit: European Centers for Disease Control

Amid a growing coronavirus crisis, experts in all fields have begun compiling massive datasets to track the impact of the contagion. These datasets capture everything from society-wide virus response information to medical needs data, available medical resources across the country, and buyer interest for medical equipment that could drive

financing for new production.

To make constructing these datasets as accurate and timely as possible, Prof. Michael Cafarella is leading an NSF-funded project that will build high-quality auxiliary datasets to enable automatic quality checking and fraud detection of the new data. These safeguards are imperative to making sure coronavirus decision-making is driven by clean, [accurate data](#).

Rapid analytical efforts by policymakers, scientists, and journalists rely on coronavirus data being complete and accurate. But like all [dataset](#) construction projects, those chronicling the coronavirus are prone to shortcomings that limit their effectiveness if left unaddressed. These issues include messy or unusable data, fraudulent data, and data that lacks necessary context.

Automatically checking coronavirus datasets against the pertinent, related datasets provided by Cafarella's team can make them more effective and insightful. For example, an auxiliary database about hospitals might contain data about the hospital's staff count, so a hospital resource allocator can test whether resources requested for coronavirus treatment are consistent with the level of staffing.

Cafarella's proposed datasets would be easy to combine with the fast-moving coronavirus data construction projects.

The team will build two large auxiliary databases. The Unified Medical Institution Auxiliary Database will be a database of all known United States medical institutions, and will include rich background information for quality-checking, as well as an easy method for data integration. The Unified Government Office Auxiliary Database will be a [database](#) of all known government offices in the United States, such as city halls, courts, or licensing offices, at any level of government. The team will release

both databases regularly, and the first release will be approximately one month after the project begins.

Provided by University of Michigan

Citation: Building better coronavirus databases with automatic quality checks (2020, April 17)
retrieved 26 April 2024 from

<https://medicalxpress.com/news/2020-04-coronavirus-databases-automatic-quality.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.