

LFS 302

# AWS re:INVENT

Real-World Evidence Platform to  
Enable Therapeutic Innovation

November 27, 2017

AWS  
re:Invent



Historically, there has been an information asymmetry in pharmaceutical R&D where the biopharmaceutical companies had the deepest understanding and knowledge about their products and how they helped and interacted with patients. Now, there's new, real-world data that exists from regulators, health plans, government authorities, and patients, which is helping pharma companies to understand how their therapies and their innovations drive value and impact in patient populations. There are imperatives to leverage that data, create new partnerships in their ecosystem, and get access to that data in an ethical way to derive insights to both fuel innovation and drive discovery. In this session, you learn best practices from Deloitte and Celgene about strategy, operating models, and execution frameworks when implementing a real-world, evidence data platform.

Let's talk about data

We need to continuously use data to drive outcomes

# The Data-Value Chain



AWS  
re:Invent



## Data Lake as a Source of Truth

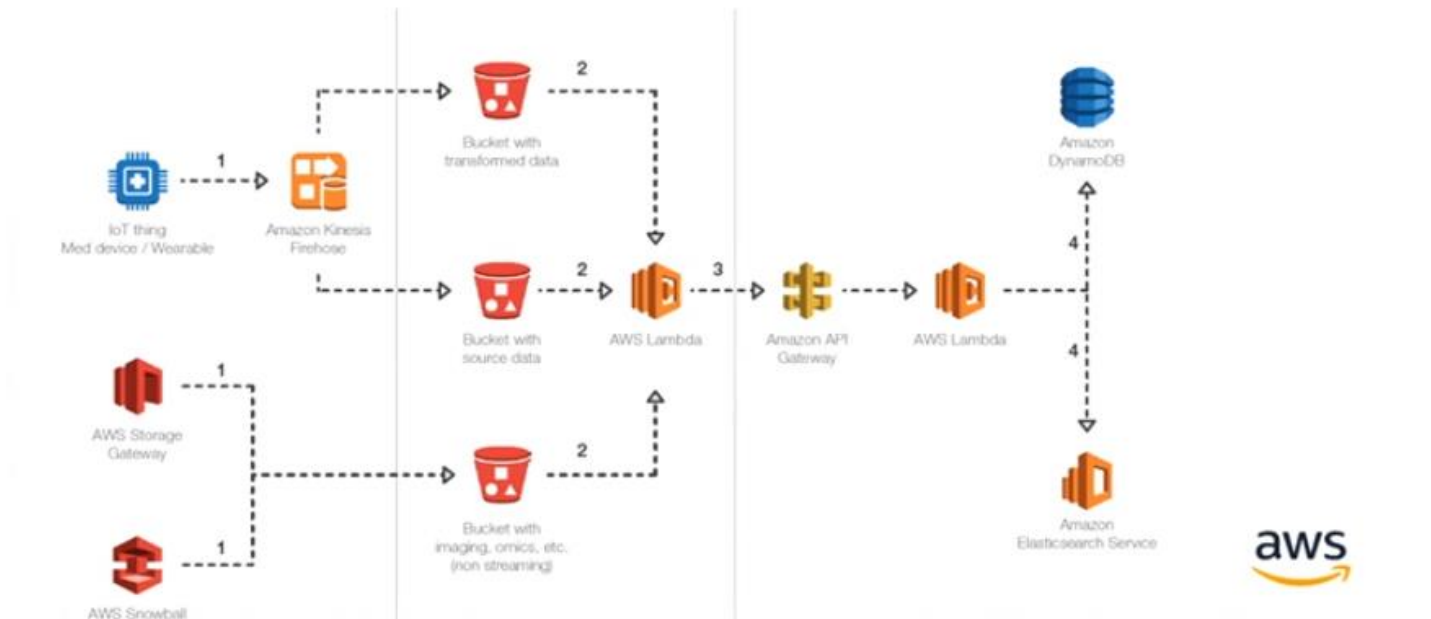


# Data Acquisition

Acquire from Wearables,  
EHRs, Genomics, etc.

Store Streaming, Structured  
and Unstructured Health Data

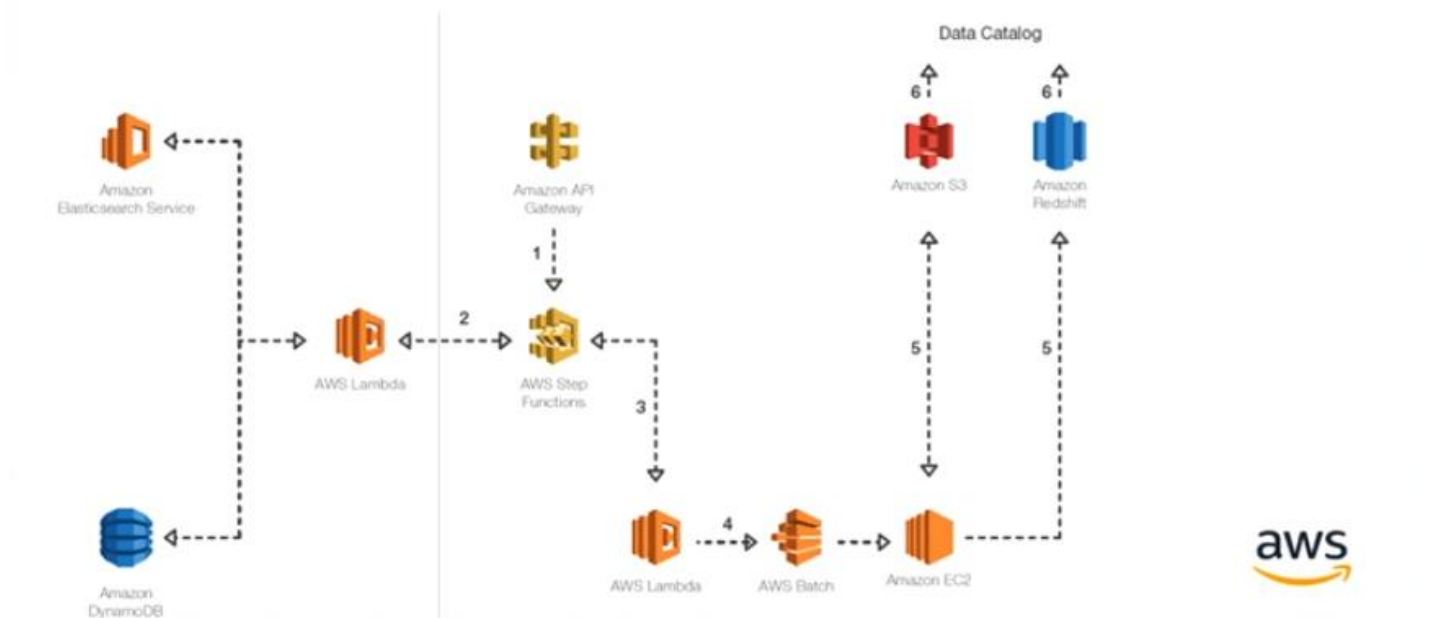
Record Metadata  
in Data catalog



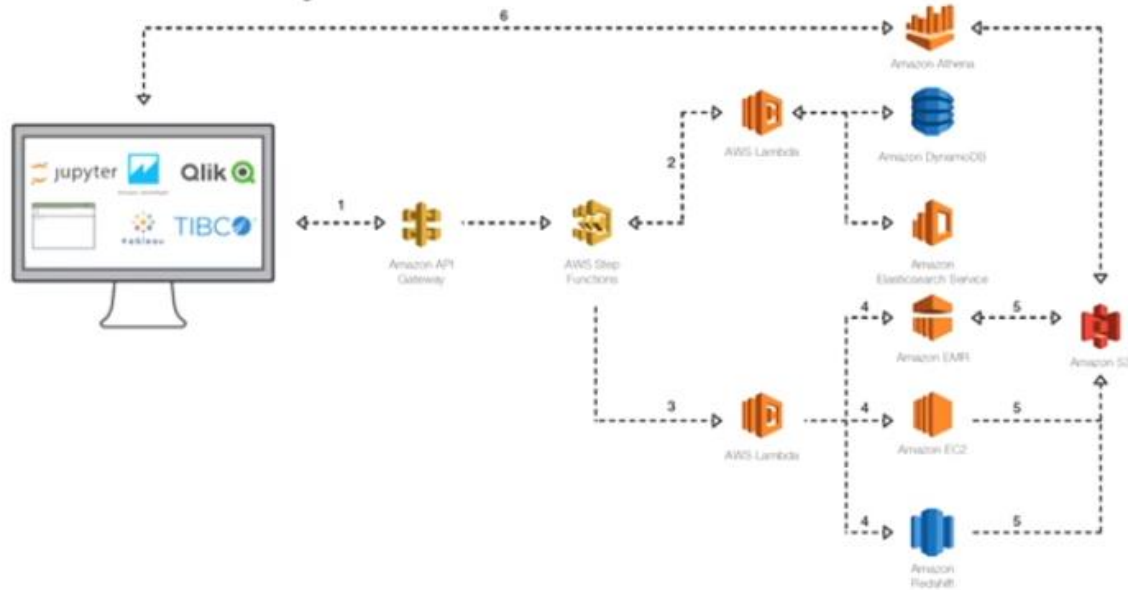
# Data Processing

Data Catalog

Data Processing and extract, Transform, Load (ETL)



# Data Analysis



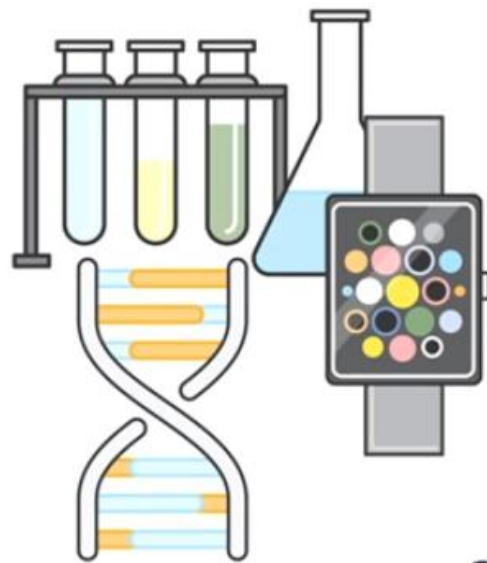
AWS re:Invent

aws

# Data in Health Care and Life Sciences



AWS re:Invent



aws



# Synapse

Celgene's Platform for Therapeutic Innovation

**AWS**  
**re:Invent**

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



## THE SPEAKERS



### DAN HOUSMAN

Consulting Managing Director,  
ConvergeHEALTH  
Deloitte Consulting, LLP

Dan Housman is a Director with Deloitte Consulting. He is a software veteran with a scientific education at MIT in Chemistry and Biology. He brings a strong knowledge combining a passion for medicine and a demonstrated track record of providing valuable and innovative product management for supporting complex distributed data analytics systems.

Dan directs ConvergeHEALTH's product innovation efforts with a focus on creating packaged products for real world evidence and bioinformatics through the Deloitte data warehouse platform and Miner Suite.

**AWS**  
**re:Invent**



### PATRICK LOERCH

Sr. Director, Data Sciences  
Celgene

Patrick Loerch is a Sr. Director with Celgene. He is a biostatistician and biochemist by training with 15+ years of experience analyzing genomics, clinical and real world healthcare datasets in roles throughout the pharmaceutical pipeline; from target identification through to commercial operations. He is responsible for building out the core data sciences capability at Celgene. In addition to building out the global team, he actively codes in R and Python and brings a strong knowledge and appreciation of the challenges associated with applying scalable data sciences in healthcare to create value for patients.



## Companies Are Prioritizing Evidence Generation...

Deloitte found in our Real World Evidence Benchmarking Survey from March 2017 that real-world evidence is a top priority for the industry, and companies are looking to transform capabilities through focus and investment

**54%** believe their current real-world evidence capability is not meeting the needs of the organization and are investing in expanding them

**80%** of companies with mature evidence capabilities utilize cloud-based systems to support their evidence work

**93%** are seeking knowledge management solutions that will enable broad sharing of information around the organization about studies conducted, evidence generated, and data available

Our survey shows that life sciences companies are making progress in using real-world evidence but still have opportunities to expand applications across the value chain, consider new channels to access evidence, and improve overall capabilities

**AWS re:Invent**



Evidence Lifecycle Management Data Lake implementation to support Real World Evidence RWE platform via data

## Emerging Trends in Application of Evidence

Major wave of investments by pharma created market learnings that are now driving a shift across industry in operating models and technology choices



Current state



Emerging model

**RWE mandates** Observing and understanding



Shaping and influencing

**IT & Informatics** Tightly coupled data and analytics solutions for technical "experts"



Standards-based, modular architectures to empower domain experts

**Source of data and tools** Data providers seen as vendors and suppliers



Data providers seen as partners and collaborators

**Value creation** Evidence generation primarily supports commercial objectives



More in-depth understanding of disease, epidemiology, and treatment standards, creating evidence to inform product strategy, design, and patient care

**AWS re:Invent**



# Achieving agile dev/ops efficiency for evidence development

Success in evidence life cycle management depends on decreasing cycle time to anticipate data supply needs from sources by iterating to find and fill gaps through collaboration with data providers

Insight Pipeline



AWS re:Invent



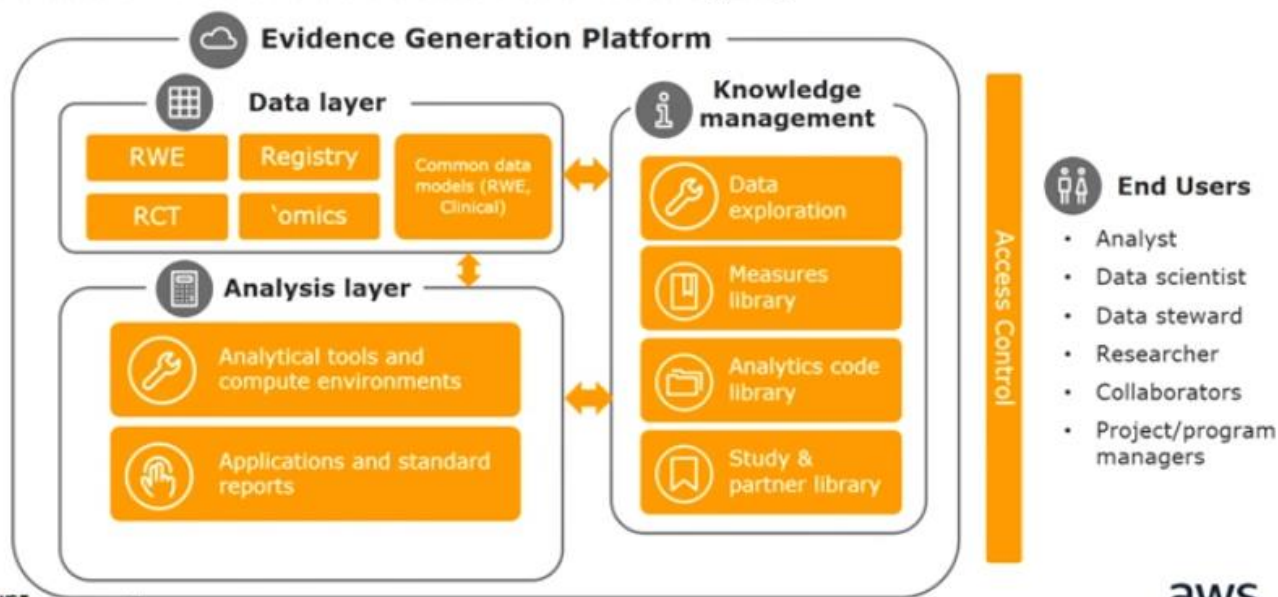
ConvergeHEALTH™



Data Providers



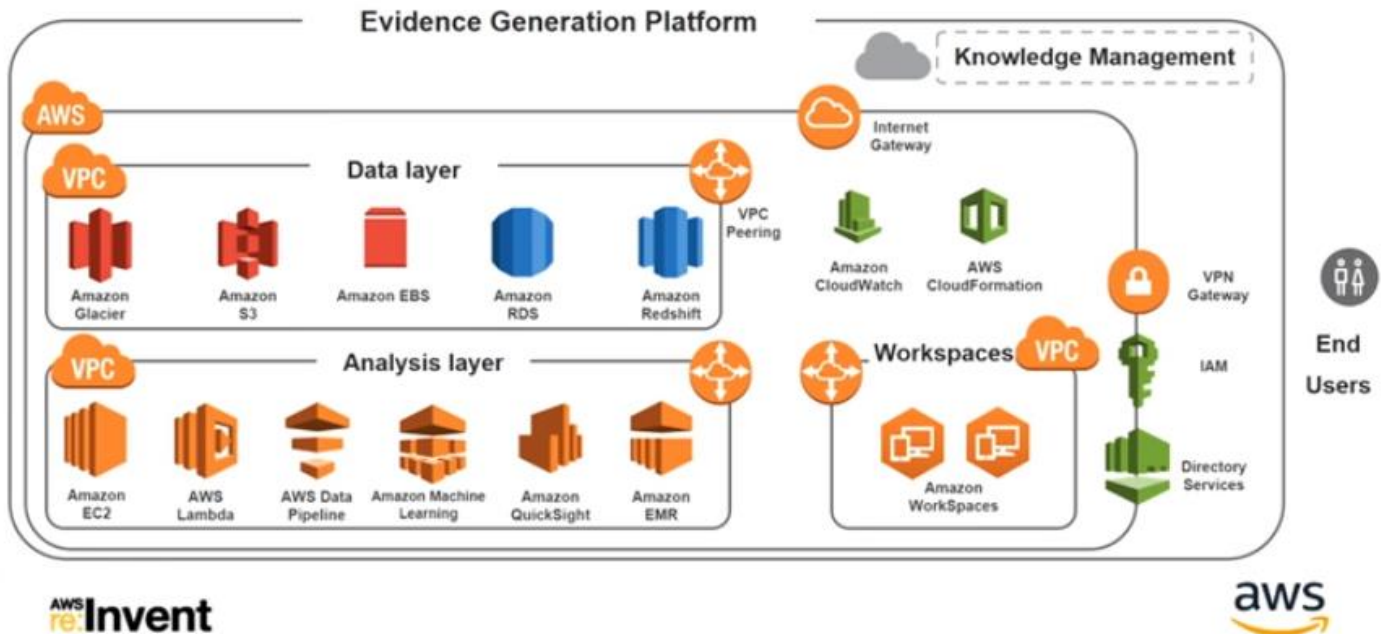
## New Integrated Analytics, Knowledge Management & Collaboration Platforms Are Emerging



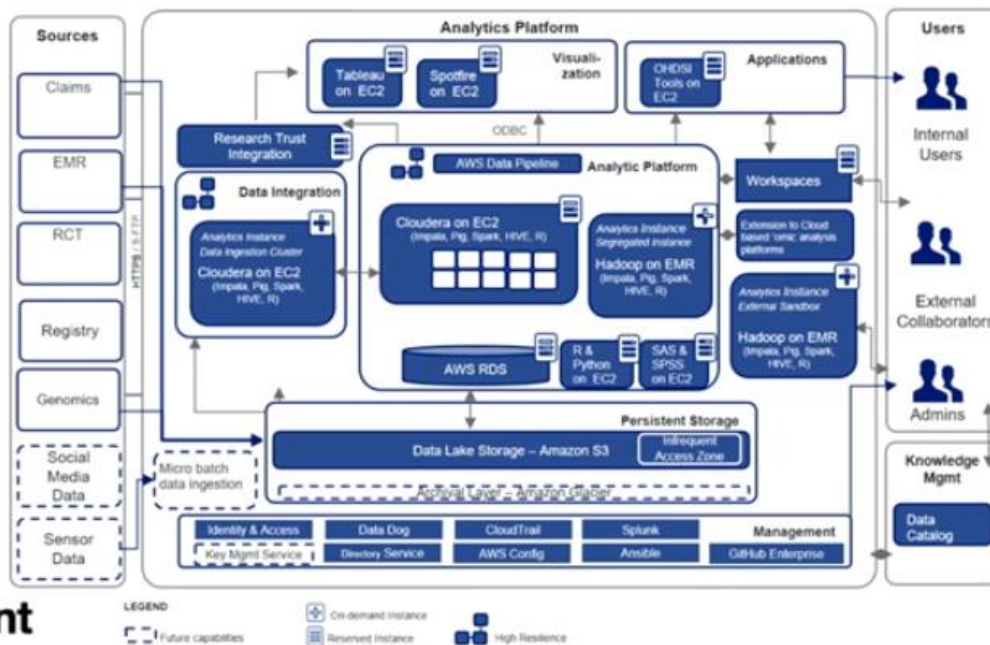
This is the logical architecture of the RWE analysis platform for de-identified patient data from different sources.



## Conceptual Technical Architecture



## Deeper Dive: The Technical Architecture



This is a HIPAA eligible backend architecture



# iKU: Transforming How We Think, Act, and Lead with Data



## Build Foundation

- Connective Technology (*Synapse platform*)
- Data Harmonization (*Governance & taxonomy*)



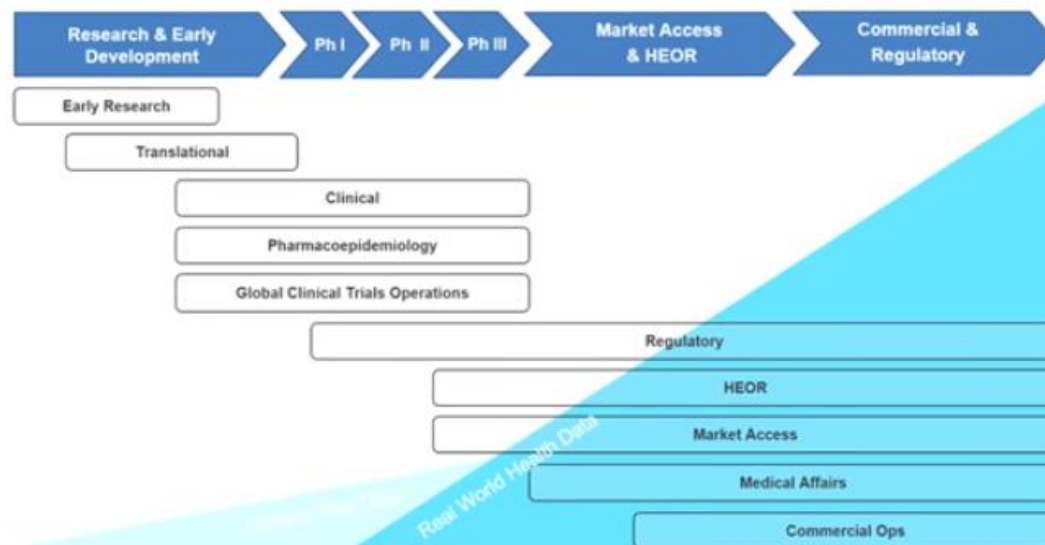
- Network of Strategic Partners
- Organizational Expertise
- Pioneering Mindset



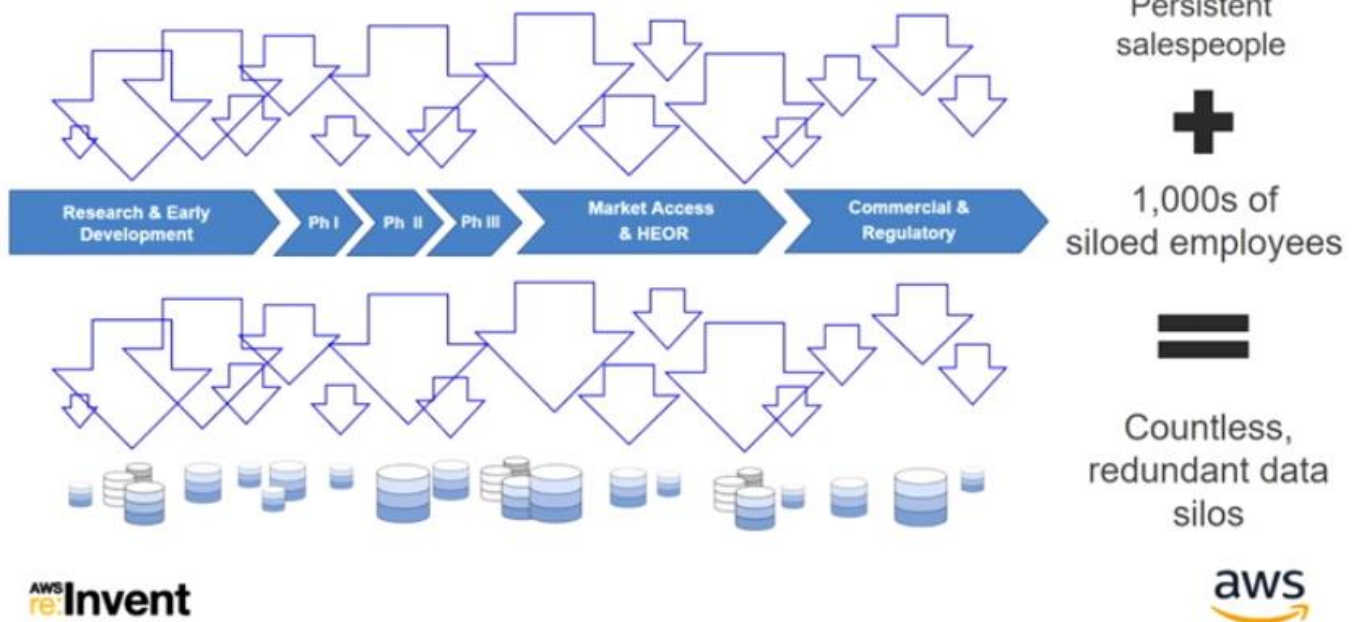
- Learn
- Collaborate
- Pioneer



## Patient Data Drives Decisions Across the Pipeline



## Then: The Rise of the External Data Vendor



## Today: What is Our Data Footprint?

	Dataset	Refresh Rate	Population Coverage	Data Type Coverage	
	Safety Data	Daily (1 day lag)	Subset of Patients	Safety Data	
	Patient Engagement	Weekly (no lag)			
	Sales Operations #1	Weekly (no lag)		Sales Data	
	Sales Operations #2	Weekly (1 week lag)	Subset of Providers		
	Sales Operations #2	Weekly (3-4 day lag)	Subset of Providers	Sales Data	
	Prescriber Network	Weekly (no lag)	Reference Data	Reference Data	
	Shipments Data	Weekly	Subset of Providers		
	Demand Data	Monthly (1 month lag)	Subset of Providers		
	Healthcare Org Services Data	Monthly (1 month lag)	Subset of Providers	Reference Data	
	Partner Data #1	Monthly (1 month lag)	Disease Population	EMR Data	
	Claims Data #1	Quarterly (4 month lag)	US Population	Claims Data	
	Claims Data #2	Quarterly (6 month lag)	US Population	Claims Data	
	Claims Data #2 (OMOP)	Quarterly (6 month lag)	US Population	Claims Data	
	Claims Data #3	Quarterly (6 month lag)	US Population	Claims Data	
	Claims Data #3 (OMOP)	Quarterly (6 month lag)	US Population	Claims Data	
	Provider Network Data #1	Quarterly (6 month lag)	US Providers	Disease Cohort	
	Provider Network Data #2	Quarterly (6 month lag)	US Providers	Refined Cohort	
	Partner Data #2	Monthly	Disease Population	EMR, Genomics & Lab Data	
	Partner Data #3	Quarterly	European Populations	EMR, Genomics & Lab Data	
	Partner Data #4	Monthly	Disease Patients	EMR, Genomics & Lab Data	
	Market Research Data #1	Monthly	Subset of Providers	Chart Abstraction	
	Market Research Data #2	Monthly	Subset of Providers	Chart Abstraction	

**Market Access** (Safety, GCRDO)  
**Medical Affairs** (R&D)  
**Comm Ops**  
**Brand Teams**  
**HEOR**  
**Real-World Partnerships**  
**Market Research Data**

**aws re:Invent** **aws**

\*to be ingested

■ Narrow Coverage  
■ Broad Coverage  
■ Slow Refresh  
■ Rapid Refresh

Every row represents data sets that currently sits within the platform, the columns are the variables about the data set. The side labels show the different groups within the company that are accessing and using the datasets. This allows us to see what data is being used by groups within the company.

# Understanding Our Data Gaps



AWS re:Invent

aws

This view shows the different diseases the company is engaged in and the number of patient datasets available for each disease area.

# Cultivating a Network of Data Partners

## Value

- Cultivating **clinical and genomic data** on previously uncharacterized patients to drive insights across the value chain
- Leverage **emerging technologies and skill-sets** to augment internal capabilities

## Approach

- Risk/balance portfolio approach with **near, mid and long-term value creation**
- Partnerships aligned with **Celgene's strategic direction**
- Deal structures tailored to **aligned interests** with partners
- Active engagement through **business development and alliance management**

FLATIRON

M2GEN

TriNetX

DANA-FARBER  
CANCER INSTITUTE

BROAD  
INSTITUTE



Deloitte  
ConvergeHEALTH™

patientslikeme®

IMIDomics

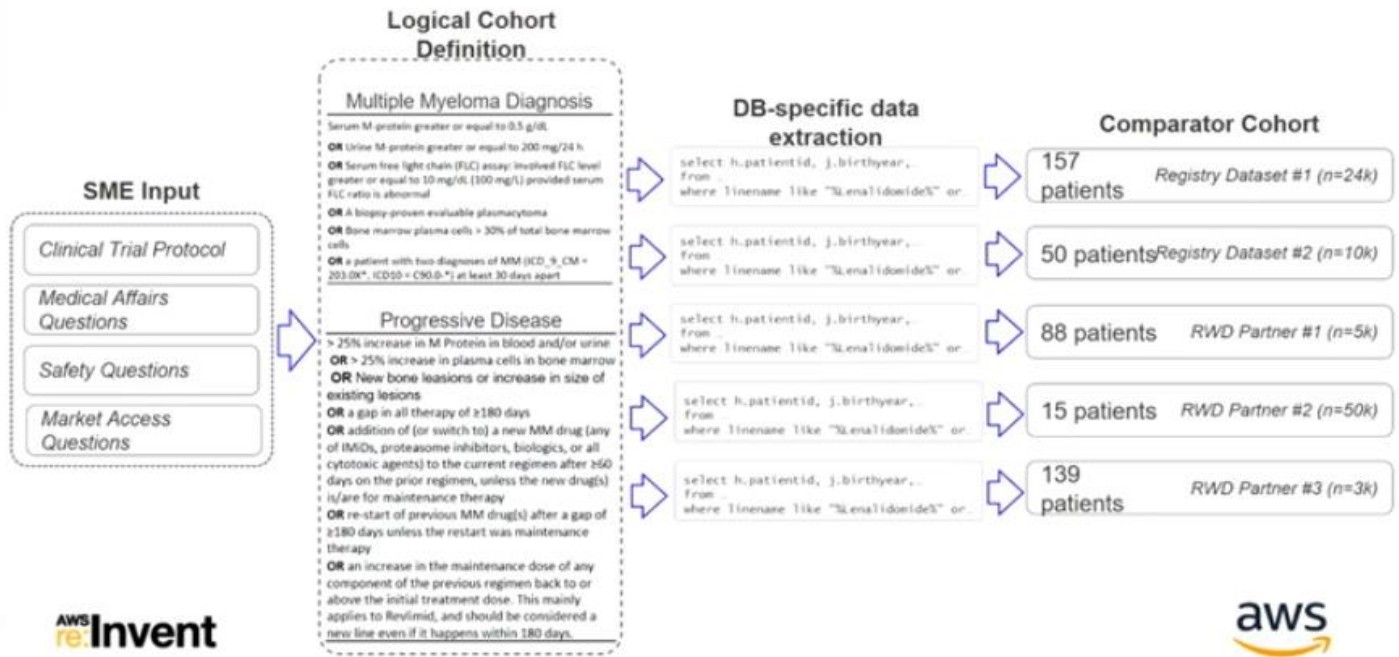
COTA  
LESS RISK. MORE VALUE.

aws

AWS re:Invent



# Execution Focused on the Development of Novel Medicines



## Knowledge Management Platform

### Summary Statistics

- 388 RWE databases tracked
- 24 DBs ingested & regularly refreshed
- ~120-150M patients per claims DB

### Features

- "Big Data" Analytics
- Data governance
- Code governance
- Interactive visualizations
- Self-serve applications (non-coders)



When you look at the platform from a user perspective, this is essentially what you will see regarding the stacks of functionalities available. There are reusable queries for getting specific data sets out of the data lake.



# Ex.: Real-World Treatment Pathways in Crohn's Disease

## Introduction

- Treatment for CD has advanced over the past 20 years with the introduction of biologics
- Despite the availability of biologics, patients may not be optimally managed

## Objective

- The aim of this study is to identify and visualize CD treatment pathways to gain insight into real-world treatment patterns

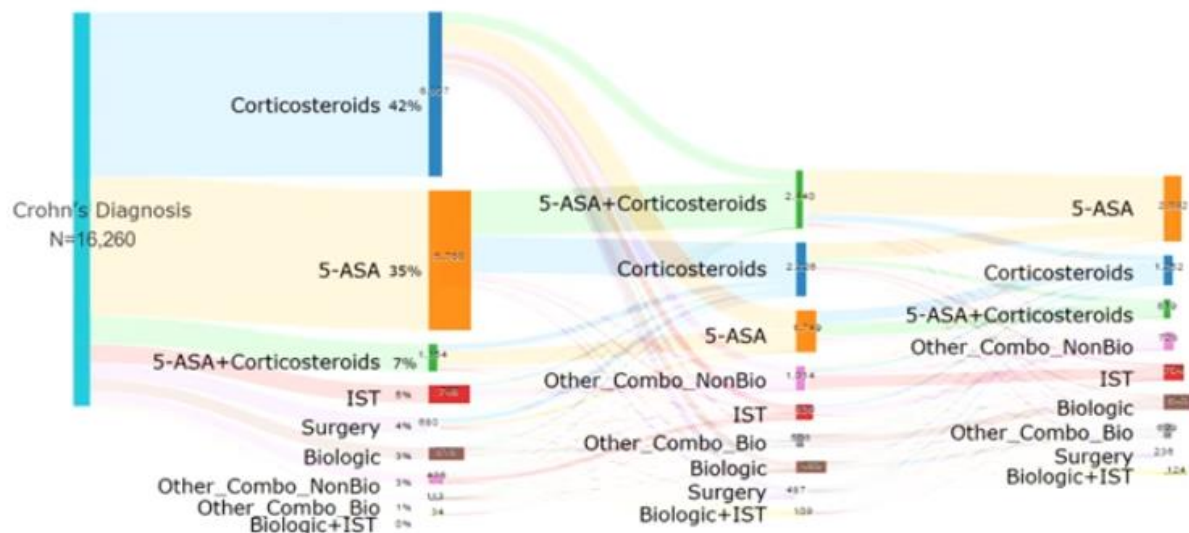
## Methods

- The MarketScan Commercial and Medicare Databases were used to assess treatment pathways in a large US insured population
  - Patients had  $\geq 2$  consecutive health claims for CD\* or UC†  $\geq 30$  days apart, with  $\geq 1$  occurrence of NDC/HCP codes for CD or UC medications from January 1, 2008, to March 31, 2016
  - Required  $\geq 3$  (1 pre-diagnosis + 2 post-diagnosis) years of continuous enrollment

AWS  
re:Invent

aws

# Ex.: Real-World Treatment Pathways in Crohn's Disease

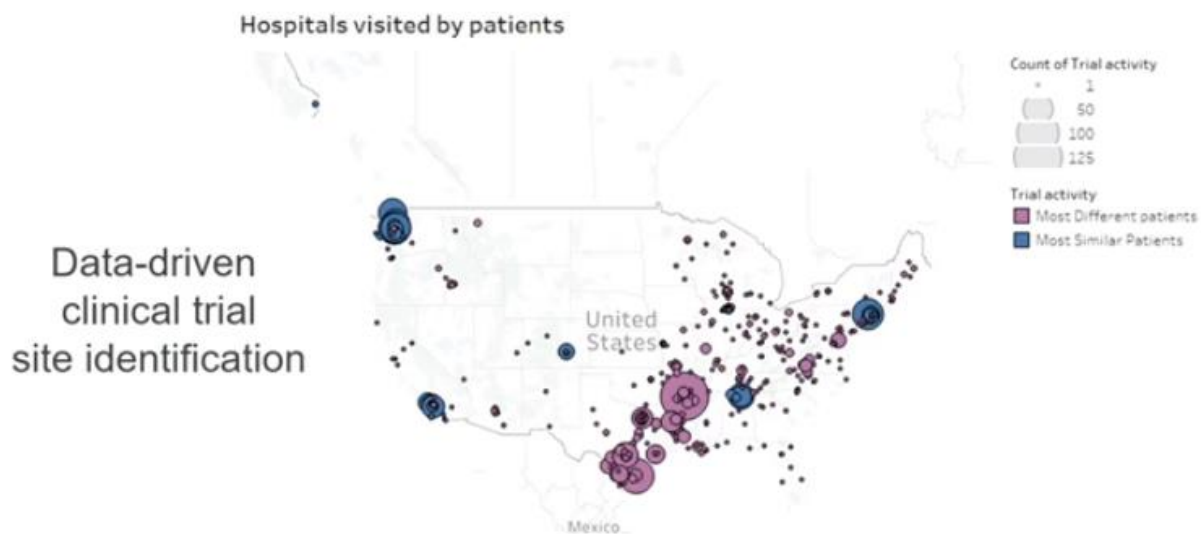


AWS  
re:Invent

aws

5-ASA=5-aminosalicylic acid; IST=immunosuppressant (i.e., immunomodulator)

## Ex.: Real-World Treatment Pathways in Crohn's Disease



AWS  
re:Invent

aws

## Achieving agile dev/ops efficiency for evidence development

Success in evidence life cycle management depends on decreasing cycle time to anticipate data supply needs from sources by iterating to find and fill gaps through collaboration with data providers



# Takeaway

- The changing health-care landscape requires pharmaceutical companies to increasingly become data-driven organizations
- Data needs to be proactively cultivated and evolve within the context of current and upcoming medicines in the discovery and development pipeline
- As in other industries, IT and information platforms need to encourage data, knowledge, and code sharing...with the appropriate access controls
- Through working with Deloitte, and leveraging AWS, Celgene has developed an industry-leading, global platform spanning from data ingestion to knowledge sharing

