COVID Information Commons (CIC) Research Lightning Talk

Transcript of a Presentation by Florence Hudson (Columbia University), June 11, 2024

Title: *CIC-E: COVID Information Commons Extension for Pandemic Recovery*

Award CIC Database Profile

NSF Award #: 2139391

Youtube Recording with Slides

Spring 2024 CIC Webinar Information

Transcript Editor: Lauren Close

Transcript

Florence Hudson:

Slide 1

Well, thank you everybody for joining. It's always wonderful to hear you listening and hear you asking questions, so please feel free to drop them in the chat. As Lauren said, we like to do them at the end because sometimes a question will actually refer to multiple of the presentations.

So the COVID Information Commons started when COVID started, pretty much. The National Science Foundation contacted the Northeast Big Data Innovation Hub in March of 2020 and asked if we could create an open portal for people to easily find NSF-funded RAPID awards. Those are the awards that the government gives out when there's a crisis. So NSF has RAPIDS, I believe NIH has RAPIDS, and so at the time, we went into the simple search of NSF and there were exactly 32 awards. I'll never forget this. And today, we have over 13,000. None of us thought it would last this long and none of us thought it would get this big. It's just going to keep going as we know. So we were really delighted to be able to help with this and to create a FAIR portal to make research Findable, Accessible, Interoperable, and Reusable. That's how the CIC got started.

We got the award in March and it was a RAPID award so it's "hurry up!" They contacted us in March [2020], we got the money in May [2020], and we launched our MVP in July. It was really very quick and at the time we invited two researchers to present in our kickoff webinar so that it wasn't just administrative and 40 more researchers asked if they could present. We were like "whoa!" So on our first call, we said, "ok, welcome to the COVID Information Commons community, we'll keep presenting to

each other and working together until we're done." And we're not done yet, so we keep it as an open resource to explore research addressing the COVID-19 pandemic and the longitudinal impacts, which, as we know, will be forever. You can search the Awards & PI, or Principal Investigator, Database. We have a really cool Machine Learning Map clustering tool we'll show you. The initial 2020 award was expanded with an extension award in 2021, which goes through October of 2025. We'll see if they still want us to keep doing this - we'll see. But we actually are archiving in Dryad, for those of you familiar with Dryad, so it has longitudinal life.

Slide 2

Next slide please. And as Lauren said, it was launched by the Convergence Accelerator, which is not the TIP directory, Technology, Innovation and Partnerships Directory, at NSF. We're very grateful for their support.

So the CIC community started from nothing in 2020 just like everything COVID-related or COVID-19 related. The community has actually grown to over 3,600 individuals in over 800 organizations across the U.S. and 37 other countries. In this past year, we had a 25% increase in members. I think a lot of that has to do with the great work Lauren has done - I shouldn't talk too much about this, but she has started the COVID Information Commons Researcher Working Group and Student Working Group. There are these very interesting data visualization and data science projects students can do and it really has grown the community a lot. The good news is that even though COVID-19 is not a wonderful thing to be living through or have lived through now, the students are getting more involved in understanding what we did learn from the data. How do we leverage the data? What we're hoping for is that in the future when they get to worry about this when there is COVID or whatever comes next, that they will have some skills and tools that will enable them to help us. They're going to be our future. So the community continues to grow.

Slide 3

Next slide please. So we were working with NSF and we're very grateful for their support and funding, but this is a disease so we felt like we should be working with the NIH as well. NSF agreed and said that's a good point. That's when they asked us if we would expand the CIC and so we got this CIC Extension award. THey gave us another $2 million for four years. Since then, we went from 732 NSF awards to now over 13,000 NSF and NIH awards that are in the corpus.

Slide 4

We update it monthly, so this is a view if you were to go in on our main webpage you could see the COVID Awards and PI Database. You would go in here and put someone's name or a key phrase like "epidemiology" or "bats" or whatever. We have actually had presentations on bats! So this has a lot of awards in it. There's information on the researchers you can click on. From the NSF ones you can click through directly to their award on the NSF website, with NIH it brings you to the reporter tool and then you get to put in the information. There are also PI profiles. For some of the Principal Investigators who

choose to fill that in, and as you can see, there's a faceted search so you can search by funder (which is NSF or NIH), search by organization, a bunch of other things.

Slide 5

This is the Machine Learning Maps tool. This is powered by Lingo 4G from Carrot Search in Poland. They actually take the 13,000 awards and then cluster them by topic so it gives you a nice colorful view and allows you to find people that might be doing research in your area of interest or the interest that you're actually researching. It allows you to look by funder, allows you to look by institution, it allows you to look by PI name so that you can find the humans and institutions you want to collaborate with in your topic of interest.

Slide 6

Next slide please. So after the Machine Learning Maps tool, I wanted to share about some of the other research resources we have. We have over 80 open source COVID-19 data sets. Lauren just did a scan to make sure that they haven't been archived - we have had some data sets archive, as you can imagine, related to COVID. We try to keep an eye on that so we just took another scan through it this week. We have a listing of Publications, Research Groups and Guides, and then Research Funding listings. There are still some funding resources as we look at Long COVID and SARS-CoV-3, as we're going to hear about.

Slide 7

Next slide please. We do these quarterly webinars. We started out with monthly talks and we gave people 5-7 minutes because they hadn't done any research yet in 2020, but they wanted to talk about it to find each other. Now we're quarterly and we offer 10-12 minute talks because people have actually done the research so we get to learn from each other. We've had over 130 individual presentations to date from NSF, NIH, and CDC researchers. If you know of any researchers that have been funded by these government agencies, please let us know or have them contact us. We'd be happy to have them present on a future webinar.

Slide 8

Next slide please. We also have a Meet the Researchers tab on the website so you can browse the 130+ research lightning talks. Lauren has actually categorized them by theme: COVID and Biology, COVID and Data, etc. If you wanted to subset them more easily, we also have the presentations, not just the video, but Lauren and the student team actually transcribe them into written English. These are not just the quasi translations or transcriptions that are done on the internet, but the team actually vets each one and then we have students that have translated them into Spanish and French. We're going to be translating them into Hindi as well, which increases accessibility for people who may think in one language, learn in another language, speak in a different language. We really try to make it accessible. Meet the Researchers includes the PIs, such as those on the call with us today, NSF, NIH, and CDC, but also the students who win the Paper Challenges. The first one that you see on this screen is Aditya Kulkarni. He was actually one of the winners of the Student Paper Challenge the first year that we had it. He was able to present on a webinar like this and we have his material. All of these are uploaded into the

Columbia Academic Commons so they get a digital object identifier, a DOI for posterity, with the researcher's permission.

Slide 9

Next slide please. We really do try to make this very FAIR - Findable, Accessible, Interactive, and Reusable. We put all of this on YouTube and we have little research spotlights where people get a taste of, for example, how did COVID impact science education? It then brings you to that video, again with these transcriptions and translations. It's a fun little less than one minute way to see what's going on and then click into a video that you might be interested in.

Slide 10

Next slide please. Another exciting thing is that I actually publish with Springer. I've done a couple of books with Springer Nature, you may be familiar with them. The books I've previously done were on a new standard that was just approved by IEEE this past week on security and privacy called TIPPSS for the Clinical Internet of Things. We're doing a new book with Springer which is going to be with the speakers, the PIs who presented on these researcher lightning talks who are NSF or NIH funded researchers. They're going to be writing chapters that talk about their projects, funders, goals of their research, the outcomes, and therefore what their recommendations are regarding the mitigation of future pandemics. The reason we're doing this is because when NSF first contacted us and said we want you to create this portal so people could find COVID research, I went looking for COVID research. I saw that the coronavirus has been around for decades but it was very hard to find anything. I feel like it's our duty to document this better so that people can actually find what we've learned and not have to start from scratch. We're excited about this - they're writing their chapters, I just got the second version of one of the chapters today. They'll be writing them this year and we plan to publish at the beginning of next year. We're excited that Springer Nature is supporting this and NSF is totally on board with it, which we're grateful for. Then we're able to give all these PIs and these authors a voice. There are about 30 chapters from about 100 authors. We really encourage them to have all their researchers involved, their students involved if they want. I love giving a voice to all of these researchers and so that's what we're going to be doing. We're looking forward to that coming out and you'll all find out about it when it gets published.

Slide 11

Next slide please. As I was mentioning and as Lauren briefly mentioned, we have a CIC Research working Group. If you'd like to join, we've been talking about doing a Long COVID project, we haven't really kicked it off yet, if you have any ideas on joint research you would like to do, please let us know. You can join us, that's always on our homepage.

Slide 12

Next slide please. Then we have the Student Working Group, which has really grown. We're really excited that students want to learn how to get information out of this data, find knowledge and insights for today and the future. We have over 600 members from 230 organizations across the U.S. and six other countries. They can participate in the projects that have already been finished from the last two

semesters. These projects are available on our Data Science Projects page on the Hub website as well as the CIC website. These projects are on data visualization and analyzing pandemic policies. They can use datasets from the COVID Information Commons, but the Analyzing Pandemic Policies project actually uses an Oxford University database, very interesting. We always use open datasets so anybody on the planet can leverage them. There will be a new project in the fall as Lauren mentioned.

The Portfolio and Network Building Group is a really cool opportunity for the students to meet each other. It could be that they can help each other think through their research or meet at a college they're planning to do their Master's degrees at, whatever it is, so that we can collaborate together. We're the Northeast Big Data Innovation Hub so our job is to be a collaboration hub and a catalyst for innovation. This really helps us do that with the students leveraging each other's content and knowledge, which is really exciting. Please invite any of your students to the networking session, the next of which will be on June 28th.

Slide 13

Next slide please. Lauren mentioned the CIC Student Paper Challenge that occurs annually. The first submissions are due July 31st so students can just go to the website. There's a bitly here. For the Student Paper Challenge, they can enroll in that and then they'll get reminders to submit the paper. Then we actually have incentives that we were able to bake into the grant. The first place prize is $400, if it's a team award, they share it. Second and third place also receive prizes. There is an undergraduate and community college cohort as well as a graduate cohort. It's global - one of our previous winners was from South Africa and she was invited to speak at the Academic Data Science Alliance conference in Texas this past year. It's a wonderful professional development opportunity for students, so please encourage them to participate.

Slide 14

Next slide please. Actually you can be a judge or mentor too - go to the website and sign up to be a judge or a mentor. We're very interested in making our content as accessible as possible. As I was mentioning, it's available on YouTube, there are translations done and we also put the information on the Columbia Academic Commons. If you know of any students or anyone that might be interested in translating our materials into another language or the languages that we currently serve, please let us know. We have a REAL program, or Research, Experience, and Leadership opportunities, and we love when the whole community participates.

Next slide please. So that's it! Thank you for having me, thank you for joining us, thank you for being a part of the CIC. You're all part of my family now so stay in the loop for upcoming events. You can join our mailing list, we send out a monthly newsletter. If there's anything else we can do for you, we'd love to know. Thank you, Lauren.