# Challenges for AI in Hospital Federations

**Yannis Ioannidis**

"Athena" Research and Innovation Center

Nat'l and Kapodistrian University of Athens

# MOTIVATION

# RARE DISEASES

- Rare disease*: Less than 5 in 10,000 people affected
- 5,000 to 8,000 rare diseases
- Estimated 1 in 15 people affected
  - 400M worldwide
  - 30M in Europe
  - 25M in US
- Mostly (80%) genetic

* 2013 WHO Background Paper 6.19 “Rare Diseases”, by S. van Weely and H.G.M. Leufkens

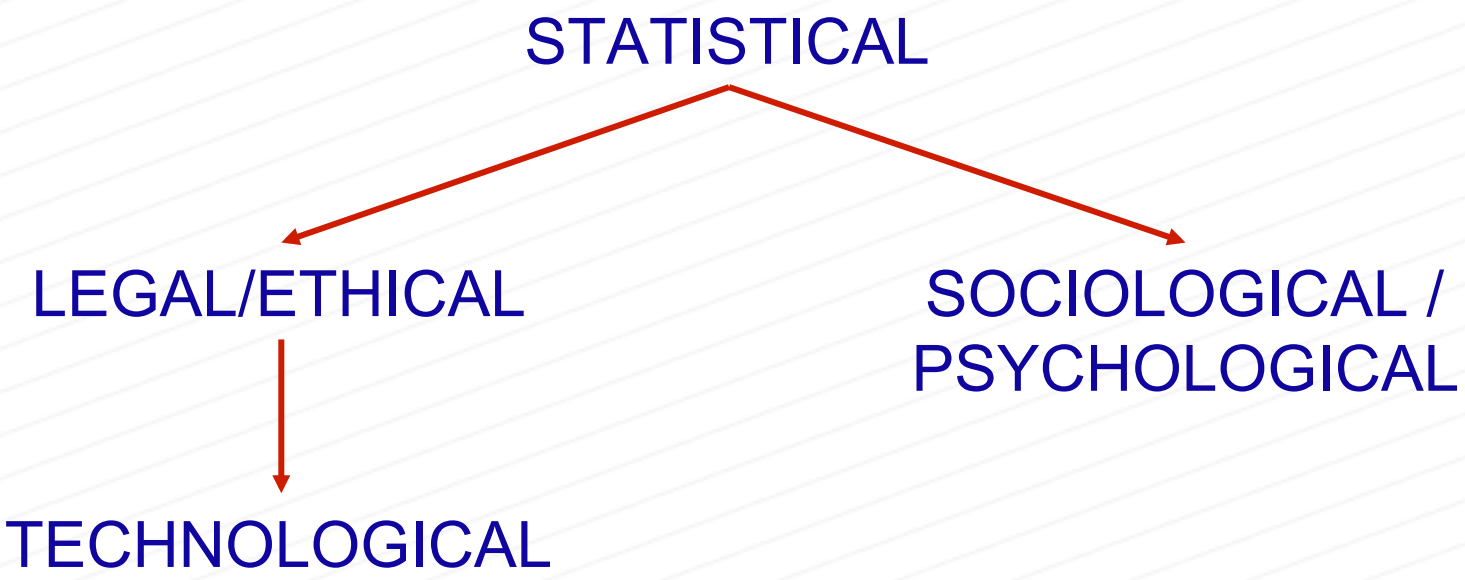
# European Reference Networks

## Treatment of patients with rare or complex diseases


https://ec.europa.eu/health/ern/work_en

# CHALLENGES

# CHALLENGES DERIVATION



# STATISTICAL

- Not enough data inside a single hospital / country
- Must study clinical data from “all” hospitals
  - Precision medicine
  - Finding biomarkers
  - Forming “Digital Twins”
  - …
- *Need hospital data federation*

# LEGAL / ETHICAL

- Privacy (laws, regulations, policies)
- Security (regulations, policies)
- *Need complete adherence in depth of time*

# TECHNOLOGICAL

- Heterogeneity
  - Laws, regulations, policies
  - Pathologies

    (in addition to data semantics, format, syntax)
- Algorithmic complexity
  - Data analytics / machine learning / modeling
  - Search
- *Need significant research & development advances*

# SOCIOLOGICAL / PSYCHOLOGICAL

- Patients losing own data
  - → (fear of) losing privacy, protection, money, …
- Clinicians losing “own” data
  - → losing discovery edge, power, money, …
- Administrators losing sleep
  - → losing protective power of hospital data center

- *Need incentives and guarantees*

# ECONOMICAL / POLITICAL

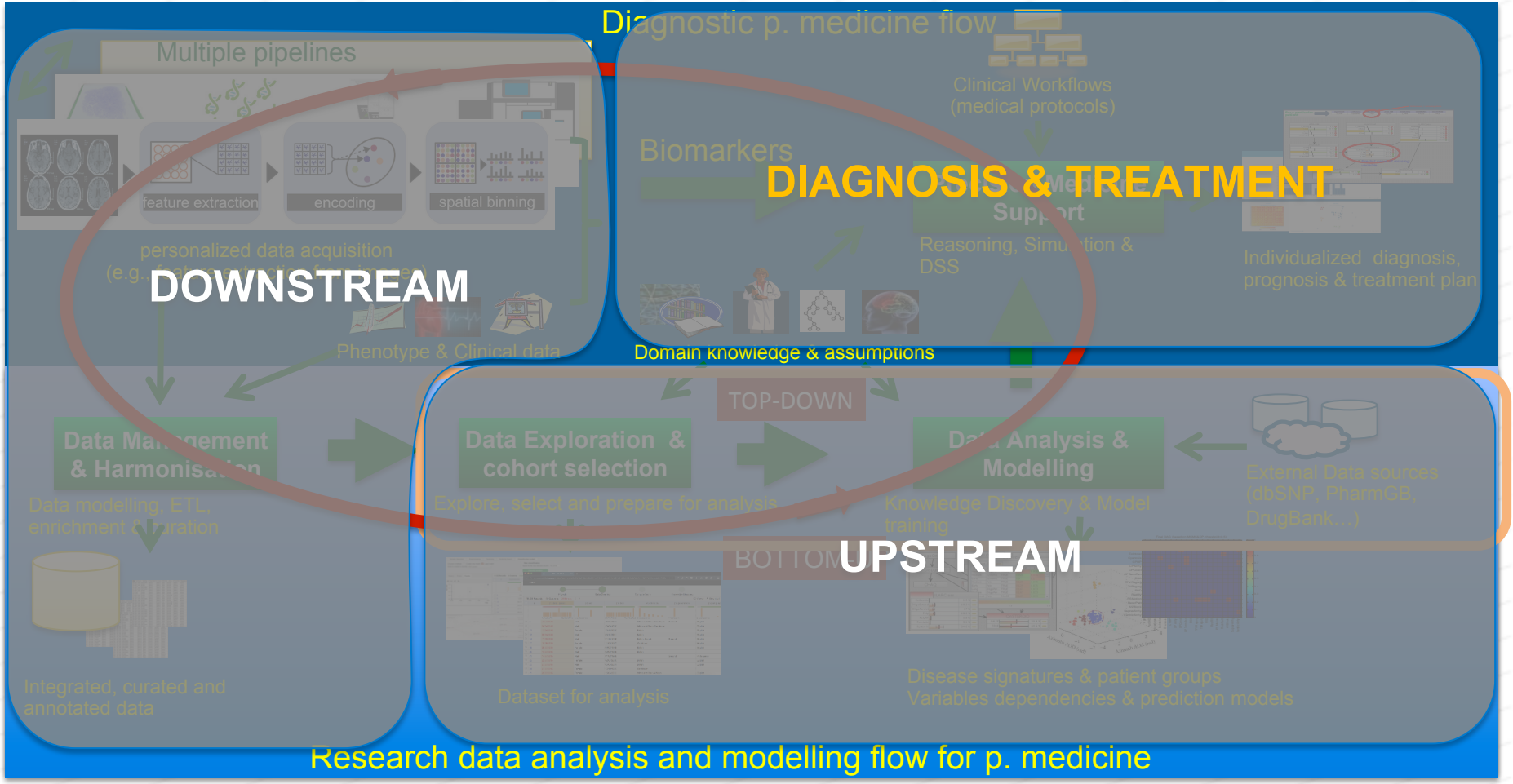
- Not enough money to be made by pharma
- Not enough votes to be gained by politicians

# VISION

# DATA DRIVEN P.MEDICINE

- Integrate patient care and medical and clinical research
- Accelerate identification of disease and effective treatment
  - biological signature of diseases
  - classifications of disorders
  - …
- Develop platform for hospitals and research centers to share and serve medical data
  - data acquisition, treatment, and analysis

# DATA BASED P.MEDICINE

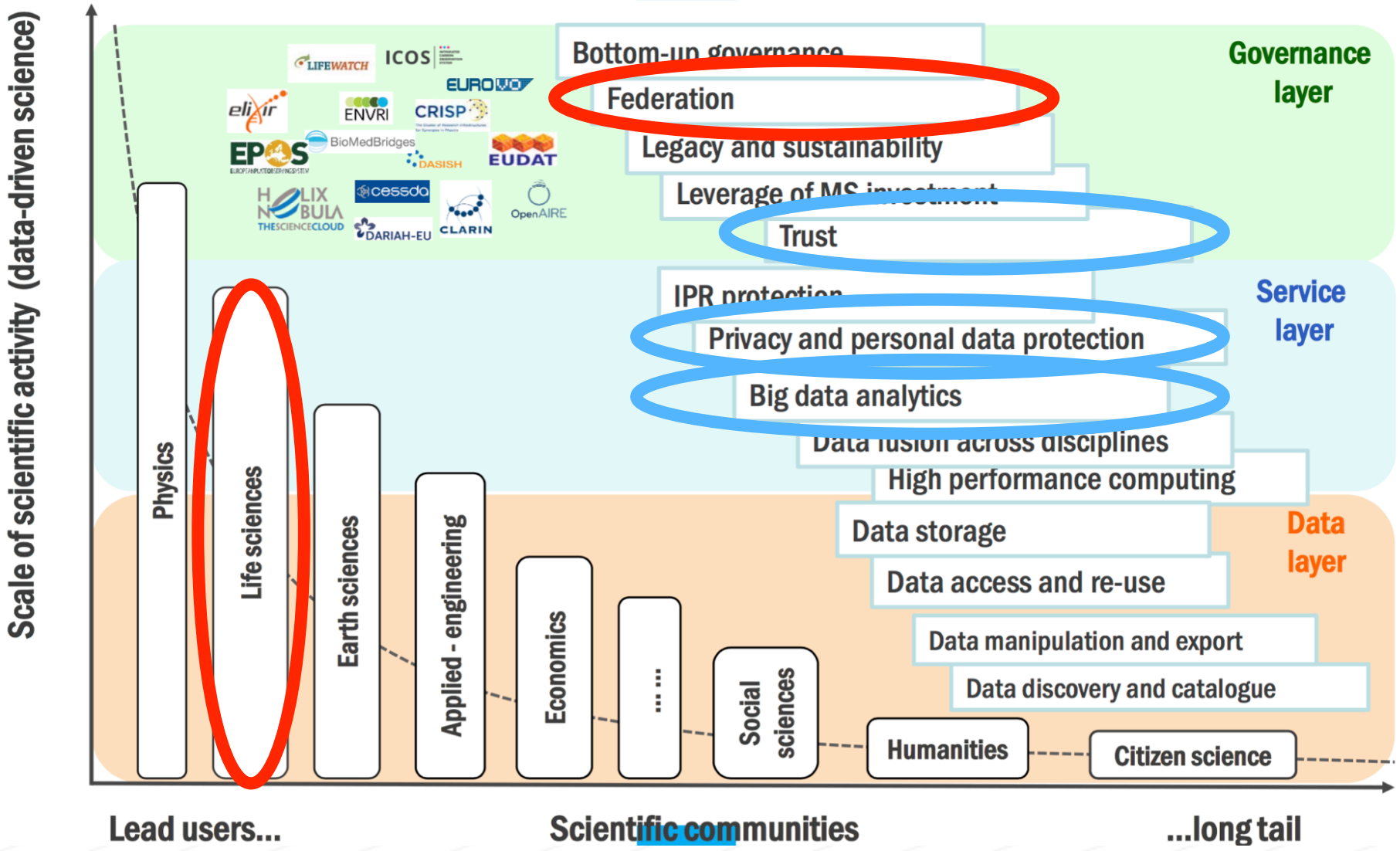


# DOWN and UP STREAMS

- Downstream (data acquisition, preparation)
  - Inside hospital of origin
- Upstream (data analysis, modeling, learning)
  - Inside hospital of origin for common diseases
  - Across hospitals for rare diseases

Build technology for the creation of data/service markets over shared hospital and individual patient data while adhering to all laws and regulations

# DATA DRIVEN SCIENCE



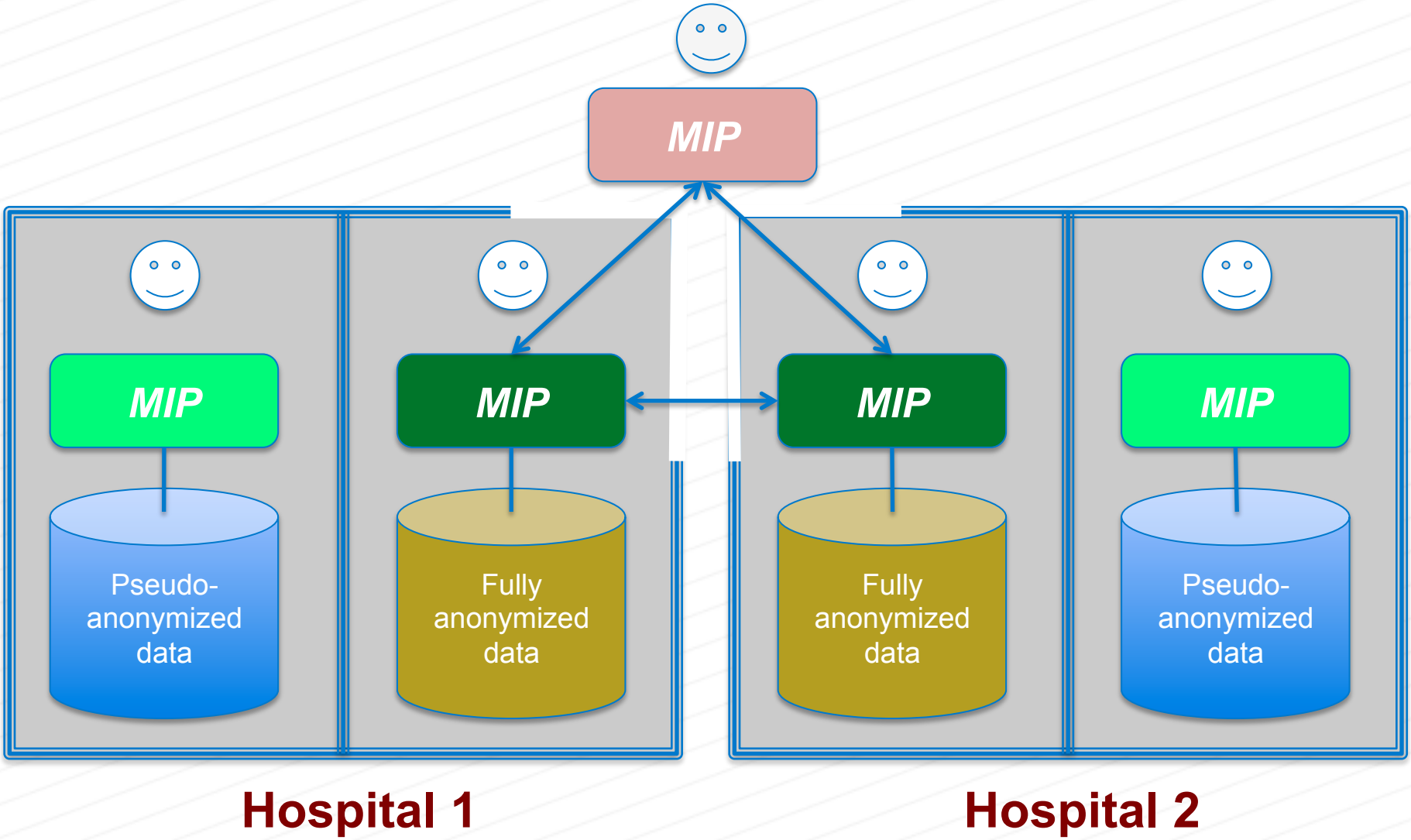
# Human Brain Project

# Medical Informatics Platform (MIP)

# MIP GOALS

- Hospitals sharing multi-modal patient data
- Brain related diseases
  - Dementia, epilepsy, Parkinson's, brain injury, …
- Bridge between brain-science research, clinical research, and patient care

# MIP APPROACH



# MIP DATA

- CHUV CLM (CHUV hospital)
- IRCCS FBF Brescia (Brescia hospital)
- LILLE (Lille Hospital)
- ADNI, EDSD (research datasets)

# MIP ALGORITHMS

- DESCRIPTIVE STATISTICS
- HISTOGRAMS
- PEARSON CORRELATION
- LINEAR REGRESSION
- LOGISTIC REGRESSION
- ANOVA
- ID3
- KMEANS
- NAÏVE BAYES (SIMPLE, WITH CROSS VALIDATION, with HOLD OUT VALIDATION)
- T-TEST (SIMPLE, INDEPENDENT, PAIRED)

# DEPLOYMENT

- Targeting 30 hospital installations by 31 March 2020

- Current status of hospital installations
  - 17 completed, with few providing data
  - 9 under way
  - 6 agreement signed

  - 32 convinced

# DEPLOYMENT

MIP Deployment Status

| MIP INSTALLED | | |
|---|---|---|
| CHUV | Lausanne/Switzerland | |
| Brescia Hospital | Brescia/ Italy | |
| Plovdiv Hospital | Plovdiv/ Bulgaria | |
| CHRU Lille | Lille / France | EPICARE |
| Niguarda Hospital | Milan / Italy | |
| Freiburg Hospital | Freiburg / Germany | EPICARE |
| Institute Mario Negri | Bergamo / Italy | |
| IRCSS Neurological Institute Carlo Besta | Milan / Italy | EPICARE |
| IRCSS Fondazione Istituto Neurologico Nazionale Casimiro Mondino | Pavia / Italy | EPICARE |
| St. Anne's | Brno/CZ | EPICARE |
| Motol university hospital | Prague/CZ | EPICARE |
| Danish epilepsy Filadelfia | Dianalund / Denmark | |
| Hospital del Mar | Barcelona/ Spain | EPICARE |
| Sahlgrenska University Hospital | Gothenburg/ Sweden | EPICARE |
| Grenoble Hospital | Grenoble / France | |
| IRCSS Don Carlo Gnocchi | Milan / Italy | |
| UKAachen | Aachen / Germany | |

| MIP INSTALLATION ONGOING | | |
|---|---|---|
| IRCCS Institute of Neurological Sciences | Bologna / France | EPICARE |
| University hospital Magdebourg | Magdebourg / Germany | |
| CHU de Liège | Liège / Belgium | |
| Charité Hospital | Berlin / Germany | |
| Max Plank Psychiatric Institute | Munich / Germany | |
| LMU | Munich / Germany | |
| Neurospin | Paris / France | |
| TLVMC | Tel Aviv / Israel | |
| IRCSS San Camillo | Venize / Italy | |

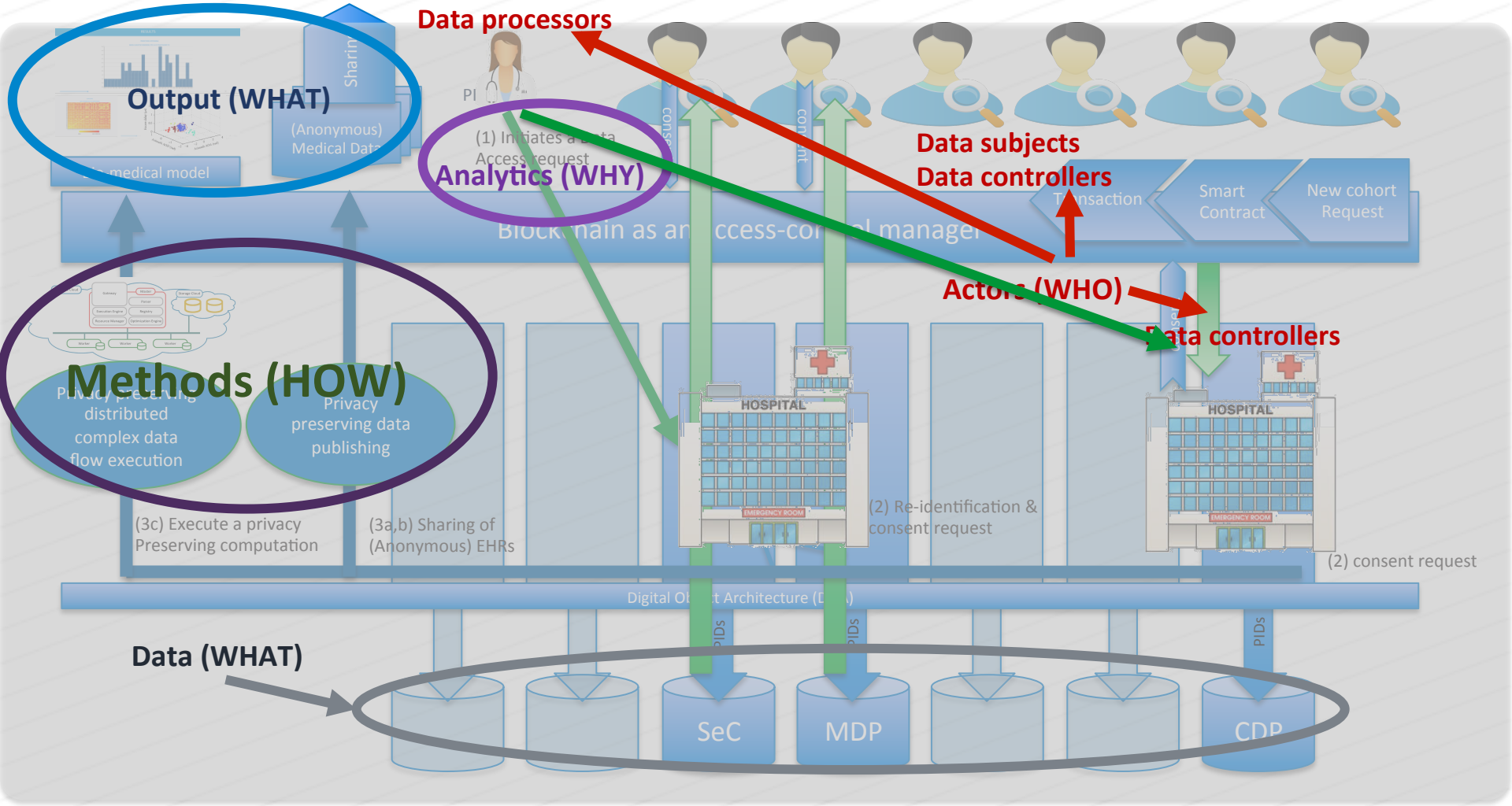
| AGREEMENT SIGNED | | |
|---|---|---|
| Centro Hospitalar do Porto | Porto / Portugal | EPICARE |
| Paracelsus Medical University | Salzbourg / Austria | |
| Rigshospitalet | Copenhagen / Denmark | |
| Karolinska Institute - INCF | Stockholm/ Sweden | |
| Moscow Research and Clinical Center for Neurospychiatry | Moscow / Russia | |
| Hospital Universitario y Politecnico La Fe | Valencia / Spain | EPICARE |

# My Health My Data (MHMD)

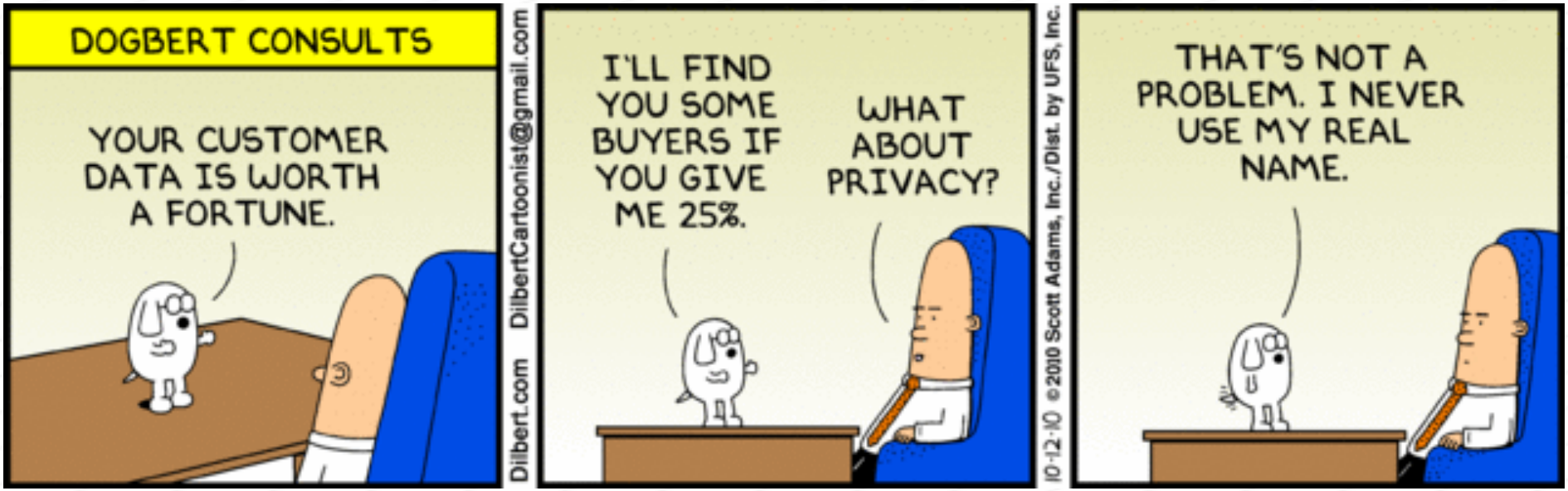
# MHMD GOALS

- Hospitals and individual patients sharing their data and offering services
- Information marketplace
- Advanced mechanisms for trust and direct, value-based relationships between stakeholders (GDPR-compliance)
  - Blockchain for distributed control and inspection of activities
  - Dynamic consent
  - Smart contracts, stored in blockchain, for transaction execution
  - Computational and output privacy techniques

# MHMD SCENARIO



# PRIVACY


DOGBERT CONSULTS
YOUR CUSTOMER DATA IS WORTH A FORTUNE.
Dilbert.com DilbertCartoonist@gmail.com
I'LL FIND YOU SOME BUYERS IF YOU GIVE ME 25%.
WHAT ABOUT PRIVACY?
10-12-10 ©2010 Scott Adams, Inc./Dist. by UFS, Inc.
THAT'S NOT A PROBLEM. I NEVER USE MY REAL NAME.

# PRIVACY CASES

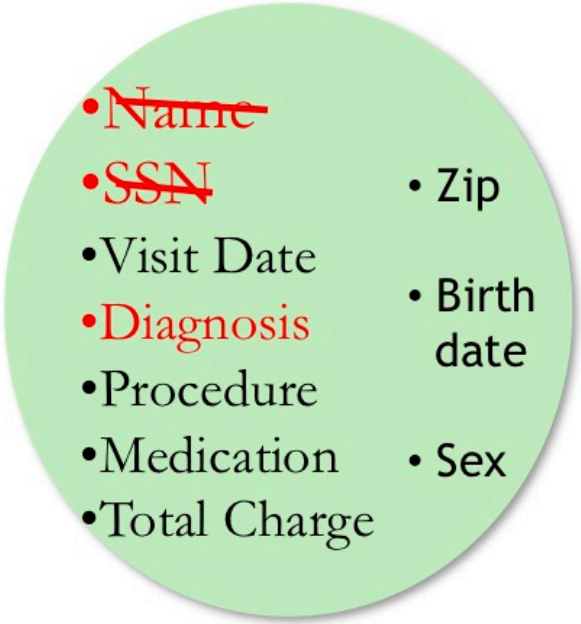
## / Private Data Publishing

- 2nd–level anonymization (over simple pseudonymization)
  → Combinatorial de-identification: k-anonymity, l-diversity (AMNESIA)
- Synthetic data with Differential-Privacy guarantees

## / Private / Secure Data Analytics

- Support interactive, private data analytics scenarios
- Secure Multi-Party Computation (SMPC), Homomorphic Encryption

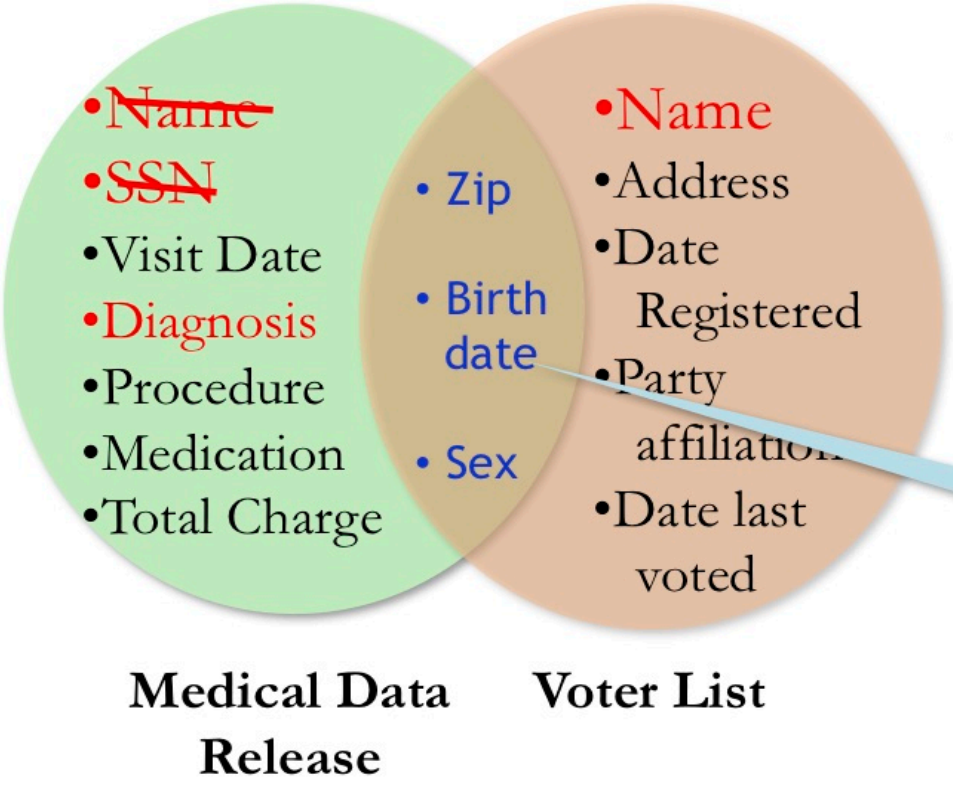
# PII MYTH




**Medical Data Release**

***Massachusetts Governor Breach***

# PII MYTH




- 87 % of US population **uniquely identified** using ZipCode, Birth Date, and Sex.

**Quasi Identifier**

# COMBINATORIAL DATA ANONYMIZATION

- Example mechanism: k-anonymity
  - Each entry becomes indistinguishable from other k-1 entries
  - Achieved through suppression and generalization

| id | Zipcode | Age | National. | Disease |
|---|---|---|---|---|
| 1 | 13053 | 28 | Russian | Heart Disease |
| 2 | 13068 | 29 | American | Heart Disease |
| 3 | 13068 | 21 | Japanese | Viral Infection |
| 4 | 13053 | 23 | American | Viral Infection |
| 5 | 14853 | 50 | Indian | Cancer |
| 6 | 14853 | 55 | Russian | Heart Disease |
| 7 | 14850 | 47 | American | Viral Infection |
| 8 | 14850 | 49 | American | Viral Infection |
| 9 | 13053 | 31 | American | Cancer |
| 10 | 13053 | 37 | Indian | Cancer |
| 11 | 13068 | 36 | Japanese | Cancer |
| 12 | 13068 | 35 | American | Cancer |

| ~~id~~ | Zipcode | Age | National. | Disease |
|---|---|---|---|---|
| ~~1~~ | 130** | <30 | * | Heart Disease |
| ~~2~~ | 130** | <30 | * | Heart Disease |
| ~~3~~ | 130** | <30 | * | Viral Infection |
| ~~4~~ | 130** | <30 | * | Viral Infection |
| ~~5~~ | 1485* | ≥40 | * | Cancer |
| ~~6~~ | 1485* | ≥40 | * | Heart Disease |
| ~~7~~ | 1485* | ≥40 | * | Viral Infection |
| ~~8~~ | 1485* | ≥40 | * | Viral Infection |
| ~~9~~ | 130** | 3* | * | Cancer |
| ~~10~~ | 130** | 3* | * | Cancer |
| ~~11~~ | 130** | 3* | * | Cancer |
| ~~12~~ | 130** | 3* | * | Cancer |

# K-ANONYMITY WEAKNESSES

#Hospital discharges in NJ of ovarian cancer patients, 2009

Counts less than k are suppressed achieving k-anonymity

| Age | #discharges | White | Black | Hispanic | Asian/Pcf Hlnder | Native American | Other | Missing |
|---|---|---|---|---|---|---|---|---|
| #discharges | 735 | 535 | 82 | 58 | 18 | * | 19 | 22 |
| 1-17 | * | * | * | * | * | * | * | * |
| 18-44 | 70 | 40 | 13 | * | * | * | * | * |
| 45-64 | 330 | 236 | 31 | 32 | * | * | 11 | * |
| 65-84 | 298 | 229 | 35 | 13 | * | * | * | * |
| 85+ | 34 | 29 | * | * | * | * | * | * |

# K-ANONYMITY WEAKNESSES

#Hospital discharges in NJ of ovarian cancer patients, 2009

| Age | #discharges | White | Black | Hispanic | Asian/Pcf Hlnder | Native American | Other | Missing |
|---|---|---|---|---|---|---|---|---|
| #discharges | 735 | 535 | 82 | 58 | 18 | **1** | 19 | 22 |
| 1-17 | **3** | **1** | * | * | * | * | * | * |
| 18-44 | 70 | 40 | 13 | * | | | | * |
| 45-64 | 330 | 236 | 31 | 32 | | | 1 | * |
| 65-84 | 298 | 229 | 35 | 13 | * | * | * | * |
| 85+ | 34 | 29 | * | * | * | * | * | * |

$= 535 - (40+236+229+29)$

# K-ANONYMITY WEAKNESSES

#Hospital discharges in NJ of ovarian cancer patients, 2009

***Vulnerable to Inference Attacks!***

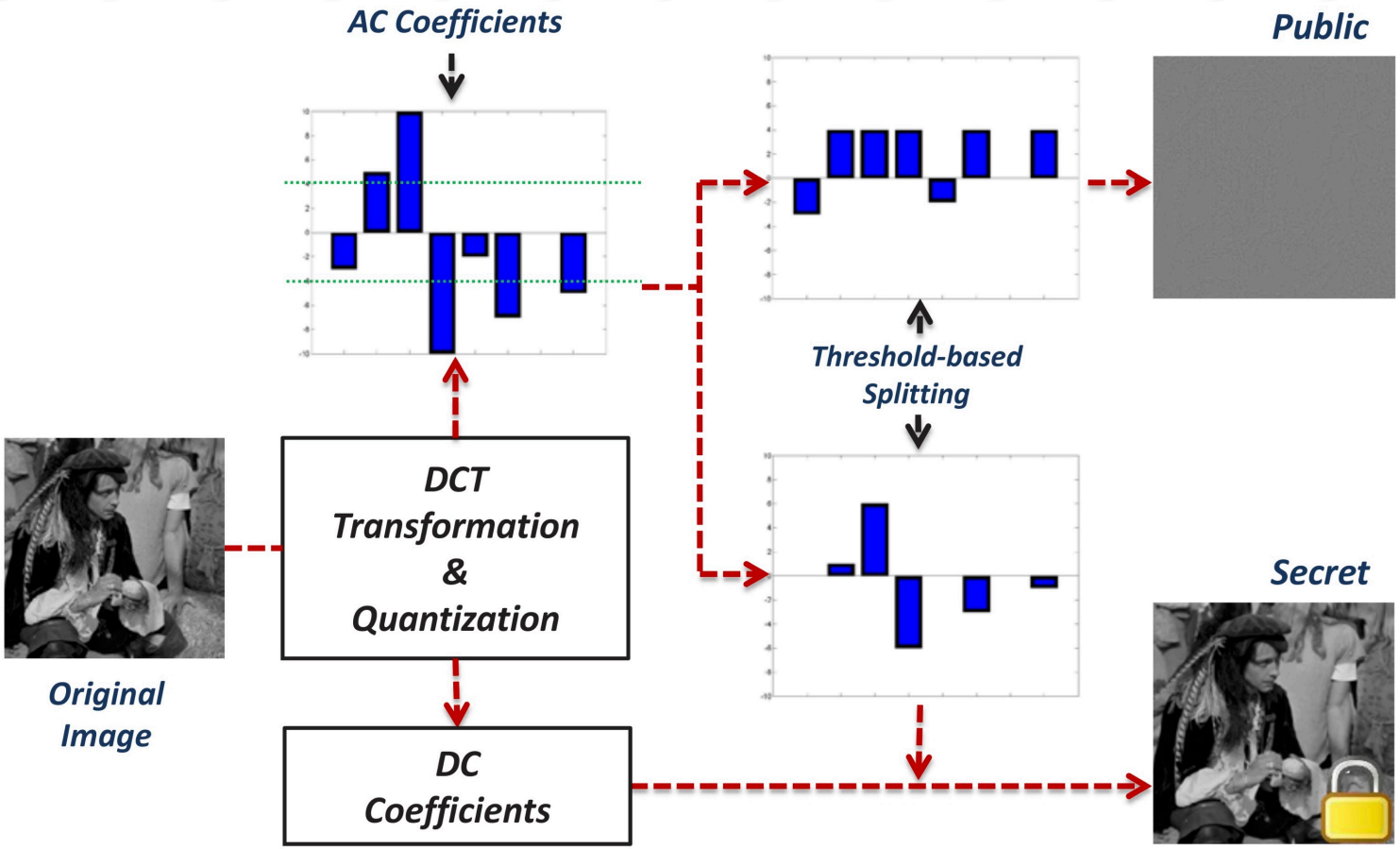
| Age | #discharges | White | Black | Hispanic | Asian/Pcf HInder | Native American | Other | Missing |
|---|---|---|---|---|---|---|---|---|
| #discharges | 735 | 535 | 82 | 58 | 18 | **1** | 19 | 22 |
| 1-17 | **3** | **1** | [0-2] | [0-2] | [0-2] | [0-2] | [0-2] | [0-2] |
| 18-44 | 70 | 40 | 13 | * | * | * | * | * |
| 45-64 | 330 | 236 | 31 | 32 | * | * | 11 | * |
| 65-84 | 298 | 229 | 35 | 13 | * | * | * | * |
| 85+ | 34 | 29 | [1-3] | * | * | * | * | * |

# NO HOPE FOR PRIVATE DATA PUBLISHING

- Published data correlated to original data ➔ Correlations discoverable/exploitable through inference attacks
- Recent work exploits Deep Learning for inference attacks
- Powerful attacks, no smarts required

- Challenge: better, more formal privacy model

# P3 Image Obfuscation



# DIFFERENTIAL PRIVACY

- Property of the algorithm, not the data
- Eliminates potential linkage attacks
- Explicitly mentioned in GDPR, already deployed (Google, Apple, US Census)

- Privacy degrades with multiple composed queries

# DIFFERENTIAL PRIVACY

**Definition (Differential Privacy)**

A randomized algorithm $\mathbf{Alg} : \mathcal{U} \to \mathcal{O}$ is $\varepsilon$-differentially private (DP) if for all $O \subseteq \mathcal{O}$, and for all pairs of adjacent datasets $D, D' \in \mathcal{U}$,

$$\mathbb{P}[\mathbf{Alg}(D) \in O] \leq e^{\varepsilon} \mathbb{P}[\mathbf{Alg}(D') \in O]$$

where the probability space is over the coin flips of **Alg**.

- Distributions of outputs for adjacent inputs are "close".

- Effect of individual records on output (analysis result) is "small".

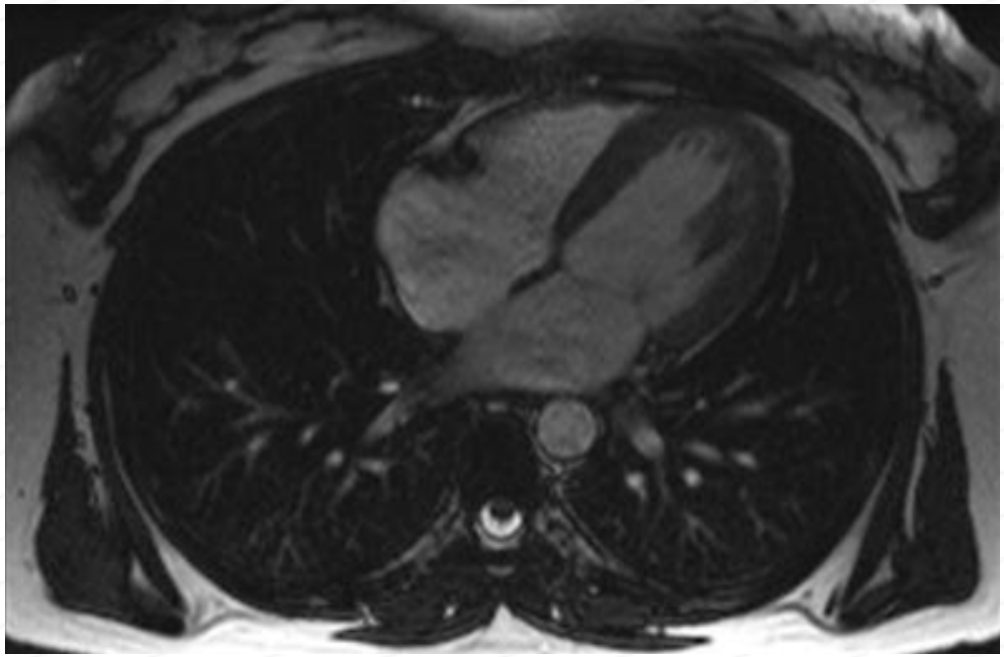
# DP SYNTHETIC DATA

- DP Generative Probabilistic Model
- Sample model points to generate synthetic dataset

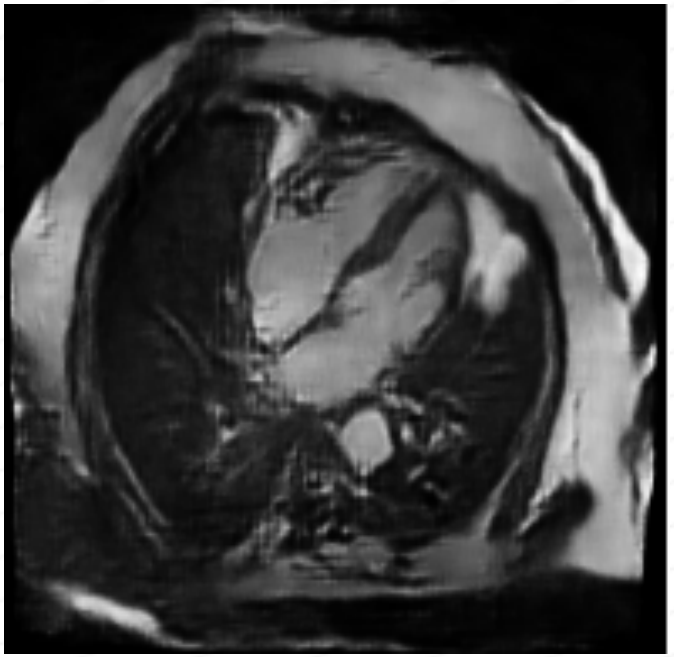
- Challenge: Distributed synthetic dataset
- Need new techniques

# DP SYNTHETIC DATA

- High-quality synthetic cardiovascular images generated by Generative Adversarial Networks (GANs)



A Magnetic Resonance Imaging (MRI) scan of the Heart



A GAN generated synthetic image

# PRIVACY CASES

## / Private Data Publishing

- 2nd–level anonymization (over simple pseudonymization)
  → Combinatorial de-identification: k-anonymity, l-diversity (AMNESIA)
- Synthetic data with Differential-Privacy guarantees

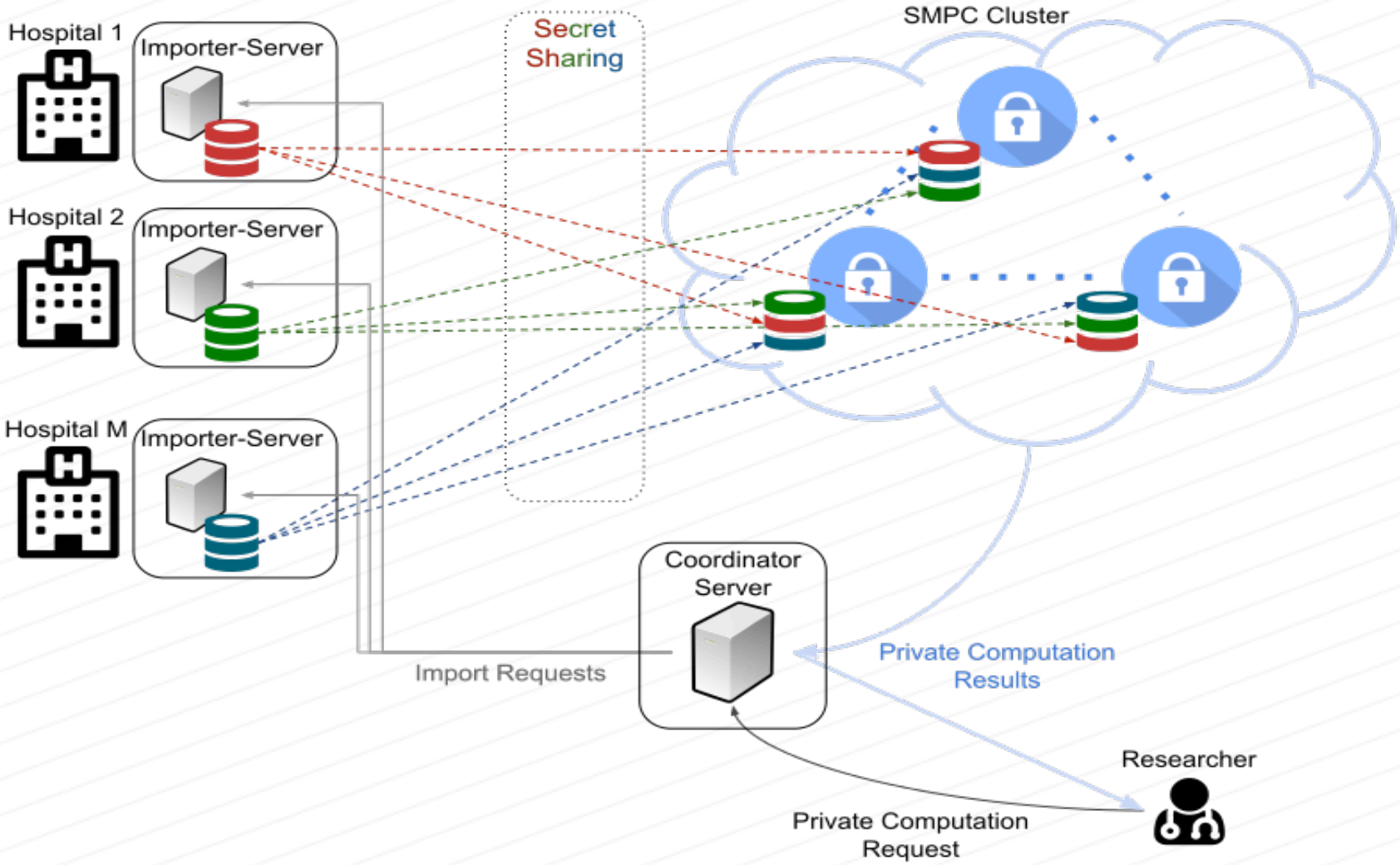
## / Private / Secure Data Analytics

- Support interactive, private data analytics scenarios
- Secure Multi-Party Computation (SMPC), Homomorphic Encryption

# COMPUTATIONAL PRIVACY

- Support private analytics over a distributed data collection
  - All that is learned is the output
- Sensitive data vertically partitioned across federation of (potentially) distrustful data controllers
- Solution:  Secure Multi-Party Computation (SMPC)

# SMPC ARCHITECTURE



# TOWARDS PRACTICAL CRYPTOGRAPHIC PRIVACY

- Practical SMPC no longer an oxymoron
  - Cybernetica, Partisia, Porticor, …
- Differential Privacy in the wild
  - Apple & Google deploy Local DP for data collection
- Numerous open challenges for genomic data

# ACCOUNTABILITY

- Blockchain ideal solution
- Currently active public attacks monitored by blockchain

# INDUSTRIAL OUTLOOK

# Human Brain Project
## Modular Strategic Support

Final Presentation
July 5th, 2019

Engagement Number: 330057622

Prepared for:

Human Brain Project



Gartner®

# DATA DRIVENT P.MEDICINE

01 Executive Summary & Recommendations

02 Market Overview & Trends

03 Market Ecosystem Overview

04 State of the Art MIP model

05 HBP MIP Maturity Assessment

06 Business Model Options

# Executive Summary

1. Limited spending and budgets for advanced technologies
2. Sprawl of analytics and digital platform propositions
3. Ambiguous and selective invest landscape
4. Interim seed funding to achieve MVP required

MVP = minimum viable product




Gartner

# CONCLUSIONS

# KEY CONCEPTS

- Rare Diseases
- Hospital Federations, Medical Data/Service Markets
- Sharing Data and Services
- Hospital and Data Incentives
- Differential Privacy, Math Foundation of Privacy, Synthetic Data
- Secure Multi-Party Computation, Homomorphic Encryption
- Federated Learning
- Blockchain