Evidence surveillance during the COVID-19 pandemic using automation and crowdsourcing

James Thomas
UCL Institute of Education, University College London, London, UK
james.thomas@ucl.ac.uk

The COVID-19 pandemic has generated a heightened appreciation of the importance of research to inform decisions. It has also highlighted major flaws in the way that findings from research can be utilised, with one article describing *how COVID broke the evidence pipeline*.[1] Over the past year, tens of thousands of empirical research and systematic review articles have been published on different aspects of COVID-19, many of which report findings that could ostensibly inform decisions about policy, practice, or future research commissioning. However, keeping track and making sense of this vast, heterogeneous, and fast-moving evidence base has tested the limits and capacity of current evidence surveillance systems, tools and workflows. One of the sessions at the 6th meeting of the International Collaboration for the Automation of Systematic Reviews (ICASR) in April 2021 featured some of the projects that have been addressing this problem, and five of them are presented as papers in this issue.

An important initiative to address information overload that features in one of the papers in this issue is the COVID-19 Research Dataset (CHORD-19). This was released in early 2020 by US technology companies with the objective of catalysing the computer science community into action to assist in datamining COVID-19 research papers. At the time of its launch, this dataset contained 28,000 research articles, but it has since grown to include nearly 600,000[2]. Researchers have analysed this dataset in many different ways, developing experimental information retrieval and extraction tools, and applications that answer questions and make conceptual linkages between papers. The "TREC-COVID task", which is the subject of the first paper in this issue, was developed to provide some structure to this experimentation. In this work, a gold standard dataset was created for biomedical researchers to use when developing and evaluating new information retrieval tools. Voorhees and Kanoulas describe the importance of this type of dataset and "task" in assisting the field to agree on common benchmarks for evaluating tool performance. Without such work, it is difficult to compare like-with-like, hampering the advance of the field.

With so many individuals and organisations responding to the pandemic at speed, some coordination of effort was required. The COVID-END initiative[3], facilitated by McMaster University, stepped in to fulfil this role, forming seven working groups to assist in planning. COVID-END is particularly concerned about making evidence available to decision-makers, and so the initiative hosts an inventory of "best evidence syntheses" and supports pre-registration of systematic review protocols in the PROSPERO database. However, this has not prevented widespread duplication of effort in pandemic responses, including those of the international evidence synthesis community. A recent paper[4] reported that hundreds of rapid reviews, systematic reviews and overviews have been published in response to the pandemic with considerable overlap in topic. Many are of poor quality, use unclear methods, and have discordant findings, which undermines the trustworthiness of the synthesised evidence base.

1 *How COVID broke the evidence pipeline (nature.com)*
2 *CORD-19: The Covid-19 Open Research Dataset (nih.gov)* and *CORD-19 Historical Releases (ai2-semanticscholar-cord-19.s3-us-west-2.amazonaws.com)*
3 *https://www.mcmasterforum.org/networks/covid-end*
4 *https://ebm.bmj.com/content/early/2021/06/03/bmjebm-2021-111710*

While many of the tools developed by the computer science community using the CHORD-19 dataset have used new natural language processing and machine learning techniques, most of the systematic and rapid reviews on COVID-19 appear to have used conventional manual methods. This may have been a missed opportunity, with a large opportunity cost. The remaining papers in this issue describe four initiatives that have used a combination of human and machine effort to 'map' the evidence. The first paper in this section, by Shemilt and colleagues, describes how one living map of COVID-19 research moved from searching conventional databases to find relevant studies (MEDLINE and Embase) to using a single, comprehensive source, based on web crawling technology (Microsoft Academic Graph). They found many more records using MAG and, in conjunction with using machine learning tools, this has made the workflow more efficient. One of the first maps of COVID-19 research to appear early in 2020 was produced by Keenan and colleagues. They used a novel automation tool – a Twitter Bot – to find and disseminate research and overcame bias in conventional English language search sources by collaborating with a team in China. The papers by Hair and Noel-Storr describe the application of crowdsourcing and automation to locate relevant studies. Hair and colleagues undertook significant custom software development to fine-tune a study identification and publication system in R. They also developed a detailed classification tool for coding records in detail based on full text reports. (Most other maps described research based on titles and abstracts alone.) Thanks to the use of automation for study identification, including the use of machine learning, and the coding being done by a crowd, they found this to be a sustainable workflow for keeping up with the evidence. Finally, Noel-Storr and colleagues describe the work of a major pre-existing "crowd" in assisting with pandemic response in Cochrane. Here, the crowd contributed to a range of evidence synthesis workflows, from identifying studies that could be relevant for specific reviews on COVID-19 to helping to maintain the Cochrane COVID-19 Study Register and other, study type specific, datasets. The team also used machine learning for some workflows and found that the pandemic enabled them to further understand the ways that crowdsourcing can contribute to maintaining a surveillance of the evidence base.

The papers thus describe considerable innovation in maintaining databases of research on COVID-19. One other presentation at ICASR featured the Epistemonikos database, which also uses a combination of human and machine effort, to identify research on COVID-19. These databases are not the only activities concerned with identifying and "mapping" COVID-19 research of course. For example, the World Health Organisation maintains its own database, which has evolved over the course of the pandemic. As with the example of systematic review production highlighted above, there appears to be some duplication of effort in tool development and data curation across the various COVID-19 databases of research. While this may have been difficult to avoid in every case, due to the rapidity with which organisations needed to react and the different users and funders they supported, some lessons can still be drawn.

First, while these projects have indeed used novel methods and tools, it is worth observing that they have not been working from CHORD-19 or using the TREC-COVID data for evaluation as outlined by Voorhees and Kanoulas. Thus, even though the field of evidence synthesis has been struggling to cope with the "infodemic", it does not appear to have taken advantage of the innovation emerging from the information and computer science community; and perhaps one of the key messages of the session at the ICASR meeting was that the evidence synthesis and computer science communities have been engaged in similar, but parallel tasks. The lack of collaboration is striking, and worthy of more detailed examination and reflection. It may be that the evidence synthesis community does not find the tools developed by computer science to be suitable for its work; or it might be simply unaware of the potential of existing tools. Whichever is the case, more collaboration is likely to result in more efficient working practices.

Second, each database contains similar, but different, classification schema for describing the studies they contain. To some extent these reflect different perspectives and organisational objectives. However, it is also

clear that the same records have been examined by multiple different people across the projects, with similar (and in some cases, precisely the same) classifications being applied. Reducing duplication of effort between two or more ongoing database projects is not straightforward, as sharing workflows requires coordination and agreement on classification schema, mutual trust in quality assurance standards, and the ability to integrate data across tools. Work to facilitate better data sharing has therefore been discussed in the COVID-END working groups, and the COVID-19 Knowledge Accelerator project[5] led by Brian Alper has been working throughout the pandemic to develop standards for the detailed description of COVID-19 evidence.[6]

Third, it is important to bear in mind that, sometimes, the most effective way to avoid duplication of effort is simply to stop and leave the work to others. After being one of the first maps to appear in the early stages of the pandemic, while others were still establishing workflows, this is the decision that Keenan and colleagues made. They could see that other groups had more sustainable production models, and they decided to cease work on their map and to focus their effort elsewhere.

All the work described in these papers has had impact, with the tools and datasets being used globally in response to the pandemic. Some of the lessons learned are already bearing fruit in the various tools and workflows described. Combining human effort and automation has been of demonstrable value in helping us to keep pace with such a huge volume of research; and further reflection on what has worked, and what can be improved, will help the field to continue to innovate in this area.

[5] https://confluence.hl7.org/pages/viewpage.action?pageId=97468919
[6] https://www.sciencedirect.com/science/article/pii/S1532046421000149?via%3Dihub