

[Centro de Información de COVID \(CIC\): Charlas científicas relámpago](#)

Transcripción de una presentación de Peter Rose (UC San Diego), 5 de mayo de 2020



Título: [COVID-19-Net: Integración de datos de salud, patógenos y ambientales en un gráfico de conocimiento para el seguimiento, análisis y pronóstico de casos](#)

[Perfil de Peter Rose en la base de datos del CIC](#)

Subvención de La Fundación Nacional de Ciencias (NSF, por sus siglas en inglés) #: [2028411](#)

[Grabación de YouTube con diapositivas](#)

[Información del seminario web del CIC de Julio 2020](#)

Editor de la Transcripción: Rhyley Vaughan

Traductor: Kenia Pujols

Transcripción

Diapositiva 1

De acuerdo. Gracias, Florence, por invitarnos. Déjenme contarles un poco sobre los antecedentes de COVID-19-Net. Fuimos parte del programa de red de conocimiento abierto de la NSF. En particular, nos interesaba vincular tres tipos de datos: datos biomédicos, datos ambientales y datos sociodemográficos. Luego, en enero llegó el COVID. Pensamos que este es realmente un gran caso de uso donde necesitamos integrar datos multidisciplinarios, así que si podemos ir a la siguiente diapositiva.

Diapositiva 2

Eso te muestra lo que estamos tratando de lograr, así que, si piensas en COVID, realmente hay tres áreas principales. Uno es el huésped, que puede ser humano o puede ser animal, y así sucesivamente. Luego, tenemos el patógeno, el virus, y luego todo lo demás, que llamamos el medio ambiente. Queremos permitir que los investigadores examinen básicamente esta interacción entre esas diferentes áreas. Por ejemplo, observando la interacción huésped-patógeno: ¿cómo afecta el ambiente a las infecciones y demás? Eso era básicamente el fondo, y empezamos. Muestra la siguiente diapositiva, por favor.

Diapositiva 3

Cuando solicitamos esta beca RAPID, contamos con colaboradores de la NSF en el 'Open Knowledge Network'. Teníamos colaboradores de la Universidad de California en Santa Bárbara. Está en verde. UCSF está en naranja, y nosotros en azul. Lo dividimos en áreas particulares. Nos enfocamos en las áreas azules como características de la población, datos de salud, información de patógenos e información ambiental, mientras que nuestros colaboradores de UC Santa Barbara (esto es otro RAPID) se enfocan en el transporte y la cadena de suministro. UCSF se centra en la biomedicina. Próxima diapositiva por favor.

Diapositiva 4

Este es nuestro prototipo de gráfico de conocimiento. En un gráfico de conocimiento no tienes silos de información, pero vinculas todos los datos. Esto es sólo un prototipo. No tenemos mucha información aquí todavía. En el lado derecho, se ve información sobre el patógeno, el huésped. Tenemos información sobre la epidemiología, sobre este brote aquí. Luego nos centramos mucho en el área biomédica. Tenemos más de 30.000 cepas diferentes en este gráfico de conocimiento, pero todo está unido. Para cada cepa, conocemos todas sus variantes de mutaciones, el efecto sobre los genes y las proteínas. Conocemos las interacciones proteína-proteína. Vinculamos esto con las publicaciones. Luego, lo más importante, también lo vinculamos a la geolocalización. Mapeamos toda la jerarquía geográfica del mundo, para poder mapear cepas o casos a cualquier lugar del mundo - hasta el nivel del tracto censal. Bien, siguiente diapositiva por favor.

Diapositiva 5

Cuando comenzamos este proyecto, queríamos que fuera un proyecto automatizado que pudiera ser expandido por otros, por lo que tenemos un flujo de trabajo muy transparente y reproducible. En primer lugar, empezamos con datos de acceso abierto. Queremos ser capaces de redistribuir la información, por lo que empezamos con repositorios de datos públicos de confianza, y luego creamos un proceso que extrae automáticamente la información e integra esa información. Esa es la clave: la integración de toda esta información. Gastamos mucho [...] aquí es donde va la mayor parte del trabajo. Con COVID, las cosas cambian a diario, así que tenemos un proceso de actualización diario. Tenemos software de código abierto que está en, y sobre una base diaria, actualizamos la información, integramos la información, y luego la cargamos en un gráfico de conocimiento que luego se puede consultar. Tratamos de seguir los principios justos, y todo es abierto y de fácil acceso. Puedes acceder a todo nuestro software. Es reutilizable y así sucesivamente, así que, con eso, tal vez ir a nuestra próxima diapositiva.

Diapositiva 6

Usted sabe que una vez que creamos este gráfico de conocimiento, hay algunas cosas que podemos (o el usuario final) puede hacer. En primer lugar, puede consultar y navegar por el conocimiento donde encuentro información. Lo que se muestra aquí en la parte superior izquierda. Esto está explorando la interacción proteína-proteína entre las proteínas del virus y las proteínas humanas, por ejemplo. Entonces, ya que está en forma de gráfico, puedes interactuar si realmente exploramos esto. En verde, hay varias cepas del virus. En la geolocalización particular, puedes ver mutaciones específicas y ver cómo se comparten entre diferentes cepas,

por ejemplo. Este es un análisis más interactivo que un análisis exploratorio, pero si quieres hacer un análisis más cuantitativo en profundidad, también puedes acceder a todos esos datos fácilmente en cuadernos computacionales, como nuestro estudio o cuadernos Jupyter para un tipo de análisis más reproducible. Como he mencionado, el mapeo en la geolocalización es obviamente muy importante para las corbetas, por lo que también accedemos a esta información a través de paneles de control. Arriba a la derecha, mostramos, por ejemplo, casos actuales del Condado de San Diego o conteos de casos predichos. En el centro aquí, nos enfocamos en ciudades específicas y vemos, por ejemplo, varias condiciones preexistentes en esas áreas como: ahora, ¿cuál es la prevalencia de cáncer, diabetes, enfermedades cardíacas, etc., y cómo afecta a la población en riesgo? Podemos profundizar más usando esta herramienta de oleaje. Podemos profundizar, por ejemplo, todo el camino hasta el nivel del censo y luego analizar más a fondo sus poblaciones de salud en riesgo, la estructura de edad, etc. Queremos darle las gracias. Próxima diapositiva por favor.

Diapositiva 7

Queremos agradecer a la NSF por la financiación, y también al programa de 'Open Knowledge Network' del que formamos parte, y a los colaboradores. Obviamente, también estamos buscando colaboradores. Estamos buscando un número de colaboradores, así que si tiene conjuntos de datos abiertos que desea compartir con nosotros, nos gustaría discutirlos. Nos gusta integrarlos. Si usted tiene código que extrae datos de varias fuentes de datos que pueden ser de interés para nosotros, y obviamente si desea utilizar cualquiera de nuestros datos (y creo que allí ya hablamos con un número de personas en este programa), usted sabe que queremos saber de usted también. Creo que eso es todo, así que en la ventana de chat voy a pegar un enlace a nuestro gráfico COVID para que todo esté disponible en línea, para que pueda explorar eso usted mismo. Gracias.