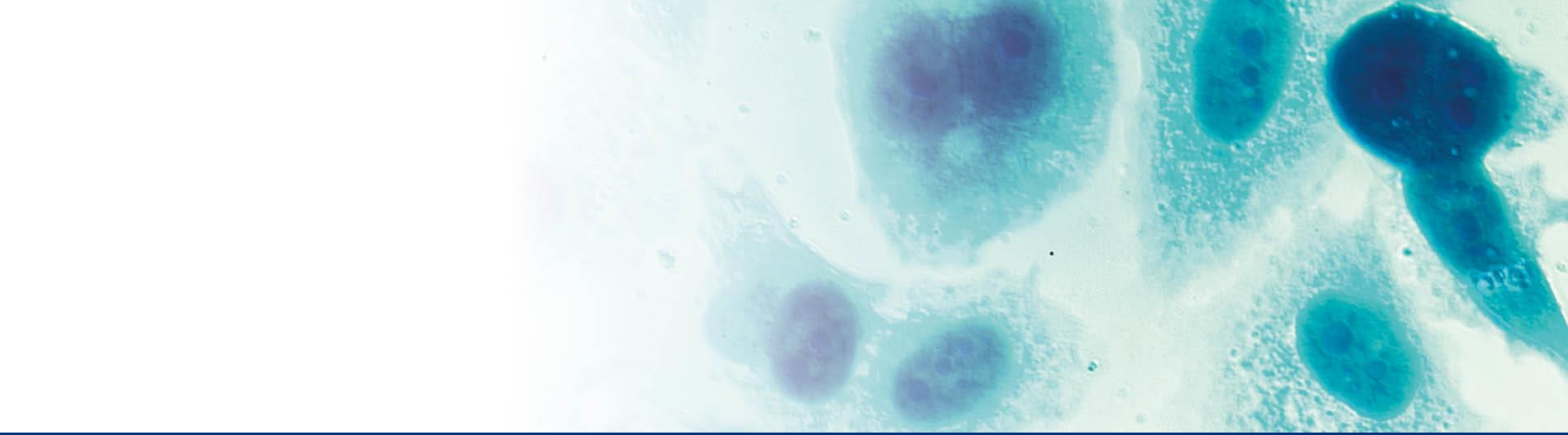
**Mr. Bart Vannieuwenhuyse**
**Data Sciences Lead Benelux campus**

# Big Data and its added value in clinical research

CONGRESS '19
EAHM
BELGIUM - GHENT



# Big data and its added value in clinical research

Pictured above: Ulcerative Colitis

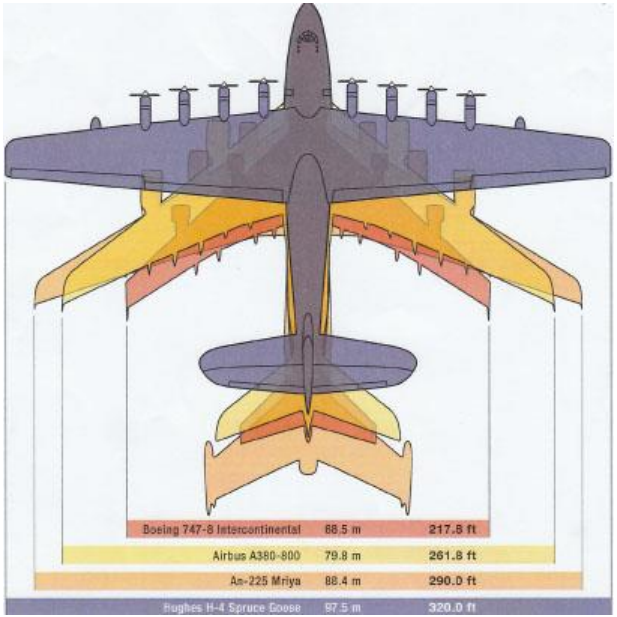
Bart Vannieuwenhuyse
September 2019 | Janssen Clinical Innovation

## “The patients are waiting!”

To eliminate disease through developing highly innovative medical solutions for people around the world

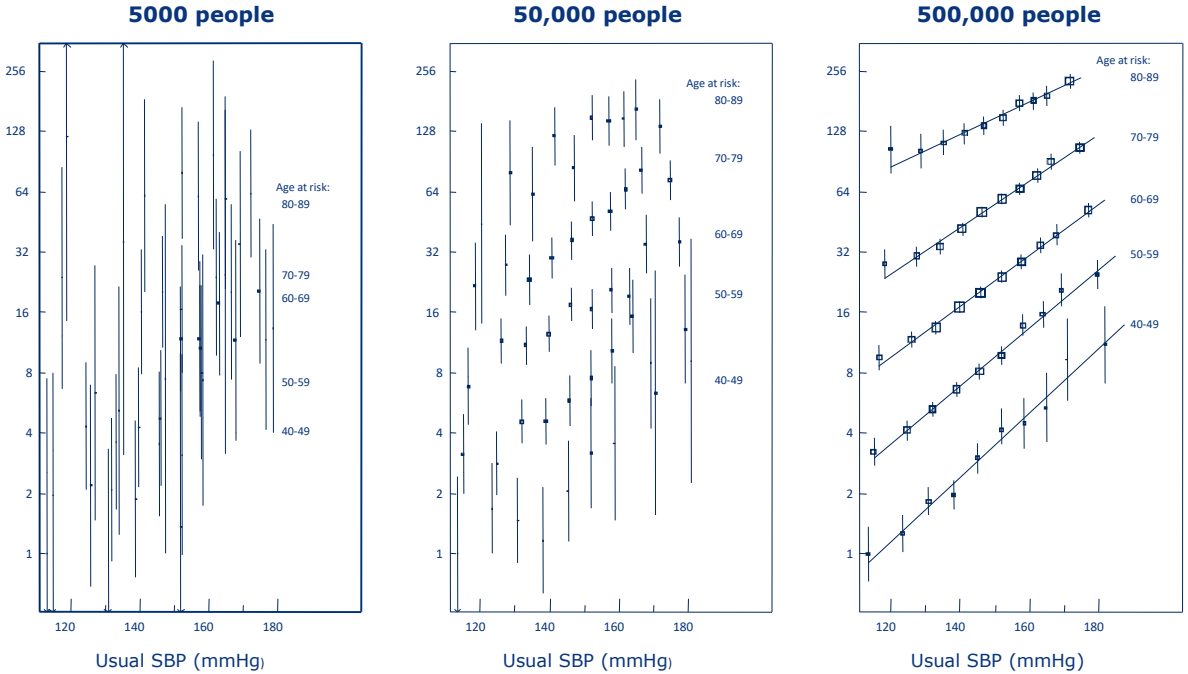
Dr. Paul Janssen

# Bigger = Better ??



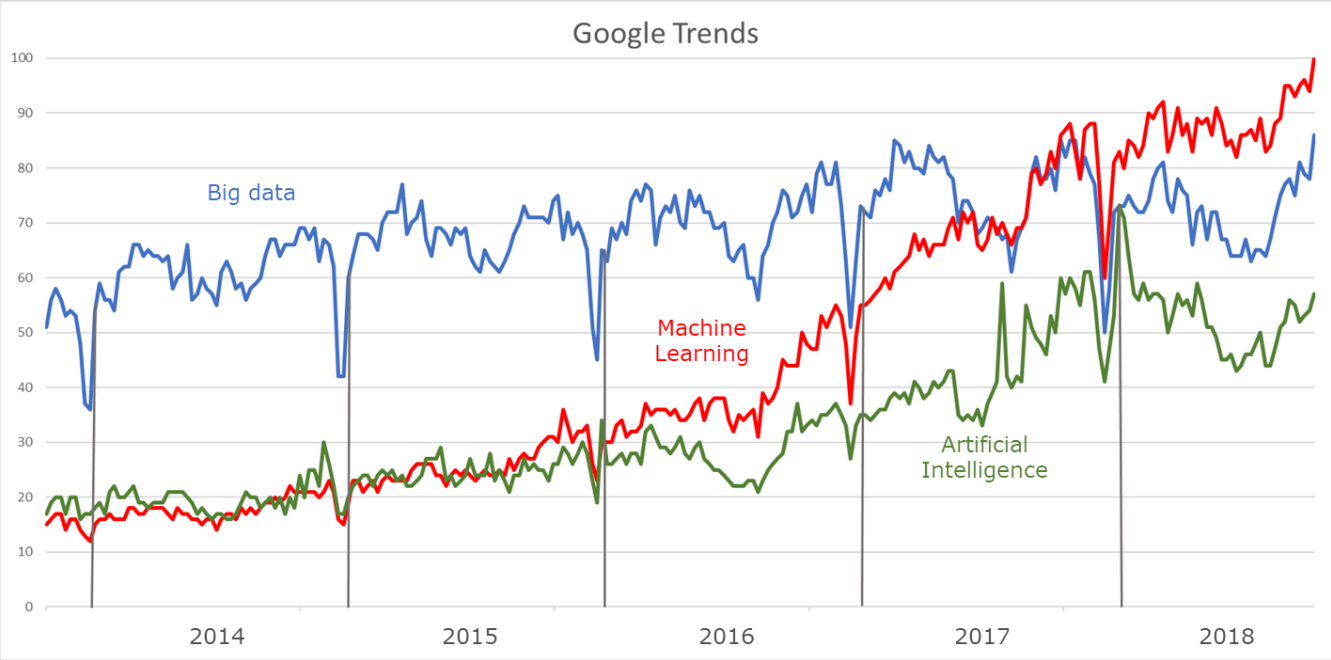

Howard Hughes' plane only flew one mile …

# Power in numbers …

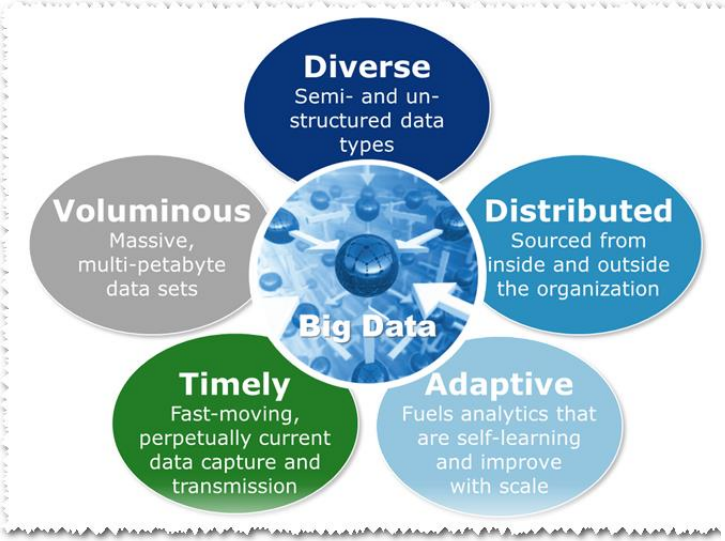



**The Prospective Studies Collaboration: Lewington et al. 2002**

# Machine Learning has overtaken Big Data

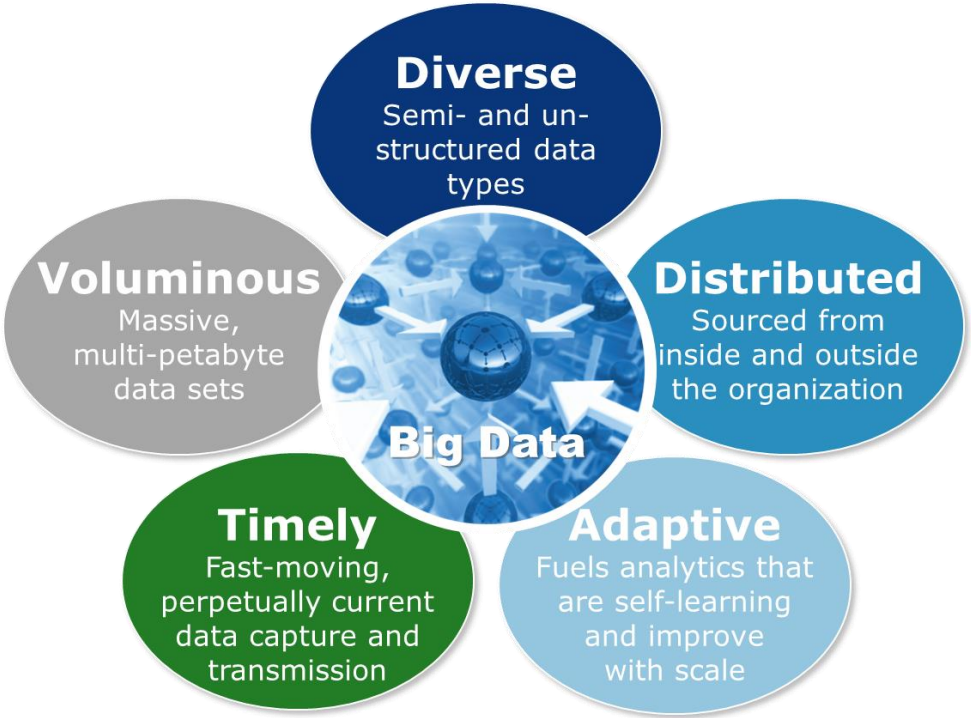


# Big data vs Real World Data



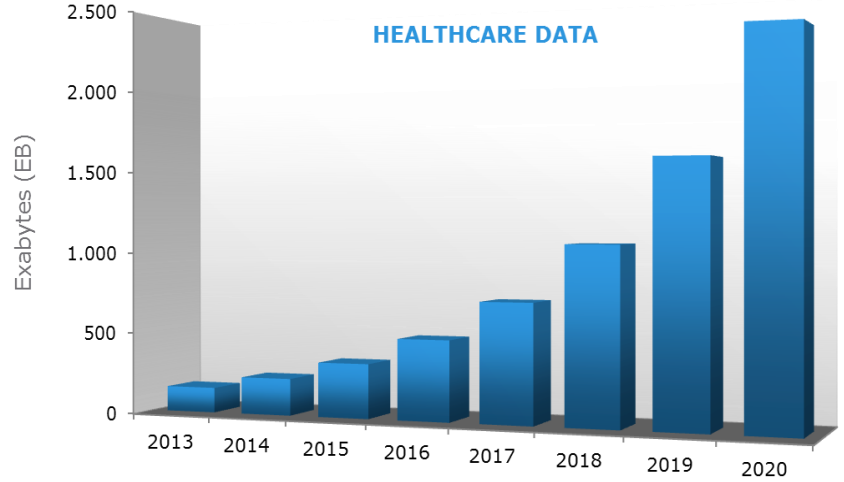

- RWE is generated using data typically collected in **usual health care settings**. RWE is most commonly **generated using a range of non-interventional (observational) studies**, including:
  - **Primary data** collections such as **registries** collecting prospective and/or retrospective data, or **surveys** collecting cross-sectional or retrospective information.
  - Analyses of **secondary data that includes (electronic) medical records, insurance claims data, and government databases** which provide data typically used for retrospective analyses.

# Big Data Defined

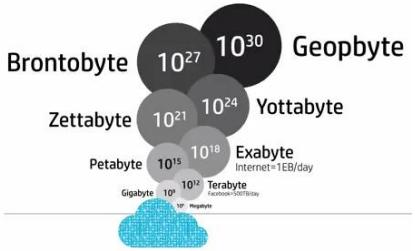


# Explosion of healthcare data

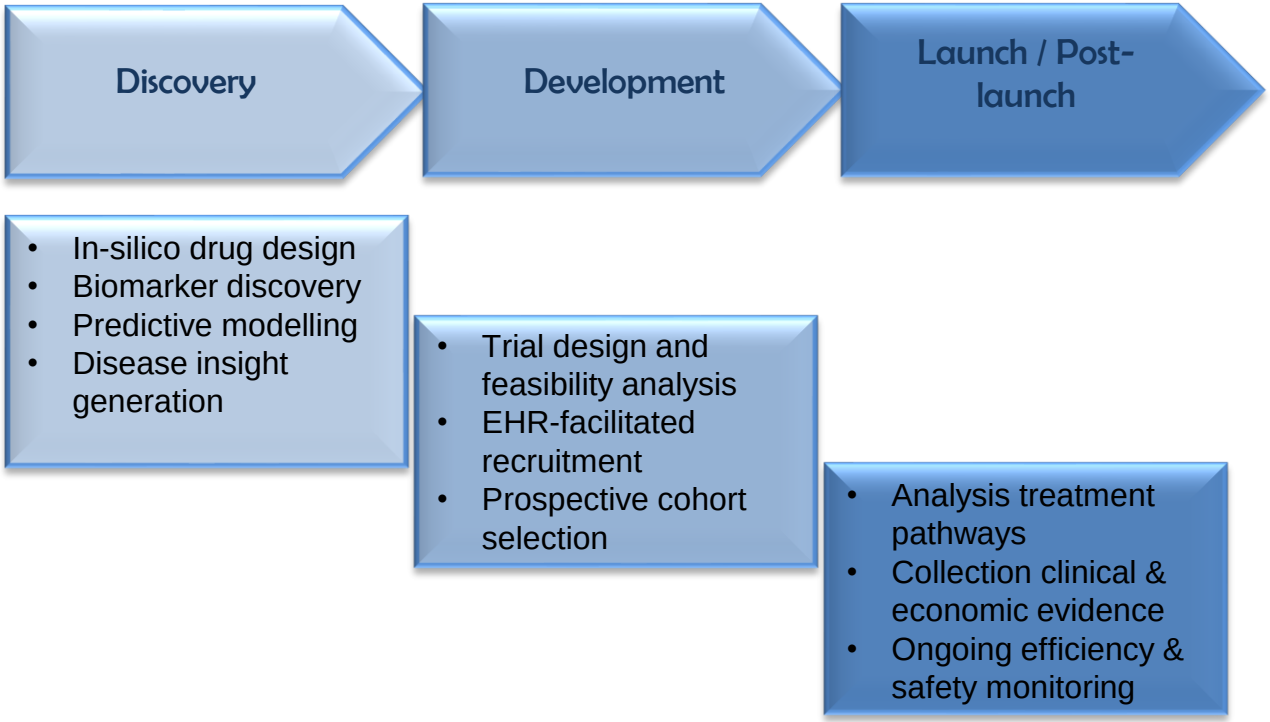
Healthcare data growth is one of the fastest across many industries.
A 48% annual growth rate will lead to 2,314 Exabytes of data in 2020.



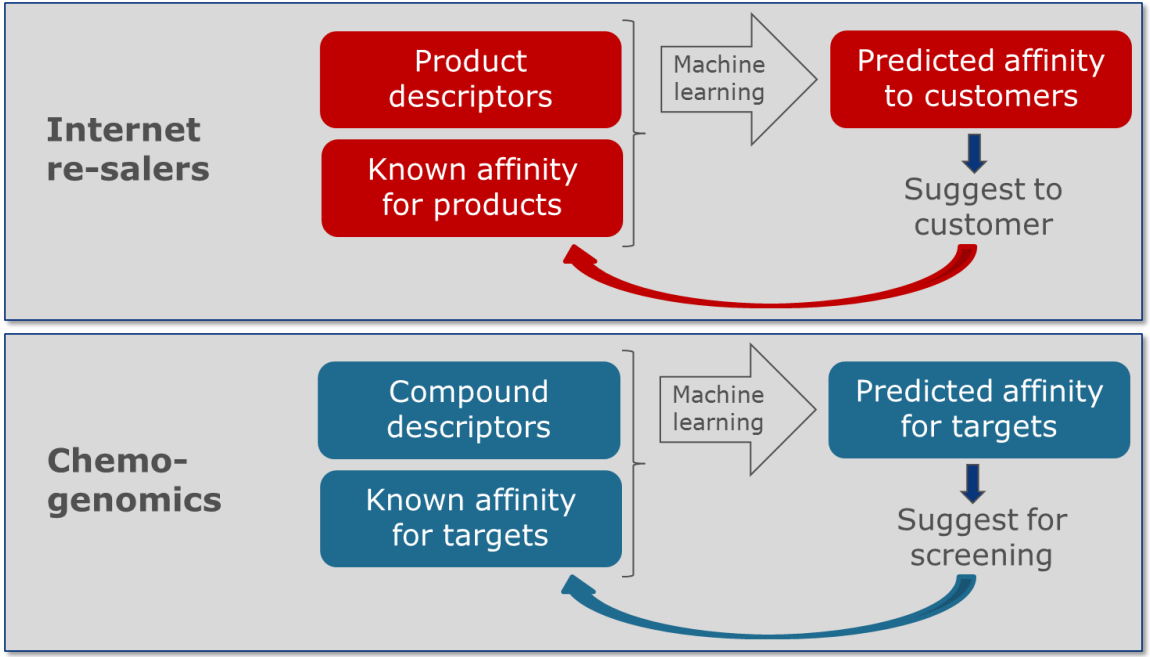

Enterprise Strategy Group 2011, Research report: North American Healthcare Provider Information



# Opportunities for Real/Big data



# Developing in-silico drug design

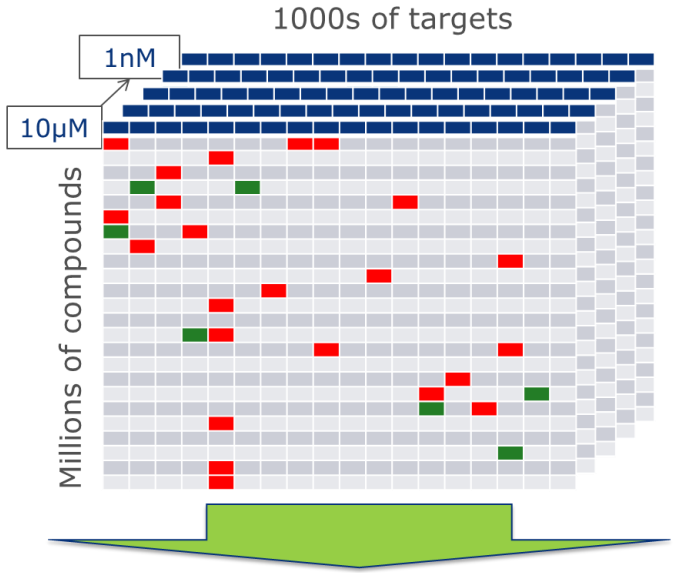



THE NETFLIX-IZATION OF DRUG DISCOVERY

# >100M data points with biochemical activities of tested compounds available for training

ABCD
GOSTAR
GVK BIO Online Structure Activity Relationship Database
THOMSON REUTERS
ChEMBL
OpenTox

>100M training points




Identify and select compounds with good biochemical activity for target(s) of interest.

MELLODDY
https://twitter.com/MELLODDY_IMI

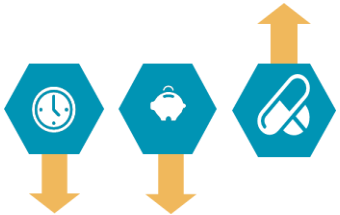
New IMI project in which 10 pharma companies collaborate

THE NETFLIX-IZATION OF DRUG DISCOVERY

# Challenges in trial execution

The percentage of studies that complete enrolment on time:
**18%** in Europe,
**7%** in the US

**Potential improvement through use of Real World Data**



**1/3** of protocol amendments are avoidable, at a cost of **$0.5m**

Almost **half** of all trial delays caused by patient recruitment problems

Only **1/3** of the sites engaged in a multicentre study manage to enrol the requisite number of patients

**50%** of today's clinical trials fail to achieve the target recruitment

Source: IMI-EHR4CR project

# EHR4CR

## The EHR4CR project

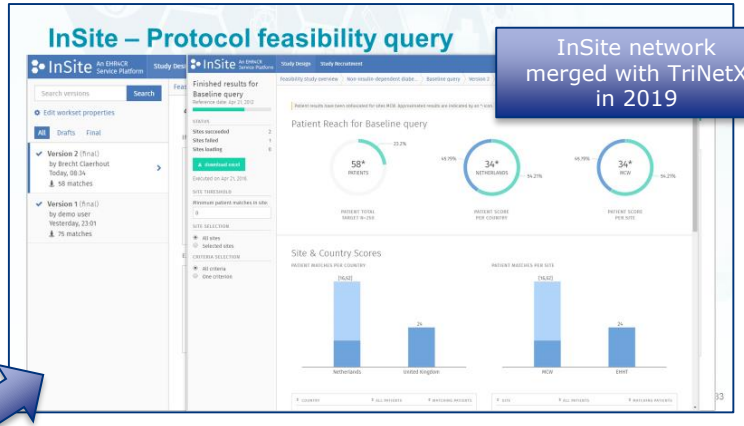
- EHR4CR – Electronic Health Records for Clinical Research
  - 4+1 year project (2011-2016), 35 partners, budget >17M€
- Objectives & Scope
  - Provide a platform for **trustworthy re-use of EHR data** to support innovation in clinical research and healthcare operations.
  - Securely reusing **health data** for optimising clinical trials
  - **7 pilot sites across Europe**
- Status
  - Extended into 2016 for making the transition to a sustainable platform
  - Initiated a **Champion Programme**, connecting hospitals to an operational platform, building up experience with pharma
  - The **European Institute for Innovation through Health Data** – an independent governance body

For more information:
http://www.ehr4cr.eu/

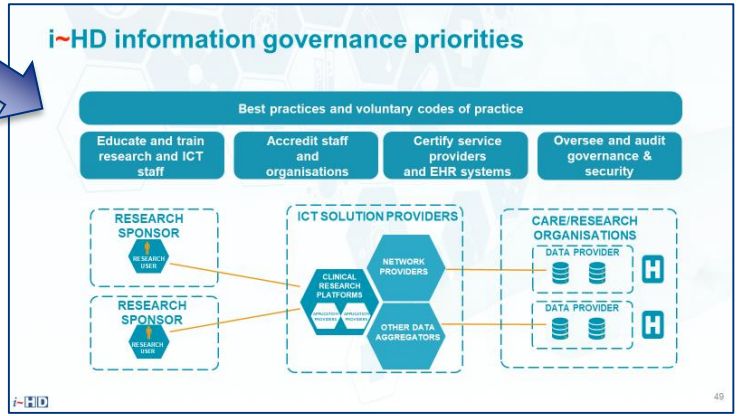
efpia

imi

i~HD 25



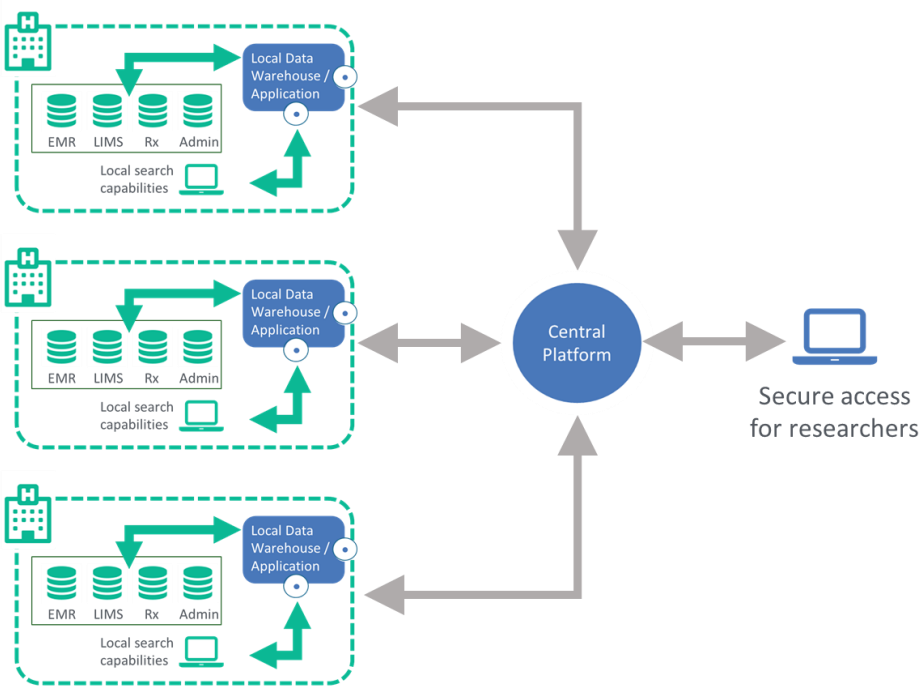

For profit – InSite platform




Non-profit – iHD institute

Source: presentation Prof Dipak Kalra

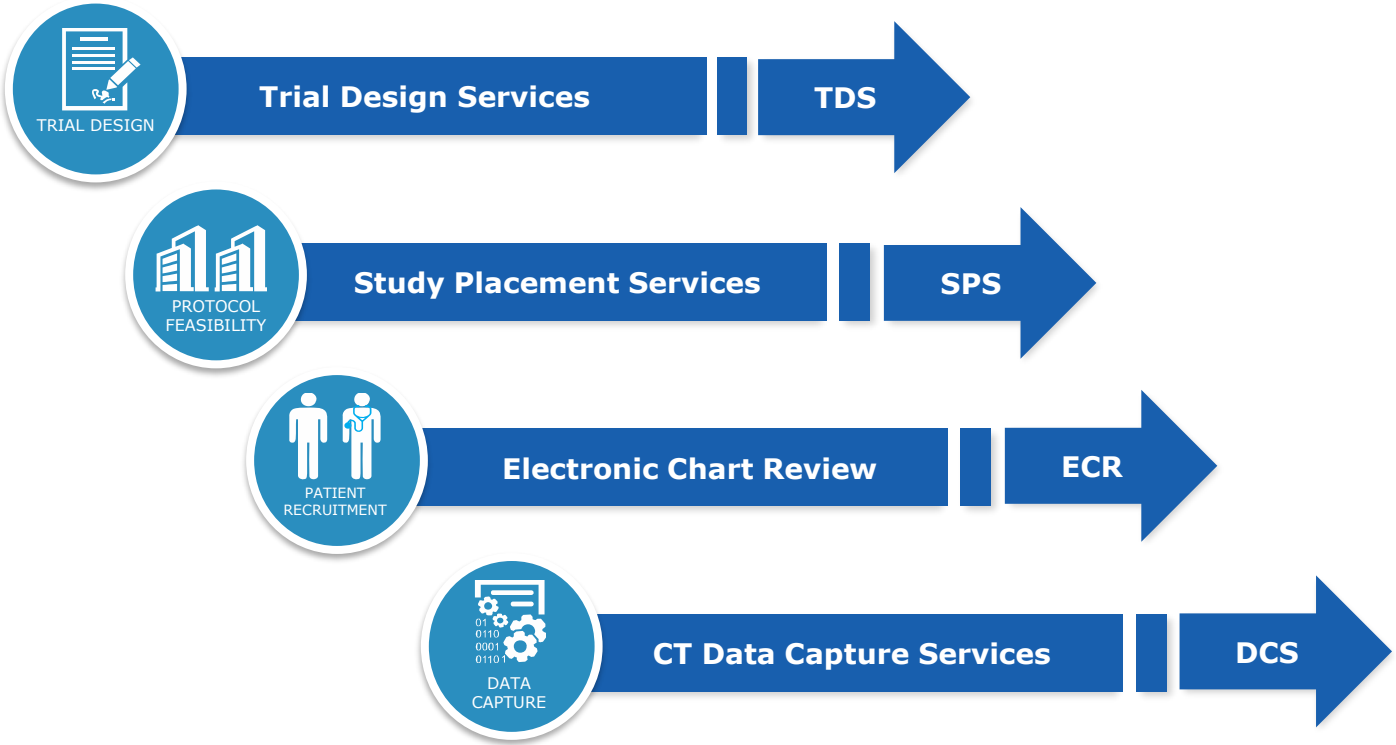
# Stimulating the flow of data through federated networks



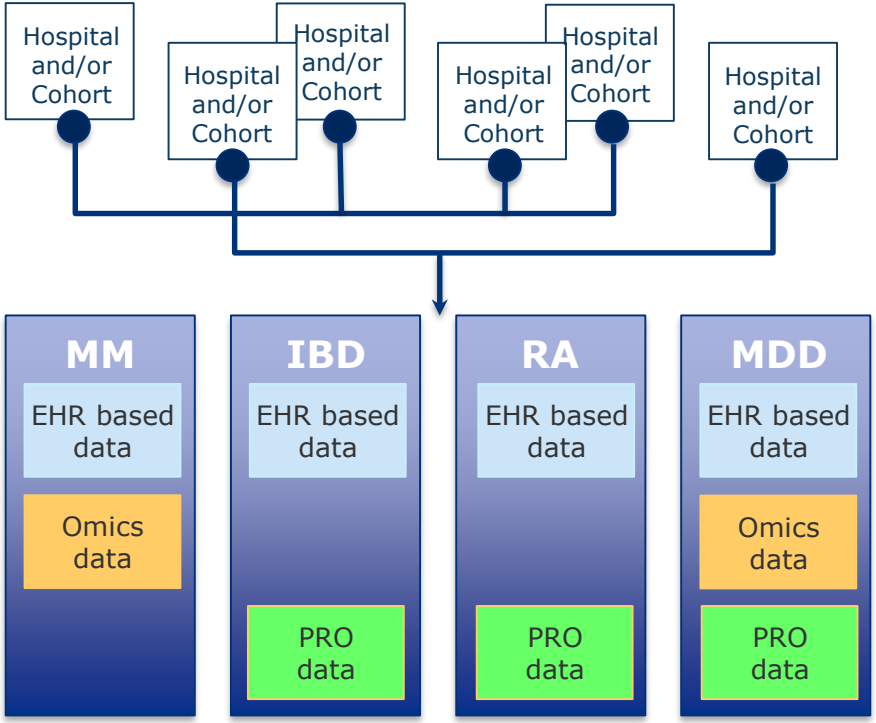

## Benefits of federated networks

- Data remains under the control of the data owner
- Locally required legal and ethical approvals apply
- No patient level data leaves the owner's site, only aggregated counts, thereby ensuring patient privacy
- GDPR – *'Privacy by Design'*
- Analysis is "brought to the data" rather than creating central data repository
- Use of common data model allows for efficient search / analysis across multiple data sets
- Requires close collaboration with data owners which builds trust

# Using federated data in clinical research



# From “broad” to “deep” data

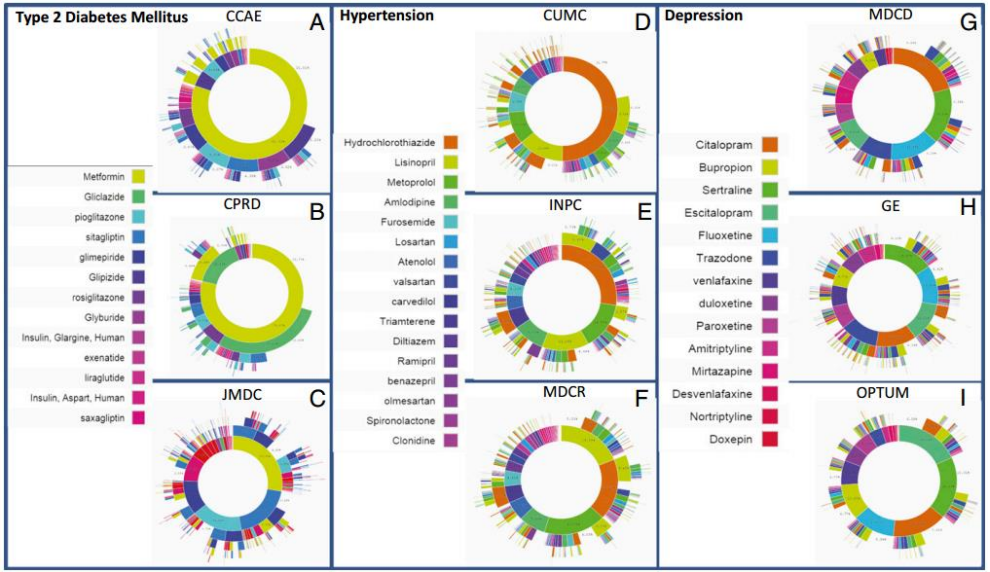



The broad network of RWD sources (hospitals, cohorts, other) can serve as starting point for enrichment (deepening).

By disease area, original RWD can be enriched based on the needs for (clinical) research, thereby creating a so called “T-shaped” data ecosystem (combining broad and deep data)

# True open science collaboration

# Power of distributed data --




Collaboration with
11 data sets representing
**255Mio** subjects

**Characterizing treatment pathways at scale using the OHDSI network**

George Hripcsak[a,b,c,1], Patrick B. Ryan[c,d], Jon D. Duke[c,e], Nigam H. Shah[c,f], Rae Woong Park[c,g], Vojtech Huser[c,h], Marc A. Suchard[c,i,j,k], Martijn J. Schuemie[c,d], Frank J. DeFalco[c,d], Adler Perotte[a,c], Juan M. Banda[c,l], Christian G. Reich[c,j], Lisa M. Schilling[c,m], Michael E. Matheny[c,n,o], Daniella Meeker[c,p,q], Nicole Pratt[c,r], and David Madigan[c,s]

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

# New IMI project

EHDEN
EUROPEAN HEALTH DATA & EVIDENCE NETWORK


- Harmonised 100 million records, mapped to the OMOP common data model
- Ecosystem of certified/qualified SMEs harmonising data sources, supported by €17mio fund via open series of calls
- Federated analytics via OHDSI tooling, ARACHNE and EMIF platforms
- Impact: Improving trust and reusability of real world data for key stakeholders

# Conclusions

- Big data offers opportunities along the full product life cycle
- Specific analytical skills and methods are required
- Analysis and approaches to big data need to take patient / subject privacy challenges into account
- Federated approaches can help to mitigate privacy challenges


Bart Vannieuwenhuyse
bvannieu@its.jnj.com