Bringing scientists to data

(to accelerate discoveries and improve human health)

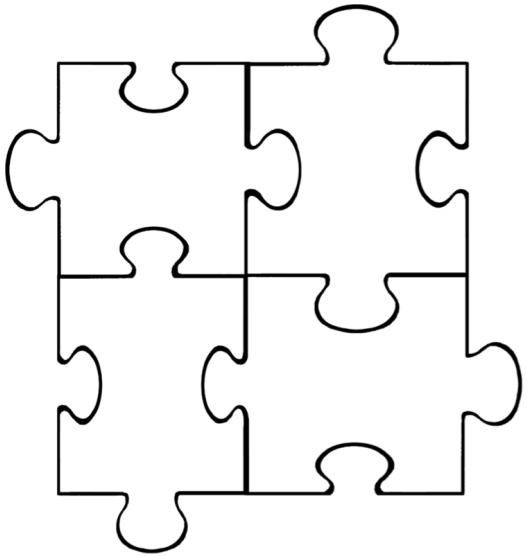
Somalee Datta, PhD
Director Research IT,
SoM Information Resources & Technology

Can we enable exploration on bigger data, faster analysis and cheaper hypothesis testing?

System engineering approach to “bigger/faster/cheaper”

Findable, Accessible, Interoperable, and Reusable data

Privacy protecting data sharing across silos



Research ethics, and compliance

Secure workspaces to interact with data

Ethics and Compliance

&

Privacy protecting data sharing

Ethics & Compliance

In Sep 1999, teenager Jesse Gelsinger died in a gene therapy trial at UPenn.

“But even one lapse is one too many.” - Donna Shalala, Secretary of the U.S. Department of Health and Human Services, NEJM, 2000

ars TECHNICA BIZ & IT TECH SCIENCE POLICY CARS GAMING & CULTURE FORUMS

SCIENCE —

Bucking FDA, Peter Thiel funds “patently unethical” herpes vaccine trial

ars TECHNICA BIZ & IT TECH SCIENCE

WHO NEEDS HUMAN SAFETY, ANYWAY? —

University could lose mil... “unethical” research backed by Peter Thiel

With nudge from federal regulators, an internal investigation found big problems.

BETH MOLE - 11/14/2017, 4:10 PM

Why is privacy a hard problem?

Probability of re-ID w/ gender known:

	5 digit zip code	County
YYYY	0.2 %	0.0%
MM / YYYY	4.2 %	0.2%
MM / DD / YYYY	63.3 %	14.8%

P. GOLLE: Proceedings of the 5th ACM Workshop on Privacy in the Electronic Society. **2006**: 77-80.

iDASH Security & Privacy Workshop 2017

Track 1: Develop **privacy preserving patient linkage** technique

Track 2: Develop scalable solutions using **SGX secure hardware** to enable genome variants search

Track 3: Develop **Homomorphic Encryption based ML techniques** over encrypted data

More @ http://med.stanford.edu/gapp.html

Biomedical data analysis brings novel benefits to human health

Predictive Analytics with 10-20 years of EHR data

MIT Tech Review
@techreview
Following

Researchers from Stanford have designed a deep neural network that can look at a patient's health records and estimate the chance of mortality in the next three to 12 months.

Improving Palliative Care with Deep Learning

Anand Avati*, Kenneth Jung†, Stephanie Harman‡, Lance Downing†, Andrew Ng* and Nigam H. Shah†
*Dept of Computer Science, Stanford University
Email: {avati,ang}@cs.stanford.edu
†Center for Biomedical Informatics Research, Stanford University
Email: {kjung,ldowning,nigam}@stanford.edu
‡Dept of Medicine, Stanford University School of Medicine
Email: {smharman}@stanford.edu

More than 60% of deaths in the US happen in an acute care hospital. This predictive model helps the Palliative Care team to be engaged early enough to ensure meaningful services.

Genome based diagnosis coming to you

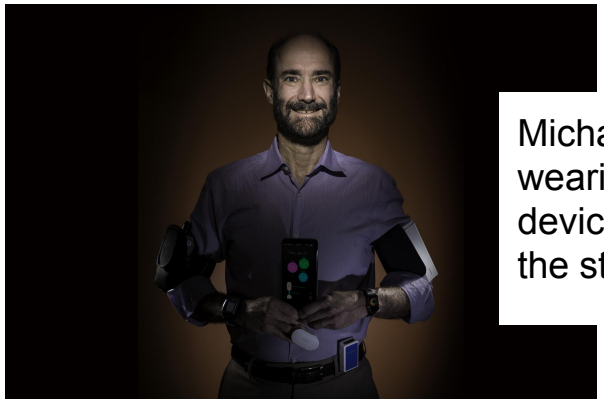
Built on a platform that allows rapid turnaround from experimentation to productization

- Combines research excellence with operational delivery
- Public-Private partnership
- Pilot launched in Jan 2014
- Production in Jan 2018
- Genome and exome sequencing
- Three major disease areas:
  - Pediatric and adult syndromes
  - Heritable cancer predisposition
  - Heritable cardiovascular disease

Future is now!

Built for the Future. Study Shows Wearable Devices Can Help Detect Illness Early

Posted on January 17, 2017 by Dr. Francis Collins



Michael Snyder wearing multiple devices used in the study

More @ http://stan.md/2AldJLE

Apple Watch will alert heart-study participants if they have an irregular beat

Edward C. Baig, USA TODAY Published 3:40 p.m. ET Nov. 30, 2017 | Updated 4:20 p.m. ET Nov. 30, 2017

AFib, the leading cause of stroke, is responsible for approximately 130,000 deaths and 750,000 hospitalizations in the US every year.

American Well

BioTelemetry inc

More @ http://stan.md/2iw97f0

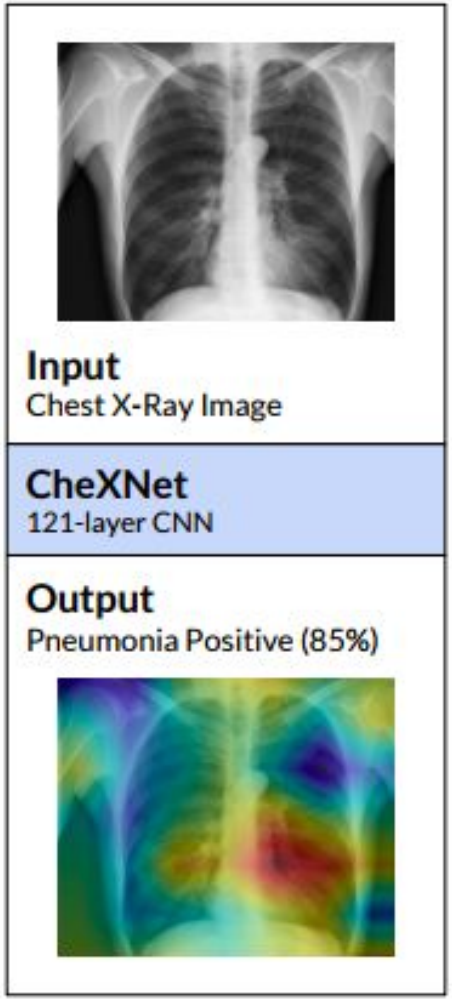
Imaging data - Radiology

CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning

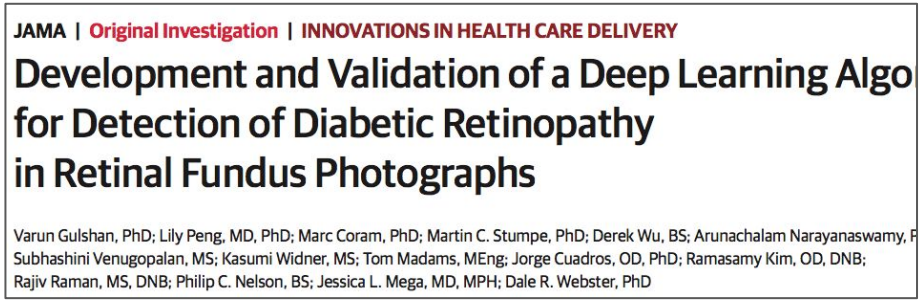
Pranav Rajpurkar [*1] Jeremy Irvin [*1] Kaylie Zhu [1] Brandon Yang [1] Hershel Mehta [1]
Tony Duan [1] Daisy Ding [1] Aarti Bagul [1] Curtis Langlotz [2] Katie Shpanskaya [2]
Matthew P. Lungren [2] Andrew Y. Ng [1]

(Algorithm + Radiologist) is better than (Radiologist)

Stanford researchers develop convolutional neural network algorithm that can diagnose up to 14 types of medical conditions based on chest x-ray image.



Potential to make planet scale impact



JAMA | Original Investigation | INNOVATIONS IN HEALTH CARE DELIVERY

Development and Validation of a Deep Learning Algo for Detection of Diabetic Retinopathy in Retinal Fundus Photographs

Varun Gulshan, PhD; Lily Peng, MD, PhD; Marc Coram, PhD; Martin C. Stumpe, PhD; Derek Wu, BS; Arunachalam Narayanaswamy, P Subhashini Venugopalan, MS; Kasumi Widner, MS; Tom Madams, MEng; Jorge Cuadros, OD, PhD; Ramasamy Kim, OD, DNB; Rajiv Raman, MS, DNB; Philip C. Nelson, BS; Jessica L. Mega, MD, MPH; Dale R. Webster, PhD

TOM SIMONITE BUSINESS 06.07.17 3:50 PM

GOOGLE'S AI EYE DOCTOR GETS READY TO GO TO WORK IN INDIA

WIRED

ARAVIND EYE CARE SYSTEM

400 million diabetic patients in world; 70 million in India; 30% estimated to develop DR

"Piloting this AI-based diagnosis is Aravind Eye Hospital, the largest eye care provider in India. Aravind Eye Hospital contributed to Google's research by providing images of DR patients." - Forbes, Sep 5, 2017

Standing upon shoulders of giants ...

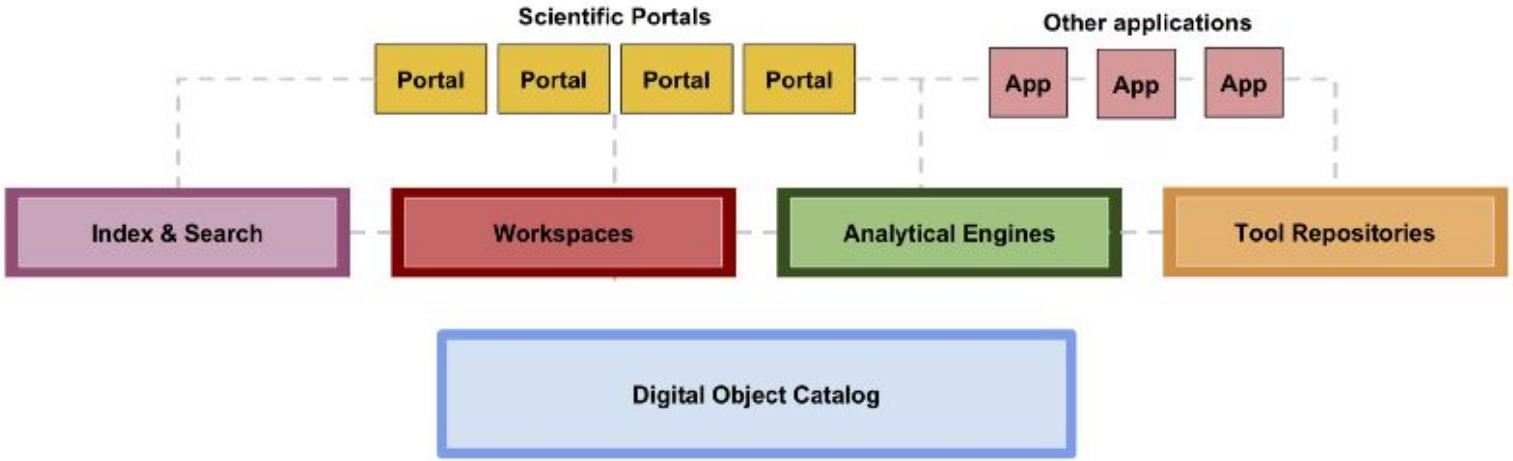
NIH BD2K “Data Commons”

Monday, November 6, 2017

NIH awards to test ways to store, access, share, and compute on biomedical data in the cloud

NIH Data Commons Pilot Phase to seek best practices for developing and managing a data commons.

- 12 awards form nucleus of an NIH Data Commons Pilot Phase Consortium
- $9 Million in 2017
- Using public clouds



Multi-Cloud Data Biosphere Vision @ http://bit.ly/2zNdBBL

Sage Bionetworks approach to Open Research

First, design for data sharing

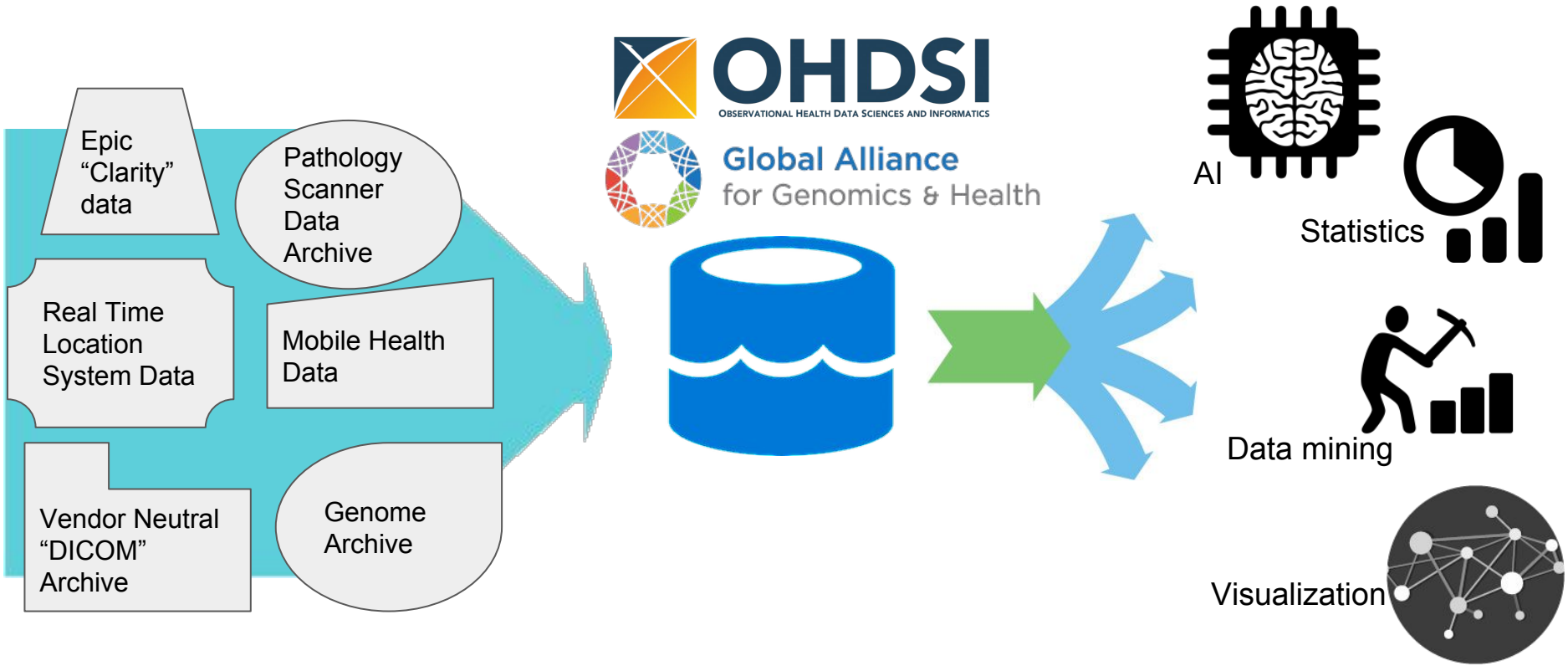
John Wilbanks & Stephen H Friend

To upend current barriers to sharing clinical data and insights, we need a framework that not only accounts for choices made by trial participants but also qualifies researchers wishing to access and analyze the data.

NATURE BIOTECHNOLOGY VOLUME 34 NUMBER 4 APRIL 2016

Stanford approach

Findable, Accessible, Interoperable, Reusable



Secure workspaces to interact with data

Google Cloud Platform

Stanford Data Center at SLAC

gatk

H2O.ai

jupyter

snorkel

APACHE Spark

XNAT

MATLAB

STATA

RStudio

TensorFlow

docker

Research Privacy, Ethics and Compliance

For a single dataset:

- Data Use Agreement
- Access “life-cycle” management
- Compliance requirements e.g. IRB approval, HIPAA training
- Track use

For a single user:

- Single pane of view to data access status
- Data flyby - metadata, counts, quality checks, versions
- Access to data experts
- Access to research community

REDIVIS
data. (re) imagined.

Stanford MEDICINE | Center for Population Health Sciences

More @ https://redivis.com/StanfordPHS

Let’s build together!