

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

Transcript of a Presentation by Murat Kantarcioglu (University of Texas at Dallas), October 16, 2020



Title: [RAPID: Collaborative: A Privacy Risk Assessment Framework for Person-Level Data Sharing During Pandemics](#)

[Murat Kantarcioglu CIC Database Profile](#)

NSF Award #: [2029661](#)

[YouTube Recording with Slides](#)

[October 2020 CIC Webinar Information](#)

Transcript Editor: Shikhar Johri

Transcript

स्लाइड 1

अच्छा- मुझे लगता है कि अब कुछ के लिए शुभ दोपहर है। मेरा नाम मुरत कांटार्सिओग्लू है। मैं डलास में टेक्सास विश्वविद्यालय में कंप्यूटर विज्ञान का प्रोफेसर हूँ। यह वेंडरबिल्ट मेडिकल स्कूल के प्रोफेसर ब्रैंड मालिन के साथ महामारी के दौरान डेटा साझा करने के लिए गोपनीयता जोखिम मूल्यांकन पर हमारा रैपिड काम है।

स्लाइड 2

इसलिए आप में से कई लोग समाचारों का पालन करते हैं, जैसे-जैसे COVID-19 आगे बढ़ता है, इस संबंध में नए प्रश्न सामने आ रहे हैं कि क्या यह दौड़ से संबंधित है, क्या उदाहरण के लिए धूम्रपान COVID-19, बुरे परिणामों आदि के साथ आपके जोखिम को बढ़ाता है। इसलिए, इस तरह के सवालों के जवाब देने के लिए चाहे वह नस्ल और COVID-19 मृत्यु दर या अन्य कारकों जैसे धूम्रपान उच्च, रक्तचाप का संबंध हो। स्पष्ट रूप से, हमें अधिक डेटा एकत्र करने की आवश्यकता है और निश्चित रूप से इस डेटा का विश्लेषण करने की आवश्यकता है।

स्लाइड 3

लेकिन दुर्भाग्य से, या सौभाग्य से कुछ अर्थों में, गोपनीयता और गोपनीयता की आवश्यकताएं इस प्रयास को धीमा कर सकती हैं यदि आप जुलाई में प्रकाशित इस विज्ञान लेख को देखते हैं, तो आप देखेंगे कि कैलिफोर्निया साझा नहीं करना चाहता था- कैलिफोर्निया राज्य के अधिकारी कुछ शोधकर्ताओं के साथ COVID-19 डेटा विवरण साझा नहीं करना चाहते थे, कुछ निश्चित शोधों में बाधा डालते हैं।

स्लाइड 3

इसलिए, एक अर्थ में वे सही हैं क्योंकि यह ज्ञात है कि साझा डेटा को आसानी से फिर से पहचाना जा सकता है। यह पिछले कई वर्षों पहले दिखाया गया है कि यदि आप अलग-अलग डेटा सेटों को जोड़ते हैं, उदाहरण के लिए, यदि आपके पास एक COVID-19 डेटा है जो ज़िप कोड, आयु और लिंग की रिपोर्ट करता है, तो इसे जल पंजीकरण सूचियों से जोड़ने और उन लोगों की पहचान करने का अधिक जोखिम है जो COVID-19 रोगी डेटा में हैं। और निश्चित रूप से, यह थोड़ी देर के लिए जाना जाता है, इसलिए यह एक नई समस्या नहीं है और इन जोखिमों को समझना पिछले बीस वर्षों में सक्रिय शोध रहा है। लेकिन फिर भी, COVID-19 ने कई चीजें बदल दीं।

स्लाइड 4

एक बात यह है कि अब हमारे पास, विशेष रूप से सोशल मीडिया के साथ, हमारे पास अलग-अलग पुनः पहचान स्रोत हैं। यह द न्यूयॉर्क टाइम्स का एक उदाहरण है जहां उन्होंने अपने जीवन की कहानियों में COVID-19 के कारण मरने वाले कई रोगियों को प्रोफाइल किया।

स्लाइड 5

इसलिए, यदि आप इस हमले के परिदृश्य पर वापस जाते हैं, तो अब आपके पास अतिरिक्त स्रोत हो सकता है जहां आप जानते हैं कि कुछ लोगों का निधन हो गया है और उनकी जानकारी और आप इसका उपयोग लक्ष्यों को हटाने और पहचान हमलों को फिर से चलाने के लिए कर सकते हैं।

स्लाइड 6

इसलिए यह रैपिड वास्तव में महामारी डेटा साझाकरण, विशेष रूप से पुनः पहचान जोखिमों के कारण अद्वितीय चुनौतियों को समझने की कोशिश कर रहा है। इसलिए, एक बात जो हम महसूस करते हैं वह यह है कि पिछले काम के विपरीत, धूम्रपान या दौड़ जैसे साझा किए जाने वाले गुणों को समय के साथ बदल दिया जा सकता है। इसलिए, हमें वास्तव में इसे समझने की आवश्यकता है। इसके अलावा, केस नंबर बदल रहे हैं, जिसका अर्थ है कि कुछ गोपनीयता जोखिम समय के साथ बदल सकते हैं, इसलिए हमें समय के साथ गोपनीयता जोखिमों को वास्तव में अपडेट करने की आवश्यकता है क्योंकि नए डेटा स्रोत दूसरों के साथ साझा किए जा रहे हैं। इसलिए महामारी डेटा साझाकरण का यह तेजी से बदलता और अनिश्चित व्यवहार जोखिम मूल्यांकन ढांचे को प्रभावित करता है, और इसी कारण से हम अब मौजूदा मॉडलों को विकसित और अद्यतन कर रहे हैं।

स्लाइड 7

इसलिए, हमारा लक्ष्य वास्तव में यह देखना है कि डेटा गोपनीयता लक्ष्यों को देखते हुए हम सबसे अच्छी उपयोगिता क्या प्राप्त कर सकते हैं, जो इस मामले में यह सुनिश्चित करने के लिए कि पुनः पहचान जोखिम कम है, दूसरे शब्दों में लोगों को फिर से पहचानना बहुत कठिन है। और यह भी, डेटा उपयोगिता लक्ष्यों को देखते हुए हम सबसे अच्छी संभव गोपनीयता क्या प्रदान कर सकते हैं? और फिर, हमारा लक्ष्य यथासंभव सटीक जानकारी देना है। दूसरा विकल्प खेल में आता है क्योंकि यह तर्क है कि, आप जानते हैं, हमें वास्तव में बीमारी का मुकाबला करने पर ध्यान केंद्रित करना चाहिए ताकि डेटा गोपनीयता माध्यमिक हो सके। फिर भी, हमारे पिछले शोध से पता चला है कि डेटा साझा करने के संबंध में आपके लक्ष्यों को देखते हुए, विभिन्न विकल्पों के अलग-अलग गोपनीयता परिणाम हो सकते हैं। इसलिए, यह समझना महत्वपूर्ण होगा। बेशक, प्रमुख दिशा होगी- डेटा उपयोगिता क्या है और इसे कैसे परिभाषित किया जाए और यह हमारा चल रहा शोध है।

स्लाइड 8

निष्कर्ष निकालने के लिए, हमने पहचान जोखिम का विश्लेषण करने के लिए इस पर अपना पिछला काम बनाया और हम महामारी के कारण समय-भिन्न और बदलती जानकारी की इस अनूठी चुनौती के आधार पर तकनीकों को अद्यतन और विकसित करते हैं। और जल्द ही, उम्मीद है, एक या एक महीने की तरह, हमारे पास स्वास्थ्य अधिकारियों का मार्गदर्शन करने के लिए सामान्य डेटा साझाकरण नीतियां देने पर एक पेपर होगा जैसे कि कम जोखिम के साथ कौन सा डेटा साझा किया जा सकता है, उच्च जोखिम विकल्प क्या हैं, और इसका प्रभाव। तो कृपया हमें फॉलो करें यदि आपको डेटा साझाकरण और महामारी डेटा की गोपनीयता के बारे में कोई चिंता है। तो, मैं यहाँ बंद कर दूँगा मुझे लगता है कि समय के साथ थोड़ा सा। धन्यवाद।