

## [COVID Information Commons \(CIC\) Research Lightning Talk](#)

Transcript of a Presentation by Dominique Duncan (University of Southern California), January 2021



Title: [COVID-ARC \(COVID-19 Data Archive\)](#)

[Dominique Duncan CIC Database Profile](#)

NSF Award #: [2027456](#)

[Youtube Recording with Slides](#)

[January 2022 CIC Webinar Information](#)

Transcript Editor: Shikhar Johri

---

### Transcript

डोमिनिक डंकन:

स्लाइड 1

ठीक है, इसलिए आपको फ्लोरेंस और लॉरेन और कोविड इंफो कॉमन्स टीम के पीछे सभी को धन्यवाद। मैंने पहली बार डेढ़ साल पहले एक बात की थी और इस बात के अंत में मैं कुछ अपडेट दूंगा कि इसके परिणामस्वरूप क्या हुआ और एक नया सहयोग जो बनाया गया था। लेकिन मैं हमारे COVID-19 डेटा आर्काइव, Covid-CHR के बारे में शॉर्ट के लिए बात करना चाहता था। मैं न्यूरोइमेजिंग की प्रयोगशाला में केक स्कूल ऑफ मेडिसिन में दक्षिणी कैलिफोर्निया विश्वविद्यालय में एक सहायक प्रोफेसर हूँ।

स्लाइड 2

इसलिए, हमारी प्रयोगशाला में हमारे पास बड़े पैमाने पर, बहु-मॉडल डेटा अभिलेखागार का निर्माण करने का बहुत अनुभव है, मुख्य रूप से मस्तिष्क डेटा के लिए। लेकिन महामारी की शुरुआत में, हमने सोचा कि हम उन सभी डेटा अभिलेखागार के साथ अपने अनुभव और संसाधनों का उपयोग करने और COVID-19 डेटा संग्रह विकसित करने में सक्षम हो सकते हैं। इसलिए हमें प्राप्त हुआ - हमें कोविड -आर्क नामक इस डेटा संग्रह को विकसित करने के लिए एक एनएसएफ रैपिड अवार्ड से सम्मानित किया गया। तो हम जो करते हैं, हम विभिन्न प्रकार के COVID-19 डेटा के साथ-साथ संसाधनों को भी एकत्र करते हैं और हमने नेटवर्क और केंद्रीकृत अभिलेखागार का एक मंच बनाया है जो कि स्टोर, क्यूरेट, विजुअलाइज़ और मल्टीमॉडल कोविड -19 डेटा का प्रसार करता है। हमारे पास दुनिया भर से बहुत सारे डेटा सेट हैं। उनमें से कई सार्वजनिक रूप से उपलब्ध हैं, कुछ निजी हैं, और इसलिए हमने उन डेटा प्रदाताओं के साथ काम किया है जो डेटा उपयोग समझौतों के साथ आने के लिए हैं जो उनकी आवश्यकताओं के अनुरूप हैं। और हमारे पास वेबसाइट पर मेटाडेटा उपलब्ध है, इसलिए यदि उपयोगकर्ता उस प्रक्रिया तक पहुंचने का अनुरोध

करना चाहते हैं, तो हम उस प्रक्रिया को सुविधाजनक बनाते हैं। लेकिन तब डेटा प्रदाता वह अंतिम निर्णय लेते हैं। और बहुत सारे डेटा USC में हमारी साइट पर संग्रहीत किए जाते हैं, लेकिन कुछ डेटा सेटों के लिए वे उस साइट पर संग्रहीत होते हैं जहां डेटा एकत्र किया गया था और हमारे पास मेटाडेटा उपलब्ध है ताकि लोग इसे देख सकें। और हमारे बहुत सारे काम मेटाडेटा के सामंजस्य के आसपास रहे हैं और यह एक साइट पर ध्यान केंद्रित करने के बजाय विभिन्न साइटों पर विभिन्न प्रकार के विश्लेषण करने के लिए लोगों के लिए अलग-अलग प्रकार के विश्लेषण करने में आसान बनाने के लिए पूल्ड कॉहोर्ट्स पर शोध को सुविधाजनक बनाने के लिए है। एक दो स्लाइड्स में, मैं उन कुछ चुनौतियों के बारे में बात करूंगा और हम इसे सामंजस्य बनाने पर क्यों काम करते हैं। और हमने विजुअलाइज़ेशन और क्वालिटी कंट्रोल टूल और एनालिटिक टूल भी एकीकृत किया है। फिर से, शोधकर्ताओं की मदद करने के लिए - सिर्फ COVID-19 पर शोध में तेजी लाने के लिए। और डेटा संग्रह और सामंजस्य के आसपास के सभी कार्यों के अलावा, हम हमारे पास मौजूद डेटा पर विभिन्न प्रकार के विश्लेषण भी कर रहे हैं और हम COVID-19 के विभिन्न पहलुओं का अध्ययन करने के लिए प्रतिक्रिया सिद्धांतों और डेटा विज्ञान का उपयोग कर रहे हैं।

### स्लाइड 3

इसलिए, डेटा के संदर्भ में हमारे पास अलग-अलग प्रकार के डेटा हैं - इसका बहुत से हिस्सा केवल सीटी छवियों के साथ-साथ एक्स-रे पर केंद्रित है, लेकिन हमारे पास नैदानिक डेटा भी हैं जिसमें लक्षण, विटाल, कॉमोरिडिटीज, जनसांख्यिकी, रोगी इतिहास, जियोलोकेशन शामिल हैं। हमारे पास भी है- इमेजिंग के लिए हमारे पास अल्ट्रासाउंड और एमआरआई के साथ-साथ कुछ ईईजी डेटा भी हैं। और फिर हमने फेफड़े के मुखौटे, संक्रमण मास्क और रेडियोलॉजिस्ट एनोटेशन भी प्रदान किए हैं। और यहां आप देख सकते हैं कि हम IBM के HIPAA के अनुरूप उच्च-गति वाली फ़ाइल ट्रांसफर सिस्टम का उपयोग करते हैं जिसे Aspera कहा जाता है और यह डेटा प्रदाताओं के लिए Covid-CHR को डेटा को स्थानांतरित करने के साथ-साथ उपयोगकर्ताओं को Covid-CHR से डेटा डाउनलोड करने के लिए एक बहुत आसान तरीका है। कंप्यूटर।

### स्लाइड 4

और अभी हमारे पास दुनिया भर से 28 डेटा सेट हैं। जैसा कि आप कल्पना कर सकते हैं कि इन अलग-अलग डेटा सेटों में कुछ असंगत फ़ाइल नामकरण है, असंगत मेटाडेटा प्रारूपण, इन्फ्रास्ट्रक्चर के भंडारण में अंतर, साथ ही साथ उन डेटा सेटों में अन्य अंतर भी हैं। इसलिए हमने जो किया है, उसने इन सभी को एक साथ एक केंद्रीकृत डेटा आर्काइव में रखा है और हमने यह सुनिश्चित किया है कि लगातार फ़ाइल नामकरण और संगठन, सुसंगत मेटाडेटा प्रारूपण, और Aspera का उपयोग करके एक स्थान से कई डेटा सेट डाउनलोड करने में आसानी हो।

### स्लाइड 5

और यहाँ हमारे पास मौजूद डेटा के हिस्से का एक स्क्रीनशॉट है। मैं एक स्लाइड पर सब कुछ फिट नहीं कर सकता था, लेकिन आप बस यह देख सकते हैं कि क्या आप covid-arc.loni.usc.edu पर जाते हैं, आप पा सकते हैं कि हमारे पास क्या डेटा है। और यह आयोजित किया जाता है कि आप देख सकते हैं कि हमारे पास साइट नंबर है, वह स्थान जहां डेटा एकत्र किया गया था, तौर-तरीके, फ़ाइल प्रारूप, कोई भी मेटाडेटा जो हमारे पास है, उसके अलावा, और फिर कितनी छवियां हैं। और उनमें से कुछ कोविड के बीच विभाजित हैं और कोविड नहीं। और फिर हमारे पास यह भी जानकारी है कि यह सार्वजनिक रूप से उपलब्ध है या नहीं।

## स्लाइड 6

इसलिए अब, मैं सिर्फ उन कुछ परियोजनाओं को उजागर करना चाहता था, जिन पर मेरे छात्र काम कर रहे हैं, वे बहुत उत्पादक रहे हैं और वास्तव में रोमांचक शोध कर रहे हैं। मेरे पास अंडरग्रेड्स के लिए कुछ NSF अनुसंधान अनुभव हैं और अक्ष गर्ग उनमें से एक हैं। जब वह हाई स्कूल में था तब उसने शुरू किया था और अब वह एक अंडरग्रेजुएट है - स्टैनफोर्ड में प्रथम वर्ष के अंडरग्राउंड और पिछले हफ्ते ही उनके पेपर को एप्लिकेशन के साथ विशेषज्ञ प्रणालियों में स्वीकार किया गया था और यहां उन्होंने कोविड बनाम कोविड बनाम को अलग करने के लिए 40 कन्व्यूशनल न्यूरोल नेटवर्क आर्किटेक्चर की तुलना की। कोविड नहीं और उन्होंने पाया कि सबसे अच्छा मॉडल कुशलनेट-बी 5 में अत्यधिक उच्च सटीकता संवेदनशीलता और विशिष्टता प्राप्त हुई। और मॉडल भी नैदानिक रूप से प्रासंगिक विशेषताओं जैसे कि ग्राउंड ग्लास ओपेसिटी और समेकन पर निर्भर करता है, जो अक्सर कोविड 19 रोगियों में देखे जाते हैं।

## स्लाइड 7

एक अन्य परियोजना - एलेक्स ब्रुकहॉस एक अन्य आरईयू अवार्डी है, इसलिए यह काम जर्नल ऑफ इमिग्रेंट एंड माइनॉरिटी हेल्थ में प्रकाशित हुआ था और यहां वह और अन्य छात्र कैलिफोर्निया में टीकाकरण की गतिशीलता को देख रहे थे। इसलिए उन्होंने सोशल वलनरेबिलिटी इंडेक्स और एसवीआई नामक कुछ को देखा - उन्होंने सामाजिक आर्थिक स्थिति, घरेलू रचना और विकलांगता, आवास प्रकार और परिवहन और अल्पसंख्यक स्थिति और भाषा सहित चार एसवीआई विषयों को देखा। और उन्होंने पाया कि सबसे कम टीकाकरण कवरेज उच्च भेद्यता समूहों में था। अल्पसंख्यक स्थिति और भाषा ने कम और उच्च भेद्यता काउंटियों के बीच कवरेज में सबसे बड़ी असमानता प्राप्त की। इसलिए मुझे लगता है कि यह वास्तव में महत्वपूर्ण काम है, खासकर जब हम अधिक से अधिक आबादी को प्राप्त करने की कोशिश कर रहे हैं।

## स्लाइड 8

एक और, एक और पेपर जो प्रकाशित एलेक्स ब्रुकहॉस और अन्य छात्रों को प्रकाशित किया गया था, फिर से शुरू होने के बाद लॉकडाउन संक्रमण दरों को देख रहा था। इसलिए उन्होंने पिछले साल उच्च कोविड -19 केस काउंट्स के साथ संयुक्त राज्य भर में 83 काउंटियों को देखा। और वे विभिन्न प्रकार के व्यवसायों को देख रहे थे और वे एक पूर्ण फिर से खोलने या आंशिक रूप से फिर से खोलने के बीच अलग हो गए और वे उन पुनः व्यवस्थित होने से पहले और बाद में संक्रमण दर में बदलाव को देख रहे थे और यह देखते हुए कि संक्रमण दरों में वृद्धि में कौन से व्यवसायों का सबसे अधिक प्रभाव पड़ा। और इसलिए बार और जिम ने एक महत्वपूर्ण भूमिका निभाई - उसमें एक बड़ी भूमिका।

## स्लाइड 9

युजिया झांग, जो इस परियोजना में मेरे प्रोजेक्ट असिस्टेंट हैं, उन्होंने पिछले साल एक रिव्यू पेपर किया था जो कोविड -19 के साथ ब्लड टाइप एसोसिएशन को देख रहा था। इसलिए उसने 23 अध्ययनों को देखा, जिसमें एक जोखिम और सुरक्षात्मक कारकों दोनों के रूप में रक्त प्रकार का अवलोकन था, कैसे - कैसे कुछ रक्त प्रकारों के साथ कुछ रक्त प्रकारों के लोग सकारात्मक और गंभीरता के नैदानिक परिणामों के परीक्षण के लिए अति-संवेदनशील होते हैं। और वह आनुवंशिक संघों और संभावित अंतर्निहित आणविक तंत्रों के माध्यम से भी चली गई।

## स्लाइड 10

और फिर अज़रीन खान, जो पिछले दो ग्रीष्मकाल में एक आरईयू साथी रहे हैं, वह इस परियोजना पर काम कर रही हैं जो थ्रेशोल्ड-आधारित फेफड़े के विभाजन को देख रही हैं। तो यह मौजूदा तरीकों की तुलना में बेहतर प्रदर्शन के साथ फेफड़े की असामान्यताओं को निर्धारित करने के लिए मल्टी-स्टेप थ्रेसहोल्डिंग विधि है।

## स्लाइड 11

और जब से मैं समय से बाहर चला रहा हूं, मैं इस के माध्यम से भागने जा रहा हूं, लेकिन मैं एक सहयोग के बारे में बात करना चाहता था जो उस पहली कोविड जानकारी कॉमन्स की बात करता है जो मैंने दिया था और माइकल पाज़ानी और अल्बर्ट हसियाओ यूसीएसडी से, सैन डिएगो से, उन्होंने बातचीत भी दी और हमने उसके बाद एक सहयोग शुरू किया। हमने एक स्मार्ट हेल्थ प्रस्ताव भी प्रस्तुत किया, जो वित्त पोषित नहीं हुआ, लेकिन हमने पिछले नवंबर में इसे फिर से शुरू किया, इसलिए हम इस बारे में इंतजार कर रहे हैं। लेकिन यह केवल एक आउटरीच वेबिनार का एक स्क्रीनशॉट है जो हमारे पास दक्षिणी कैलिफोर्निया में हाई स्कूल के छात्रों के लिए था, और सामूहिक रूप से हमारे छात्रों ने अपनी परियोजना के बारे में बिजली की बातचीत दी और यह बहुत सफल था। और मैं सिर्फ आपको धन्यवाद देना चाहता था। यह लैब वेबसाइट कोविड-आर्कवेबाइट्स है, कृपया मुझे ईमेल करें [duncand@usc.edu] यदि आपके कोई प्रश्न हैं और फंडिंग के लिए NSF और NIH दोनों को धन्यवाद।