

## Announcement

[Nightingale Open Science](#), a computing platform that houses massive new medical imaging datasets for the public good, launches today, December 14, at the prominent NeurIPS 2021 machine learning conference. The [full-day launch workshop](#) includes a product demo and panel discussions with top minds in computer science, technology, and medicine: Eric Schmidt, Regina Barzilay, Eric Topol, Jennifer Chayes, Aneesh Chopra, and others.

Nightingale was co-founded by Ziad Obermeyer, a physician–scientist at UC Berkeley, and Sendhil Mullainathan, a computational and behavioral scientist at the University of Chicago, with generous support from Schmidt Futures, The Gordon and Betty Moore Foundation, and Ken Griffin, founder and CEO of Citadel. At the launch event, the Nightingale team will announce a multi-funder collaborative bringing together anchor and new funders, in a large-scale effort to add multiple new imaging datasets across a number of geographies and disease areas.

## Background

Today, health data are mostly locked up in small sandboxes, controlled by a handful of private companies or well-resourced researchers. Nightingale Open Science aims to unlock those data, securely and ethically, and make them available for the public good. Just as ImageNet jump-started the field of machine vision, Nightingale seeks to build a community of researchers working in the new scientific field of ‘computational medicine.’

We believe this effort will be transformative for medicine. Over the past 100 years, we have made enormous strides in understanding how the body works, and how it fails. But deep and unsolved mysteries remain. For example, sudden cardiac death kills 300,000 Americans every year, but even in the rear-view mirror, doctors can find no identifiable cause for the majority. The key to solving these problems will lie in the massive volumes of complex, high-dimensional data health systems produce every day: electrocardiograms, x-rays and CT scans, digital pathology images, and more. In our current health system, these data are interpreted by humans. But machines have new ways of ‘seeing’ signals and patterns in the data that humans cannot.

Nightingale is a non-profit, open platform housing these de-identified medical datasets, and making them available to a diverse, global community of researchers. Nightingale funds and collaboratively builds these datasets with health systems around the world, then makes the de-identified datasets available on a secure cloud platform. By focusing on data that link medical images with real patient outcomes – rather than doctors’ opinions – Nightingale enables the creation of algorithms that learn from nature -- not from humans — and fosters groundbreaking research bridging computer science and clinical medicine.

## Quotes

[Eric Schmidt](#) – Former CEO of Google, technologist, entrepreneur, and philanthropist  
*“Nightingale is incredibly important. It is the first large database of images that is being organized around healthcare. We saw how well this worked with ImageNet in 2011. I believe, with the Nightingale team, this repository of never-before-seen images, tied to outcomes with labeled data, will lead to revolutionary new approaches.”*

[Ken Griffin](#) – Founder and CEO, Citadel

*“Nightingale Open Science enables the world’s leading data scientists to apply powerful machine learning and predictive analytics to solve some of medicine’s most important and urgent challenges. This pioneering initiative advances the field of computational medicine and has the potential to unlock lifesaving breakthroughs for people around the world.”*

[Eric Topol](#) – Founder and Director, Scripps Research Translational Institute

*“We welcome Nightingale Open Science, a new non-profit resource that will make large medical image datasets available to the research community. Nightingale OS will unquestionably help advance and accelerate AI in medicine, just as ImageNet did for deep learning several years ago. I am thrilled to serve as an advisor to the team.”*

[Aneesh Chopra](#) – Former U.S. CTO, Co-Founder & President, CareJourney

*“We need an ‘all hands on deck’ approach to transform our care delivery system towards higher value and the work starts with open data platforms for learning. I’m grateful for the launch of Nightingale Open Science and am eager to celebrate the community of entrepreneurs, innovators, policymakers and front-line care teams putting those insights into practice for the betterment of the public.”*

[The Gordon and Betty Moore Foundation](#)

*“The majority of academic papers describing clinical AI algorithms used training datasets [from just three states](#). This reflects a high barrier of entry for health systems to contribute their data to public repositories, resulting in datasets that may not be reflective of the general population. The Nightingale Open Science platform and this collaborative funding call will support health systems that serve under-represented populations to ensure that public datasets and clinical AI tools are more inclusive of all populations.”*

### **More about the datasets**

The platform will house 5 new datasets, totaling over 40TB of images and waveforms, at the time of launch, with many more to come over the coming months. Each dataset is curated around an unsolved medical mystery that AI can help solve:

1. **Diagnosing ‘silent’ heart attack:** 49,000 ECG waveforms linked to the results of cardiac ultrasounds, to visualize scars in the wall of the heart formed by prior heart attack. This could help identify patients who urgently need drug regimens that prevent their next heart attack—or cardiac arrest.
2. **Identifying high-risk breast cancer:** 175,000 digital pathology images from 11,000 patients, linked to patient outcomes (stage, metastasis, mortality). This will allow algorithms to identify patients at high risk of poor outcomes.
3. **Subtyping cardiac arrest:** 24,000 ECG waveforms, from emergency patients who suffered cardiac arrest, as well as matched controls, linked to mortality, neurological function, and data on the cause of the arrest. This could help the emergency team figure out why the heart stopped, what they can do about it in real time, and whether a patient will survive without profound physical or neurological impairments.

4. **Predicting fracture risk:** 64,000 chest x-rays linked to data on past and future fractures all over the body, as well as data on diagnoses of osteopenia and osteoporosis. This could help target preventive care and medications to those at highest risk.
5. **Emergency triage of Covid-19 patients:** 7,000 chest x-rays from Covid-19 patients, linked to data on pulmonary deterioration (need for a ventilator) and mortality. These data could help doctors make critical triage decisions for Covid-19 patients—whether they are safe to go home, or whether they need to be monitored in the hospital.

**Anecdotes and related links:**

- [Science Friday](#) episode featuring Ziad Obermeyer discussing a study in Nature Medicine, leveraging a dataset that inspired many on the Nightingale platform. Thanks to this NIH-funded open dataset, researchers led by Emma Pierson (Cornell Tech) showed that AI was able to detect causes of knee pain in Black patients that physicians miss. See also:
  - The original study in [Nature](#)
  - A write-up in [Wired](#) and an interview in FT's [Tech Tonic](#)