

# How Did Splunk Secured Patient Clinical And Genetic Data On AWS By Creating The First HIPAA Cloud Compliant Infrastructure At Harvard

Paul Avillach, MD, PhD

Asst Professor of Biomedical Informatics, Harvard Medical School

.conf2016

splunk >

# Disclaimer

During the course of this presentation, we may make forward looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC. The forward-looking statements made in the this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not, be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

# Diagnosis of Leukemia during 18<sup>th</sup> century ?

## Humorism

*phlegmatic*

*choleric*



**Treatment:  
Bloodletting**



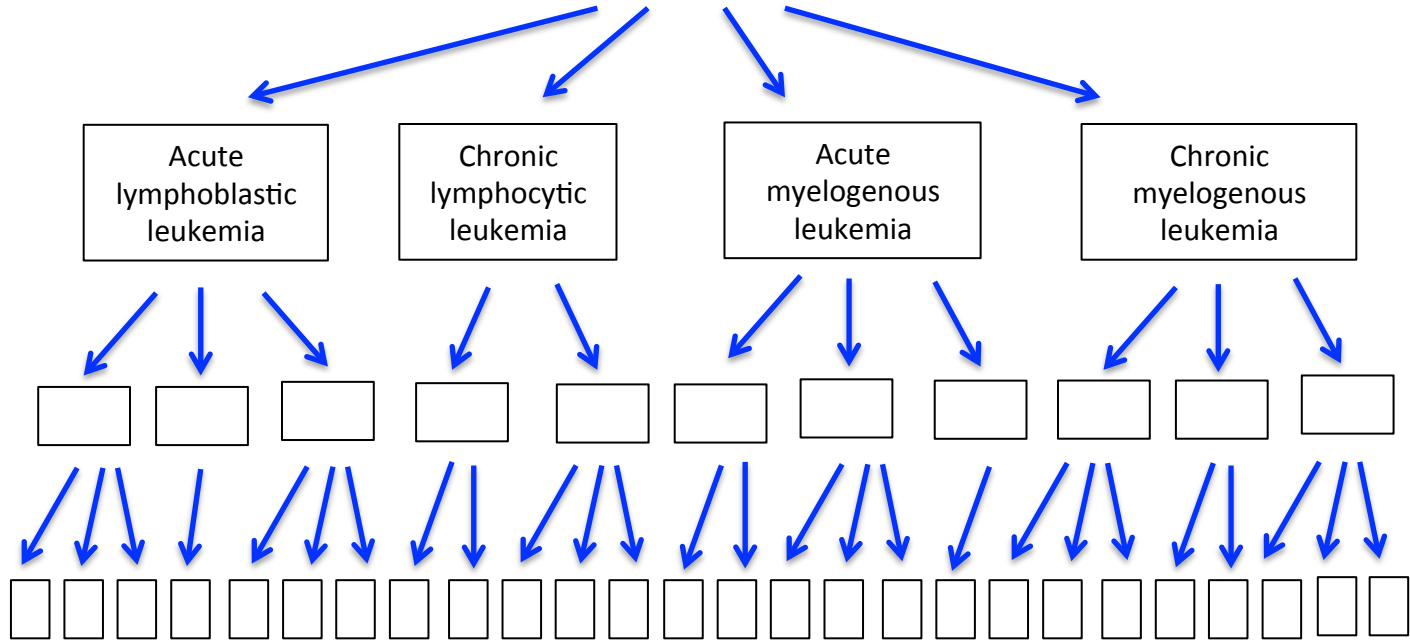
*sanguine*

*melancholic*

# Diagnosis of Leukemia:



*sanguine*



**For each subtype:**

- Diagnosis
- Prognosis
- Potential treatment

MAY 27, 2013

The IRS Mess / Syria's YouTube War / The End of Alimony

TIME

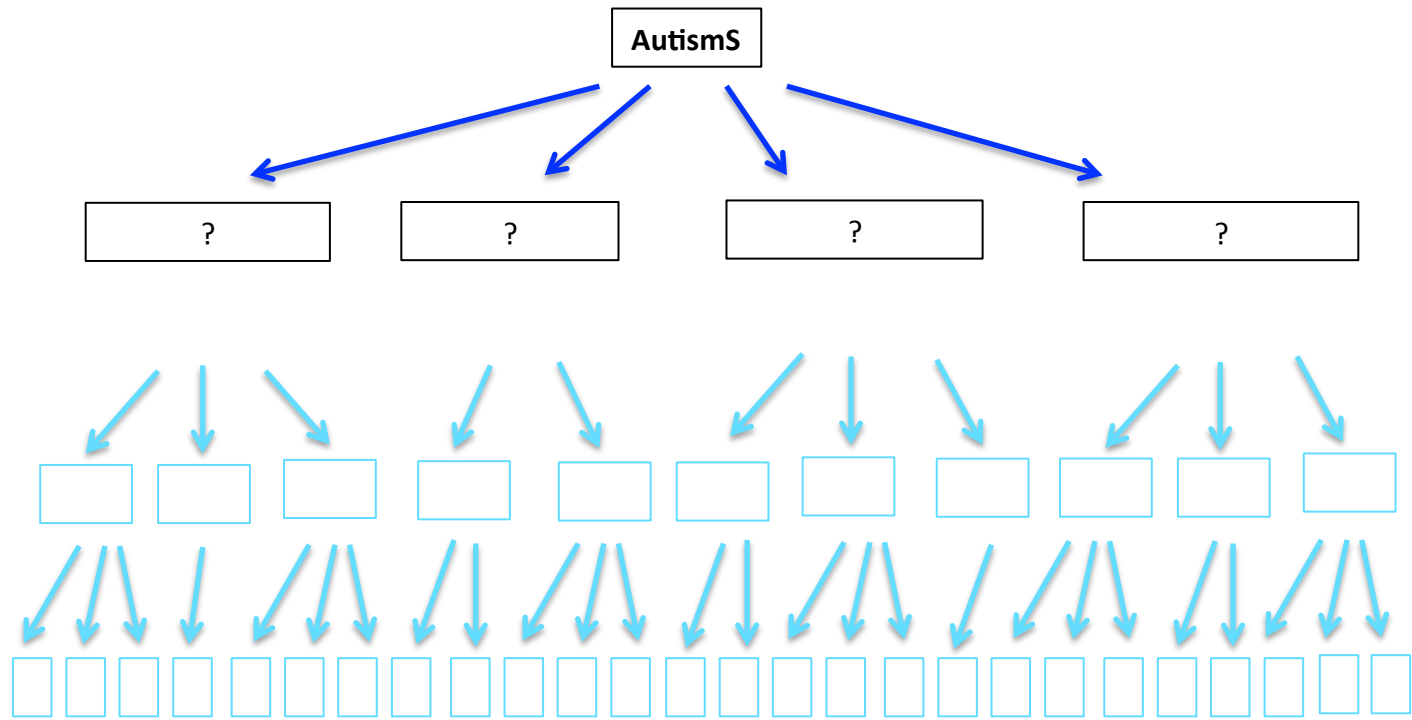


THE  
**ANGELINA**  
EFFECT

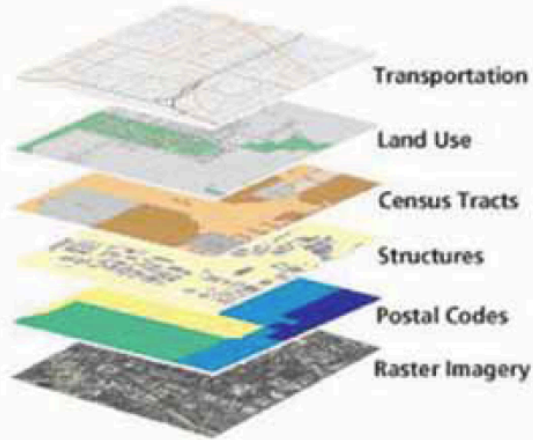
Angelina Jolie's double mastectomy puts genetic testing in the spotlight. What her choice reveals about calculating risk, cost and peace of mind

BY JEFFREY KLUGER & ALICE PARK

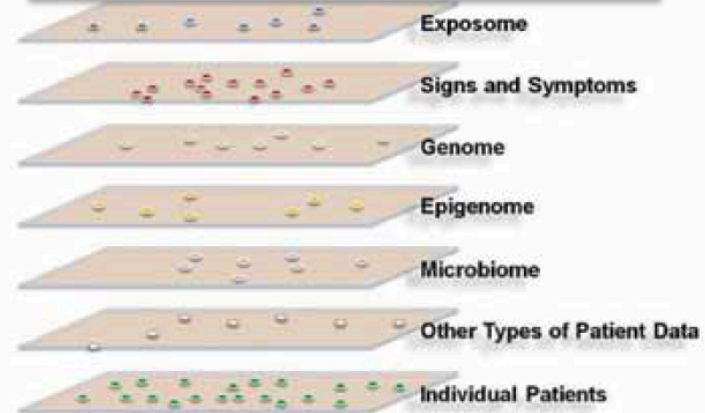
time.com



**Google Maps: GIS layers  
Organized by Geographical Positioning**

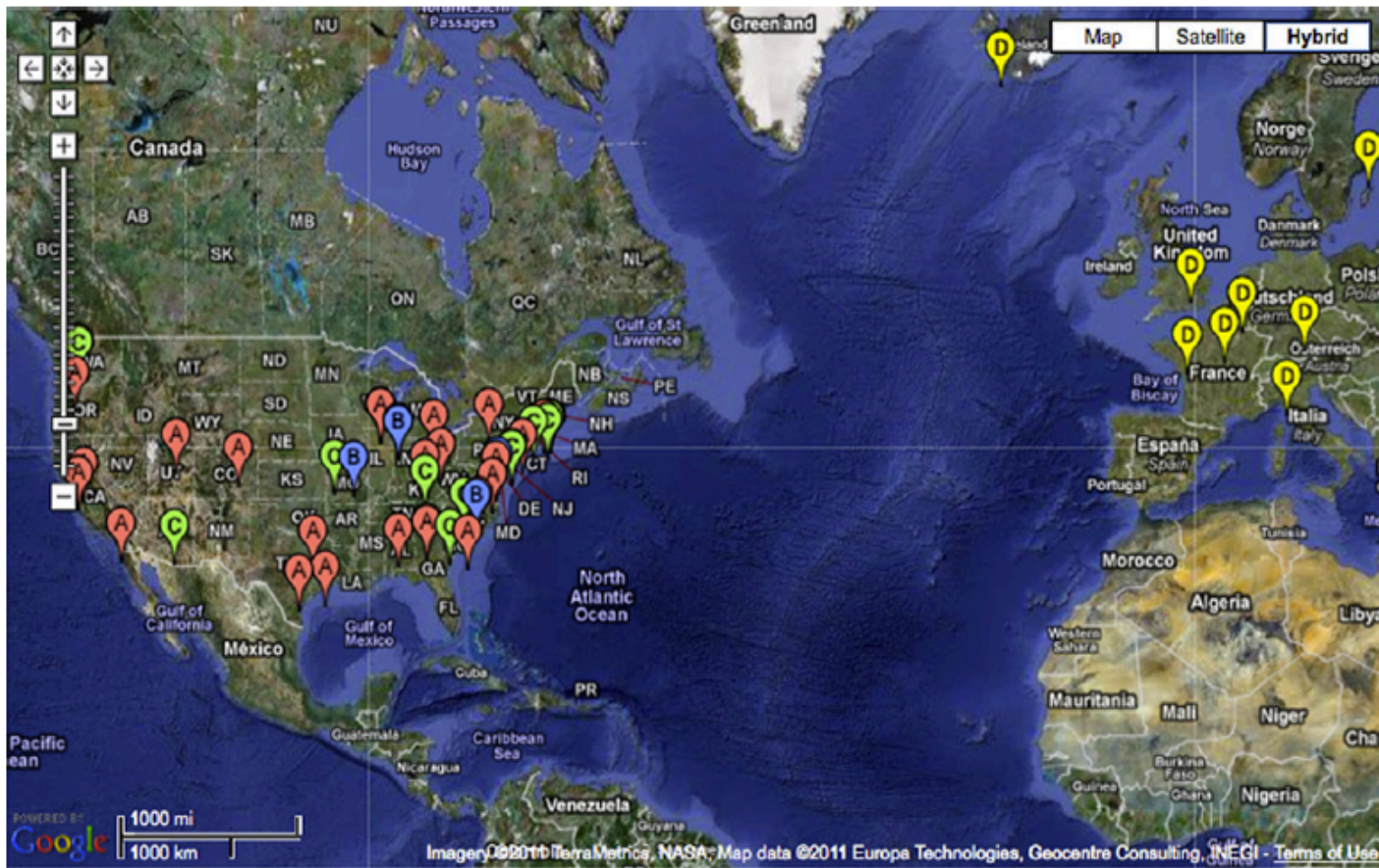


**Information Commons  
Organized Around Individual Patients**

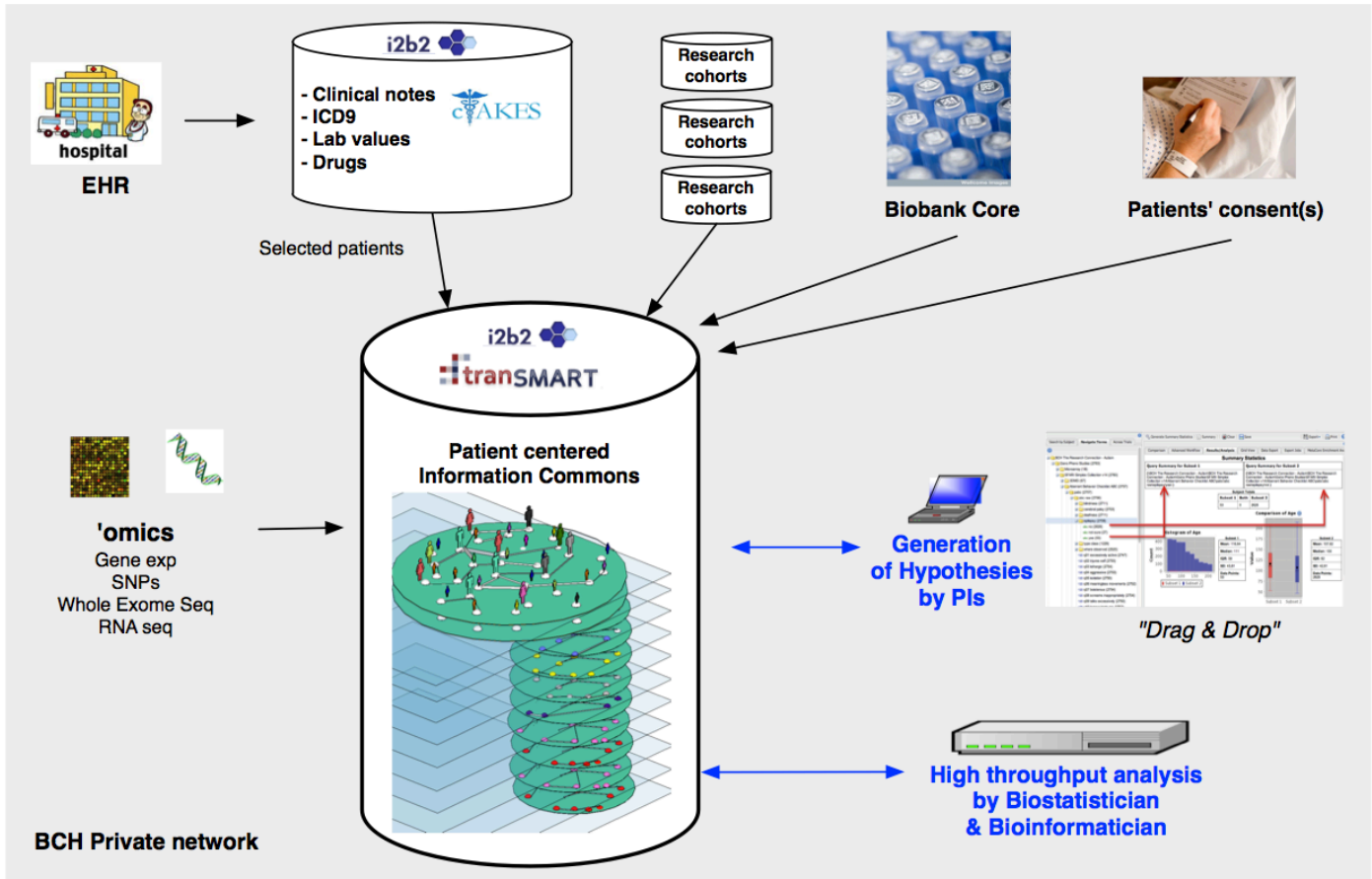


Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease

Report from National academy of science, USA, 2011







→ integration  
 ↔ analysis

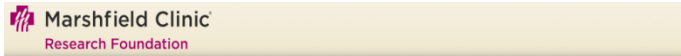


National Center  
for Advancing  
Translational Sciences

The NIH/NCATS GRDR® Program  
Global Rare Diseases Patient Registry  
Data Repository



Malignant Hyperthermia Association of the United States



Marshfield Clinic®  
Research Foundation

Clinical Registry Investigating Bardet-Biedl Syndrome (CRIBBS)



INTRACRANIAL HYPERTENSION  
RESEARCH FOUNDATION

CONTACT



CoRDS Registry  
Coordination of Rare Diseases  
at Sanford

SANFORD  
RESEARCH



National Ataxia  
Foundation



INTERNATIONAL WAGR  
SYNDROME ASSOCIATION



Pachyonychia Congenita Project

*Fighting for a cure. Connecting & helping patients. Empowering research.*



The Plastic Bronchitis Foundation

*Looking for a Cause, Working on a cure, Education and assisting*



Wolfram Syndrome International Registry and Clinical Study



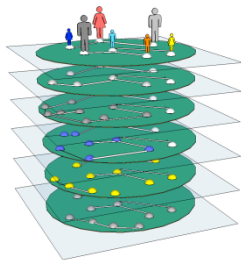
CdLS Foundation  
Cornelia de Lange Syndrome Foundation, Inc.

<https://grdr.hms.harvard.edu>

splunk > .conf2016

Rare disease registry	Patients	Variables
Clinical Registry Investigating Bardet-Biedl Syndrome	180	708
International Pachyonychia Congenita Research Registry	569	496
International Plastic Bronchitis registry	66	63
Intracranial Hypertension Registry	1,349	91
North American Malignant Hyperthermia Registry	2,122	154
Wolfram Syndrome International Registry	124	580
Coordination of Rare Diseases at Sanford Registry	2,091	40
<i>including:</i>	<i>including:</i>	<i>Additional variables:</i>
<i>National Ataxia Foundation</i>	<i>869</i>	<i>15</i>
<i>International WAGR syndrome association</i>	<i>52</i>	<i>379</i>
<i>Cornelia De Lange Syndrome Registry</i>	<i>67</i>	<i>485</i>

<b>Total</b>	<b>6,501</b>	<b>2,132</b>
--------------	--------------	--------------



<https://grdr.hms.harvard.edu>



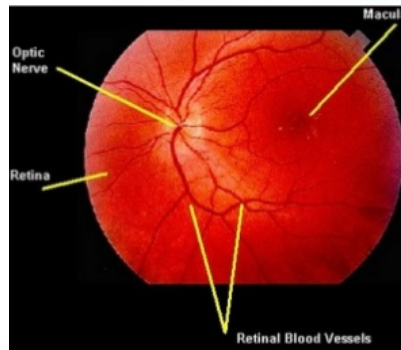
## Use case 1:

Is there an association between  
“papilledema”

*and*

“high level of intracranial pressure”?

Normal Fundus - Papilledema



<https://grdr.hms.harvard.edu>

# 1) Selection of the Subsets

**SUBSET 1**  
Exclude Enable Variant Panel X  
.../Yes\

**SUBSET 2**  
Exclude Enab  
.../No\

**SUBSET**  
\\Intracranial Hypertension Registry\2 Clinical data\5 Physical findings\Symptoms\Ophthalmological symptoms\Papilledema\Having Papilledema\Yes\

**SUBSET 2**  
\\Intracranial Hypertension Registry\2 Clinical data\5 Physical findings\Symptoms\Ophthalmological symptoms\Papilledema\Having Papilledema\No\

**Relation between Subset 1 and Subset 2**  
X

# 2) Drag & Drop the Variable “TAP1 pressure”

**Analysis of ...Lumbar tap\Tap1\TAP1 pressure for subsets:**  
Comparison of ...Lumbar tap\Tap1\TAP1 pressure

**Histogram of ...Lumbar tap\Tap1\TAP1 pressure**

**Box Plot**

Subset 1	Subset 2
IHR4	IHR4
Mean: 351.94	Mean: 312.58
Median: 335	Median: 300
IQR: 120	IQR: 100
SD: 100.55	SD: 85.86
Data Points: 1065	Data Points: 200

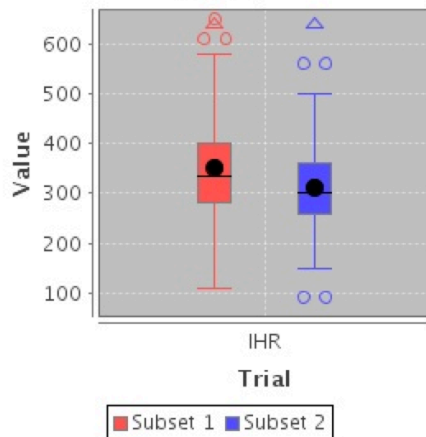
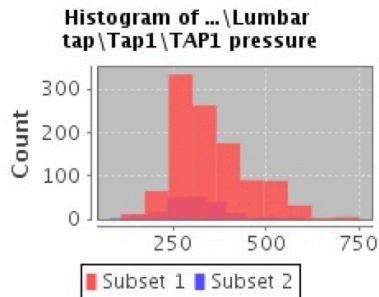
t statistic: 5.7809  
p-value: 1.8077e-08

The results are significant at a 95% confidence level.

**Variable « TAP1 pressure »**  
\\Intracranial Hypertension Registry\2 Clinical data\6 Diagnosis\Lumbar tap\Tap1\TAP1 pressure\

## Analysis of ...\Lumbar tap\Tap1\TAP1 pressure for subsets:

Comparison of ...\Lumbar tap\Tap1\TAP1 pressure



Subset 1	Subset 2
IHR	IHR
Mean: 351.94	Mean: 312.58
Median: 335	Median: 300
IQR: 120	IQR: 100
SD: 100.55	SD: 85.86
Data Points: 1065	Data Points: 200

t statistic: 5.7809  
p-value: 1.8077e-08

The results are significant at a 95% confidence level.

### ***P value < 0.0001 (T test)***

*The mean "351.94" of "Papilledema = Yes" group (subset1) > The mean "312.58" of "Papilledema= No" group (subset2)*

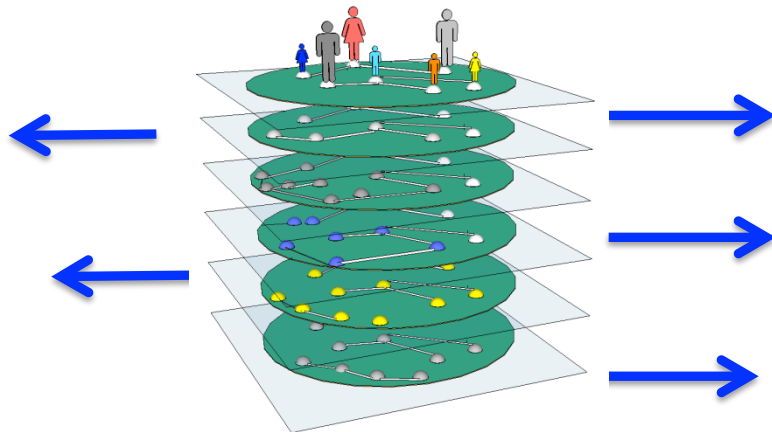
*The high level of the intracranial pressure is associated with the "Papilledema" symptom*

Morgan WH, Balaratnasingam C, Lind CR, Colley S, Kang MH, House PH, Yu DY.  
**Cerebrospinal fluid pressure and the eye.** Br J Ophthalmol. 2015 Apr 15. pii:  
bjophthalmol-2015-306705



NIH Big Data to Knowledge (BD2K)

## Patient Centric Information Commons (PIC)



**RESTful API**

Play with API: <https://bd2k-picsure.hms.harvard.edu>

Source code: <https://github.com/hms-dbmi>

# Protected Health Information (PHI)

## Identifiers

The **18 Identifiers** defined by HIPAA are:

- Name
- Postal address
- All elements of dates except year
- Telephone number
- Fax number
- Email address
- URL address
- IP address
- Social security number
- Account numbers
- License numbers
- Medical record number
- Health plan beneficiary #
- Device identifiers and their serial numbers
- Vehicle identifiers and serial number
- Biometric identifiers (finger and voice prints)
- Full face photos and other comparable images
- Any other unique identifying number, code, or characteristic

Copyright 2011 The Regents of University of California

All Rights Reserved

The Regents of the University of California accepts no liability for any use of this presentation or reliance placed on it, as it is making no representation or warranty, express or implied, as to the accuracy, reliability, or completeness of the presentation.







### Harvard IRB security levels:

Level 5 - Extremely sensitive information

**Level 4 – Very sensitive information**

Level 3 – Sensitive, or Confidential information

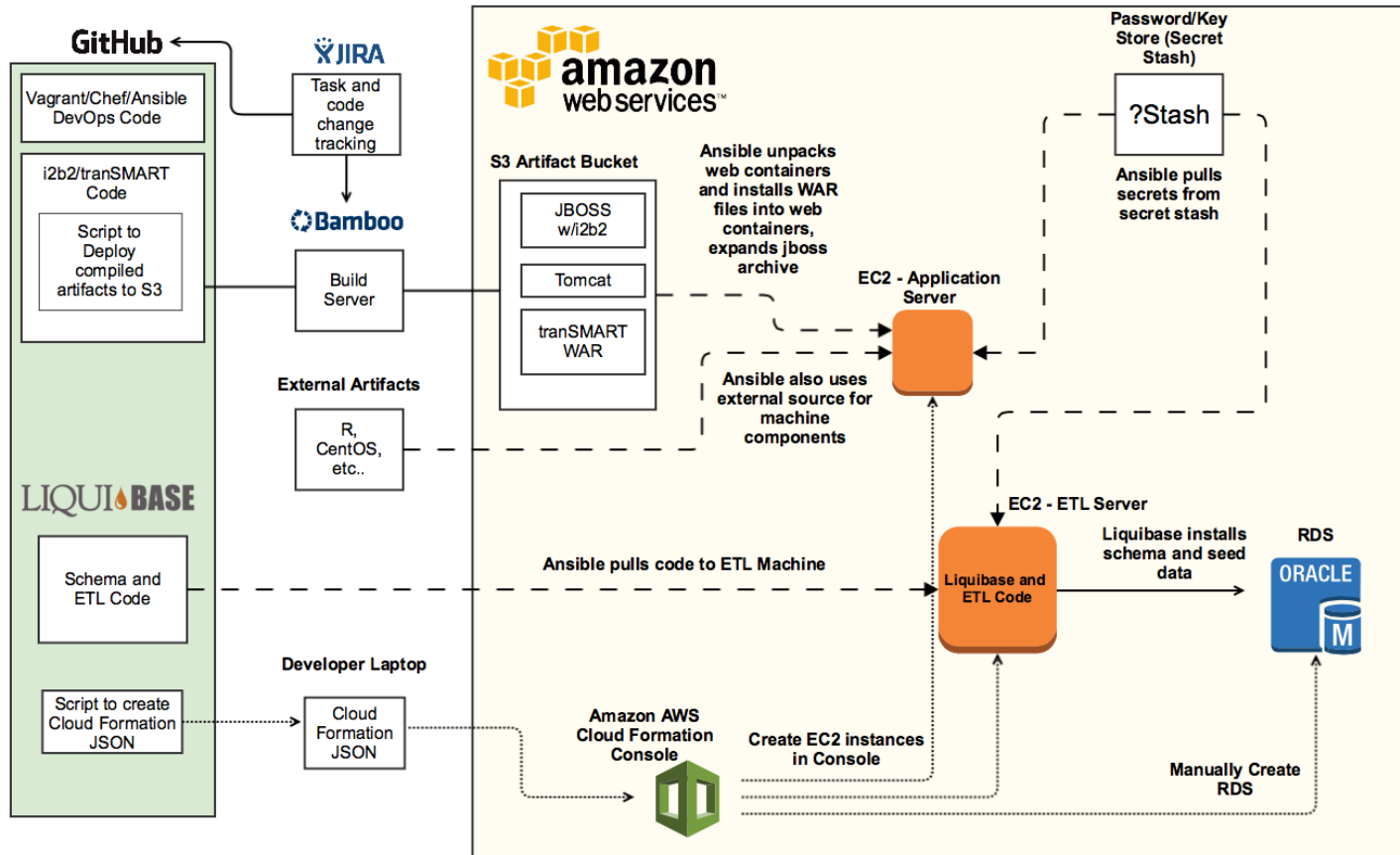
Level 2 - Benign information to be held confidentially

Level 1 - Non-confidential research information



HARVARD  
UNIVERSITY





# HIPPA Compliance On AWS



# Secured Access Control

## 1. Authentication

## 2. Authorization

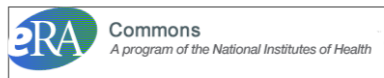
*you allowed to do?*

**Level 0** : *Authenticated BUT no access to data*

**Level 1** : *Aggregated data*

**Level 2** : *patient level data*

## Enterprise Identity Providers



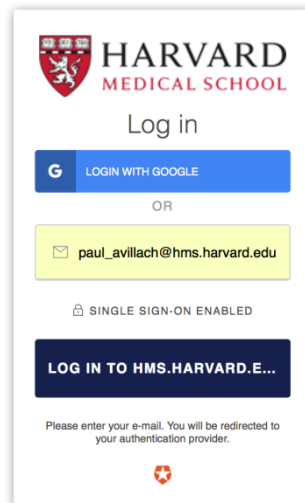
[...]

## Public Identity Providers

Sign with Google

# 1. Authentication

## Service Providers



### Application User Interface



### Programmatic Interface

RESTful API



AWS

VPC

**PIC-SURE**



NIH Big Data to Knowledge (BD2K)

VPC

**CCD - PIC-SURE**



NIH Big Data to Knowledge (BD2K)

VPC

**CountEverything**



NIH Big Data to Knowledge (BD2K)

VPC

**GRDR**



National Center for Advancing Translational Sciences

VPC

**pcori** PPRN  
PMS\_DN

.....

Until every child is well™

AWS

VPC

**Precision Link  
Biobank portal**

.....

**NGRID**

NIH National Institute of Mental Health

VPC

**splunk > enterprise**



- Deployment of a new server
- System activity on ec2 and RDS
- Application activity on ec2
  - tomcat
  - Jboss / Wildfly
  - Httpd
  - solr

- Authentication via Auth0 API
- User activity
  - What users do on the platform?
- API activity
  - What is being accessed and by whom?
- AWS console and API activity
  - Who is accessing AWS console or API and doing what?



# Monitoring EVERYTHING

- Dashboards sent by email every morning
- Alerts trigger events in Slack

<http://avillach-lab.hms.harvard.edu>

## Manager of Data Infrastructure

- Michael McDuffie, MSc

## Developers

- Jeremy Easton-Marks
- Gabor Korodi
- Thomas DeSain
- Sean Finan
- Ranjay Kumar
- Alexander Nikitin
- Ken Hoflen

## Project Manager

Cassandra Perry, MS, CGC

## Research Associates - Postdocs

- Antoine Tran, MD, MSc
- Laurie Tran, MD, MSc
- Cartik Saravana, PhD
- Li Ly, PhD
- Joany Zachariasse, MD, MSc
- Antoine Neuraz, MD, MSc

## Graduate students

- Ombeline Dorval, MD
- Maxime Wack, MD
- Claire Hassen-Kodja, MD, MSc
- Emmanuelle Sylvestre, MD, MSc
- Yuri Ahuja, MD, PhD HMS candidate

## Previous members

- Samuel Finlayson, MD, PhD HMS candidate
- Ephi Sachs, MD
- Pei Chen
- Sushma Hanawal

## Funding:



THE NANCY LURIE MARKS FAMILY FOUNDATION



## We are hiring now:

- Senior Software Developer \*2
- Postdocs \*3

# THANK YOU

.conf2016

