

Building Bridges Between Technical Data Privacy Capabilities and Legal Requirements

Bradley Malin, Ph.D.

Professor of Biomedical Informatics, Biostatistics, & Computer Science

Co-Director, Health Data Science Center

Vanderbilt University

30 November 2018

EMA Policy 0070

- Push to publicly disseminate clinical trials data
- Phase 1: Publication of clinical study reports only
- Phase 2 (future): Publication of individual-level records




EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

2 October 2014
EMA/240810/2013

European Medicines Agency policy on publication of
clinical data for medicinal products for human use

POLICY/0070
Status: Adopted
Effective date: 1 January 2015
Review date: No later than June 2016
Supersedes: Not applicable



Online access to clinical data for medicinal products for human use

Data on this website

This website contains clinical data published under the European Medicines Agency (EMA) policy on the publication of clinical data. The clinical data have been submitted by pharmaceutical companies to support their marketing applications for human medicines under the centralised procedure and have been assessed by the Committee for Human Medicinal Products (CHMP).

EMA is the first regulatory authority worldwide to provide such broad access to clinical data.

For more information on the clinical data on this website, see [Clinical data available](#).

For more information on EMA and its policy on the publication of clinical data, see the [EMA corporate website](#).

Latest clinical data published

[Ameluz](#) (5-AMINOLEVULINIC ACID)
EMA/H/C/002204/II/0020 published 12 November 2018

[Truxima](#) (RITUXIMAB)
EMA/H/C/004112/0000 published 9 November 2018

[Alecensa](#) (ALECTINIB)
EMA/H/C/004164/0000 published 17 October 2018

[Lucentis](#) (RANIBIZUMAB)
EMA/H/C/000715/II/0061 published 9 October 2018

[Movymia](#) (TERIPARATIDE)
EMA/H/C/004368/0000 published 2 October 2018

[Terrosa](#) (TERIPARATIDE)

Log in with an EMA account

EMA account holders should log in with their login credentials.

This field is required.

[Forgot username](#)

This field is required.

[Forgot password](#)[Not sure if you have an EMA account?](#)☐ Remember me

No EMA account?

New users need to create an EMA account to access clinical data on this website. Once you have created an EMA account, please return to this page to log in.

[Terms of use](#)

Clinical reports for Ameluz - Extension of indication

All information published on the website is correct at the time of publication. For the current status of this product, please see 'Find medicine' on the EMA website.

Enter a search term to identify the documents containing this term.

Search

Any text or keyword search looks for matches in both the document title and the document content.

Upgrade your access

The protocol and protocol amendments, sample case report form and documentation of statistical methods are incorporated in the documents under 'Clinical Study Report'.

Expand all

Clinical overview

The clinical overview provides a critical analysis of the clinical data in the eCTD.

Clinical summary

The clinical summary provides a detailed factual summary of the clinical information in the eCTD.

Clinical study reports

A clinical study report (CSR) on a clinical trial is a detailed document about the methods and results of a trial.

Anonymisation report

The anonymisation report describes the anonymisation process followed by the Applicant.

Product name

Ameluz

MAH

Biofrontera Bioscience GmbH

Active substance

5-AMINOLEVULINIC ACID

ATC code

L01XD04

Number of Documents

25

Procedure type

Extension of indication

Publication year

2018

Product Status

Authorised

Type

Article 58

No

EMA procedure number

EMA/H/C/002204/II/0020

See the European Public Assessment Report (EPAR) on the EMA website

methods are incorporated in the documents under 'Clinical Study Report'.

Collapse all

Clinical overview



The clinical overview provides a critical analysis of the clinical data in the eCTD.

Clinical summary



The clinical summary provides a detailed factual summary of the clinical information in the eCTD.

Clinical study reports



A clinical study report (CSR) of a clinical trial is a detailed document about the methods and results of a trial.

m5332-ala-ak-ct006-s-csr-with-a

m5351-ala-ak-ct002-s-csr-with-a

m5351-ala-ak-ct002-s-csr-with-a

m5351-ala-ak-ct002-s-csr-with-a

m5351-ala-ak-ct003-s-csr-with-a

m5351-ala-ak-ct003-s-csr-with-a

m5351-ala-ak-ct003-s-csr-with-a

m5351-ala-ak-ct007-p-csr-with-a

m5351-ala-ak-ct007-s-csr-with-a

m5353-ise-sap.pdf

m5353-ise.pdf

m5353-iss-sap.pdf

m5353-iss.pdf

m5354-ala-ak-ct005-s-app1611-protocol.pdf

m5354-ala-ak-ct005-s-app1612-crf.pdf

m5354-ala-ak-ct005-s-app1619-sap.pdf

m5354-ala-ak-ct005-s-csr-body-1-13.pdf

m5354-ala-ak-ct005-s-csr-body-14.pdf

m5354-ala-ak-ct005-s-csr-body-15.pdf

Anonymisation report



The anonymisation report describes the anonymisation process followed by the Applicant.

Number of Documents

25

Procedure type

Extension of indication

Publication year

2018

Product Status

Authorised

Type

Article 58

No

er

04/II/0020

Public Assessment

EMA website [↗](#)

Accept



In order to view the document please accept the [Terms of Use](#) by ticking the check box.

☐

OK

Clinical data available

Annual report - policy
implementation

Data protection notice

Terms of Use

Contact Us

Terms of Use

The terms of use have to be accepted in order to gain access to clinical data. They provide detailed information on the access to the information in the clinical reports, and on the intended use of such information.

There are two terms of use, in line with the category of access to the published clinical data.

The **general information purposes** terms of use are for users accessing the published clinical data for general information and other non-commercial purposes, including non-commercial research purposes.

The **academic and other non-commercial research purposes** terms of use are for users accessing the published clinical data for academic and other non-commercial research purposes. They allow more access rights compared to the "general information purposes" terms of use, but require some additional procedural steps.

Prior to viewing or downloading every document users must re-confirm their acceptance of the terms of use.

Terms of Use for general information purposes

These Terms of Use ("**Terms**") govern the access and use of clinical data, as defined in chapter 3. of the EMA policy on publication of clinical data, Policy 0070 ("**Policy**"), that are made available to Users via such Policy. By accepting these Terms and upon being granted access to the Clinical Reports, you agree to be bound by these Terms. Please read them carefully.

1. Definitions

In these Terms the terms below have the following meaning:

"**EMA**" means the European Medicines Agency.

"**Clinical Reports**" means the clinical overviews (module 2.5), the clinical summaries (module 2.7) and the clinical study reports (module 5, "CSR"), together with appendixes to the CSRs no. 16.1.1, 16.1.2 and 16.1.9 which are accessible via the EMA website as a result of the implementation of the Policy.

"**Applicant/MAH**" means the natural or legal person(s) or organisation(s) that submitted the Clinical

3. Use of the Clinical Reports

The User agrees to use the Clinical Reports according to these Terms and, in particular, that:

- a. The User may use the Clinical Reports for general information and other non-commercial purposes, including non-commercial research purposes, subject to these Terms.
- b. The User is not granted any intellectual property or other commercial rights in relation to the Clinical Reports other than as expressly set out in these Terms.

When using the Clinical Reports, the User shall:

- a. acknowledge that its source is the Applicant/MAH;
- b. not use it in a way that suggests that the Applicant/MAH endorses the User's use of the Clinical Reports for any other purpose than general information and other non-commercial purposes, including non-commercial research purposes;
- c. ensure that the use of the Clinical Reports comply at all times with applicable law;

- e. not seek to re-identify the trial subjects or other individuals from the Clinical Reports in breach of applicable privacy laws.

The User may not:

- use the Clinical Reports to support an application to obtain a marketing authorisation and any extensions or variations thereof for a product anywhere in the world;
- share the User's username, password or other account details with a third party or otherwise provide a third party with access to the User's account;
- make any unfair commercial use of the Clinical Reports.

If the User fails to accurately complete the registration process, comply with these conditions, or uses the Clinical Reports in breach of these Terms, the rights to access and use the Clinical Reports will be revoked.

4. Warranties and liability

EMA Policy 0070: Drawing a Line

- Push to publicly disseminate clinical trials data
- April 2017 – guidance on how to share data
- Risk of re-identification: **no worse than 0.09**



https://www.ema.europa.eu/documents/regulatory-procedural-guideline/external-guidance-implementation-european-medicines-agency-policy-publication-clinical-data_en.pdf



The future of health begins with you

The *All of Us* Research Program is a historic effort to gather data from one million or more people living in the United States to accelerate research and improve health. By taking into account individual differences in lifestyle, environment, and biology, researchers will uncover paths toward delivering precision medicine.

JOIN NOW

Interested in the *All of Us* Research Program?

LEARN MORE

Sign up to be notified of announcements, events, funding news and more.

SUBSCRIBE

We are building a research program of 1,000,000+ people



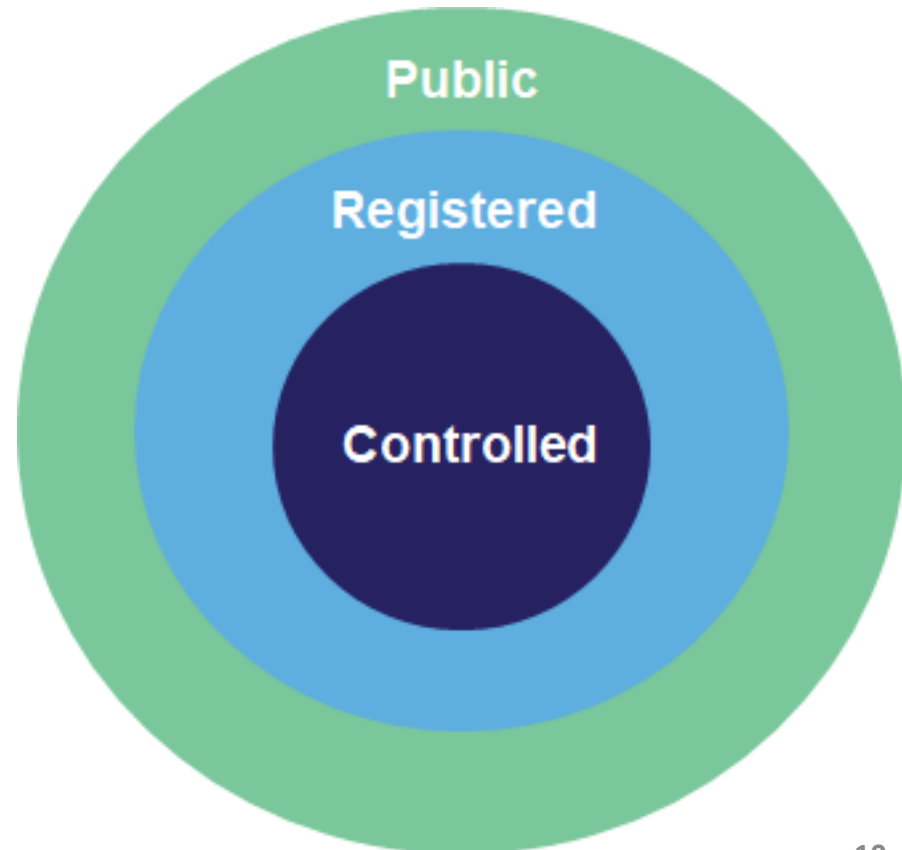
environment



lifestyle

Tiered Levels of Access

- Public
 - Can be accessed without logging in
 - Summary statistics only
- Sandbox Environments (on Google Cloud)
 - Registered
 - Individual level records with minimal risk to participant identification
 - Controlled
 - Individual level records with more risk to participant identification, but expected to be low



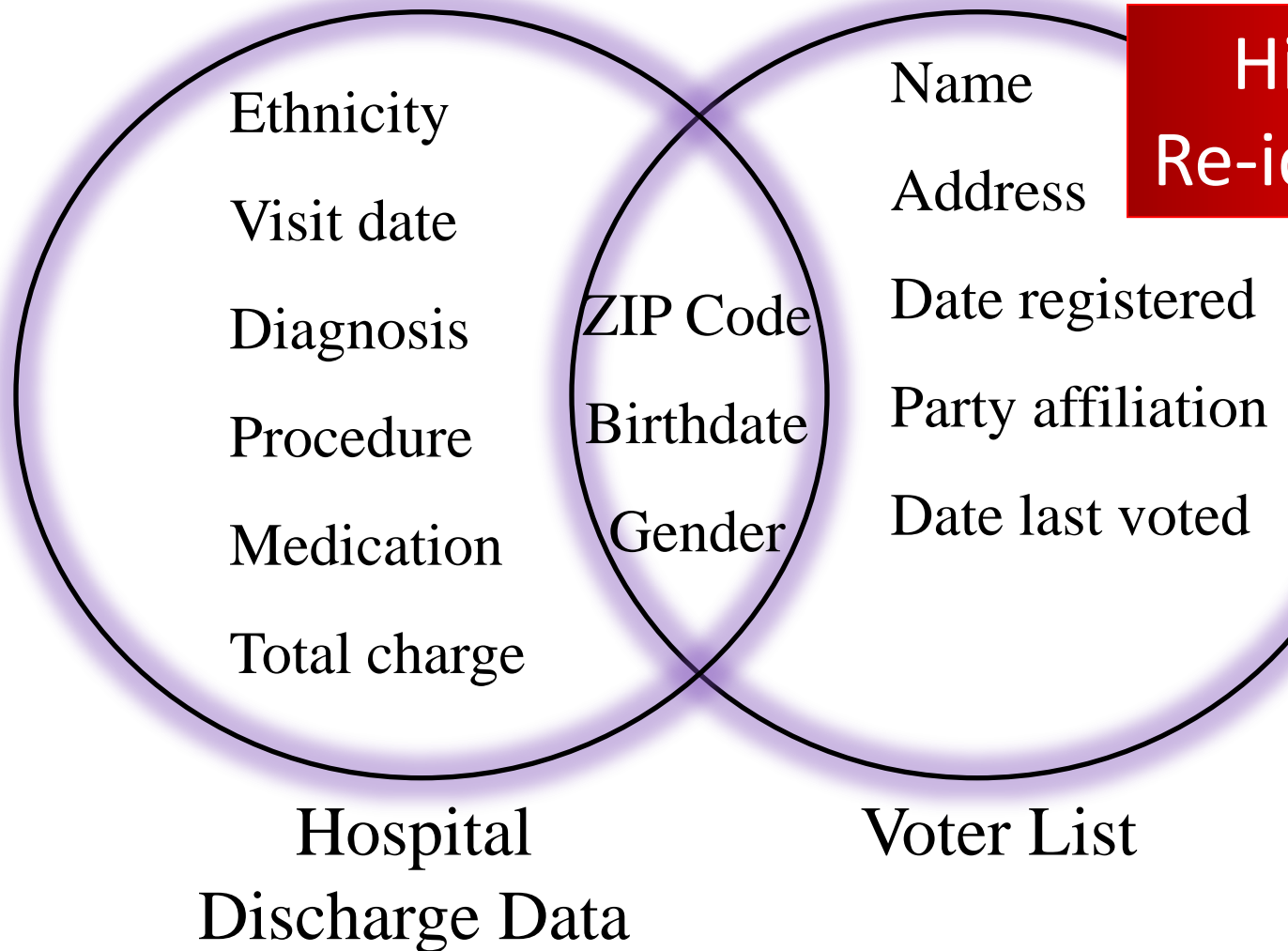


THE FACE
THAT
LAUNCHED



A THOUSAND
SHIPS

A “Quasi-identifier” Conundrum



**High Profile
Re-identification**



5-Digit US ZIP Code

+ Birthdate

+ Gender

63-87% of USA
estimated to be unique

Set the World
on



Harvard Professor Re-Identifies Anonymous Volunteers In DNA Study



Adam Tanner Contributor ⓘ

Apr 25, 2013, 03:47pm • 22,581 views

f

t

in

A Harvard professor has re-identified the names of more than 40% of a sample of anonymous participants in a high-profile DNA study, highlighting the dangers that ever greater amounts of personal data available in the Internet era could unravel personal secrets.

From the onset, the [Personal Genome Project](#), set up by Harvard Medical School Professor of Genetics George Church, has warned participants of the risk that someone someday could identify them, meaning anyone could look up the intimate medical histories that many have posted along with their genome data. That day arrived on Thursday.



Harvard Professor Latanya Sweeney

MUST READ **WINDOWS 10: THESE TWO NEW BUILDS DELIVER LOTS OF BUG FIXES**

Re-identification possible with Australian de-identified Medicare and PBS open data

Using publicly known information, a team of researchers from the University of Melbourne have claimed to re-identify seven prominent Australians in an open medical dataset.



By [Chris Duckett](#) | December 18, 2017 -- 01:01 GMT (17:01 PST) | Topic: [Security](#)

1

f 3

in



The dataset containing historic longitudinal medical billing records of one-tenth of all Australians, approximately 2.9 million people, has been found to be re-identifiable by a team from the University of Melbourne, with information such as child births and professional sportspeople undergoing surgery to fix injuries often made public.

The team, consisting of Dr Chris Culnane, Dr Benjamin Rubinstein, and Dr Vanessa Teague, warned that they expect similar results with other data held by the government, such as Census data, tax records, mental health records, penal data, and Centrelink data.

"We found that patients can be re-identified, without decryption, through a process of linking the unencrypted parts of the record with known information about the individual such as medical procedures and year of birth," Dr Culnane said.

"This shows the surprising ease with which de-identification can fail, highlighting the risky

[Your Favorite Feature] Distinguishes You!!

- Demographics (Sweeney '97; Bacher '02; Golle '06; El Emam '08; Koot '10; Li '11, Sweeney '13)
- Diagnosis Codes (Loukides '10; Tamersoy '10, '12; Heatherly '16)
- Laboratory Tests (Cimino '12; Atreya '13)
- DNA (Malin '00, Lin '04; Homer '08; Gymrek '13, Ayday '14, Huttenhower '15; Shringapure '15; Lippert '17, Erlich '18)
- RNA (Backes '16a; Backes '16b)
- Proteome (Li '16)
- Health Survey Responses (Solomon '12)
- Pedigree Structure (Malin '06, Ayday '13)

- Location Visits and Mobility Traces (Malin '04; Golle '09; El Emam '11; de Montjoye '15; Kondor '17; Murakami '17)
- Movie Reviews (Narayanan '08)
- Social Network Structure (Backstrom '07; Narayanan '09; Yang '12; Cecaj '14, '16)
- Search Queries (Barbaro '06)
- Internet Browsing (Malin '05; Eckersley '10; Banse '11; Herrmann '12, Olejnik '12; Kirchler '16; Riederer '16)
- Smart Utility Meter Usage (Buchmann '12; Faisal '15; Tudor '15)

Given Enough
Time,
Effort,
Incentive, &
Money...

Possible

doesn't imply

Probable

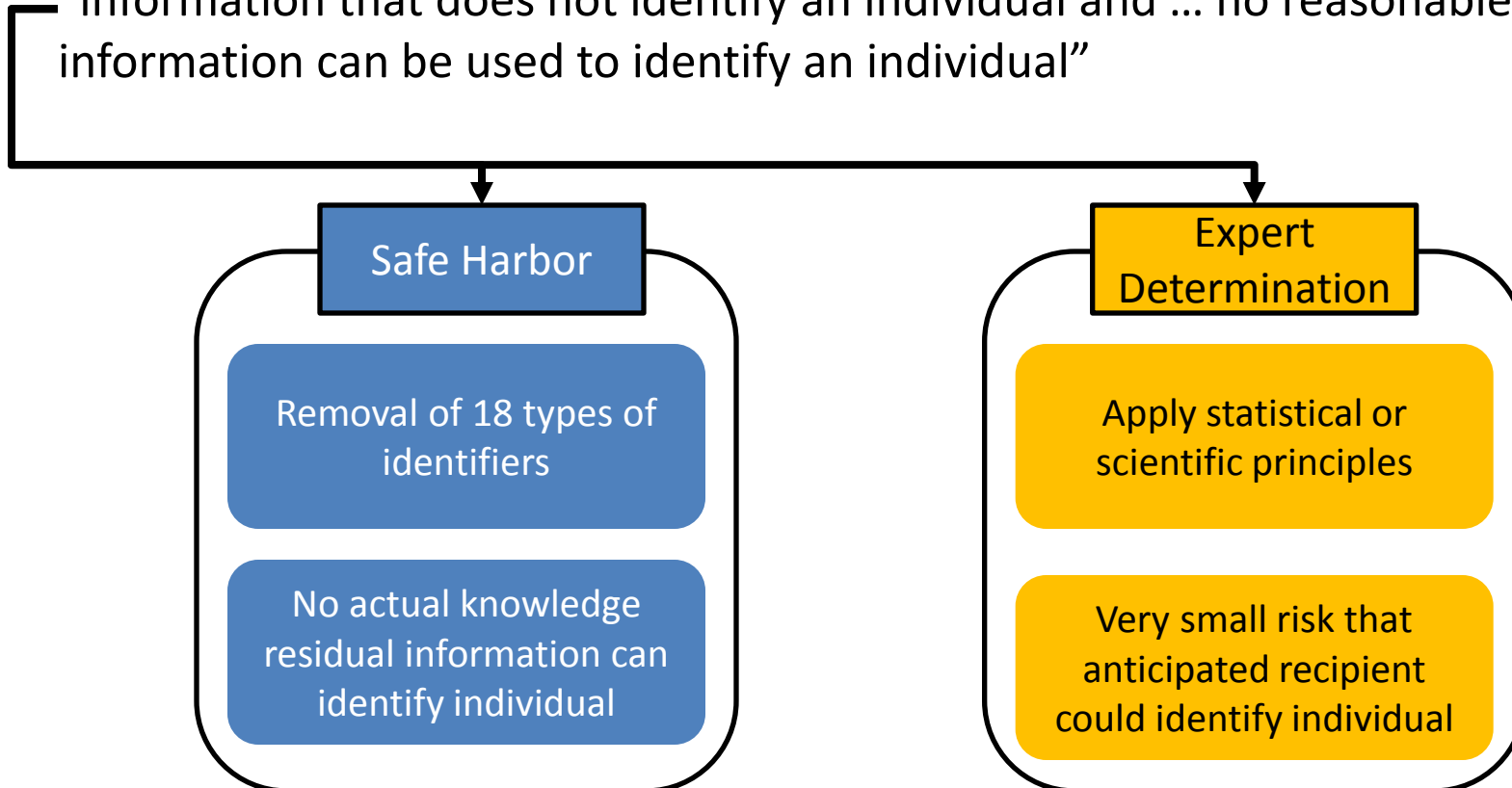
What is “Anonymisation”?

According to EU (Data Protection Directive → GDPR):

“principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable”

According to the US Health Insurance Portability & Accountability Act (HIPAA):

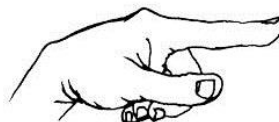
“information that does not identify an individual and ... no reasonable basis ... information can be used to identify an individual”



US National Institutes of Health Data Sharing Policies

- 2003 Final Data Sharing Policy:
 - Studies with > \$500k/yr → Investigators must have data sharing plan or explain why it's not possible
 - Recommends sharing data devoid of identifiers
- 2007 GWAS Policy and 2014 Genome Data Sharing Policy
 - Studies involving > \$0
 - Recent considerations for extending this to all sequencing data
- Identifiable?

NIH



HIPAA

HIPAA Privacy Rule Allows Secondary Uses of Data

Identified Patient Data

- Waiver of consent: data is “on the shelf”
- Consent is impractical to obtain

Limited Data Set

- Removal of 16 designated attributes
- Recipient signs data use contract

“De-identified” Data

- Option 1: Safe Harbor
- Option 2: Expert Determination

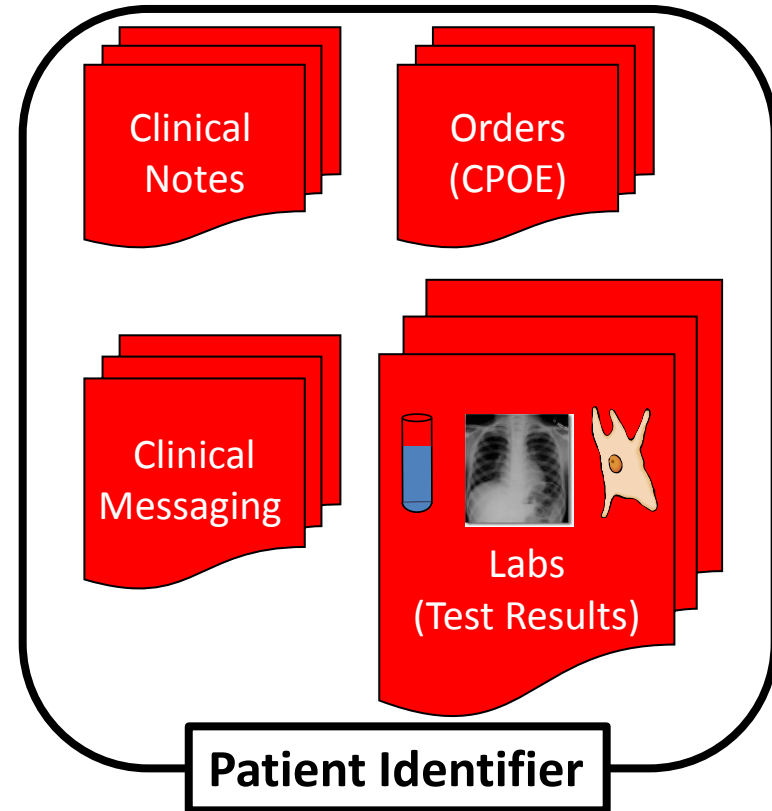
A Recipe for De-identification

Field	Detail
Names	Related to patient (not provider)
Unique Numbers	Phone, SSN, MRN, ...
Internet	Email, URL, IP addresses, ..
Biometrics	Finger, voice, ...
Dates	Less specific than year Ages > 89
Geocodes	Town, County, Less specific than Zip-3 (assuming > 20,000 people in zone)
"Catch all"	"Any other unique identifying number, characteristic, or code"

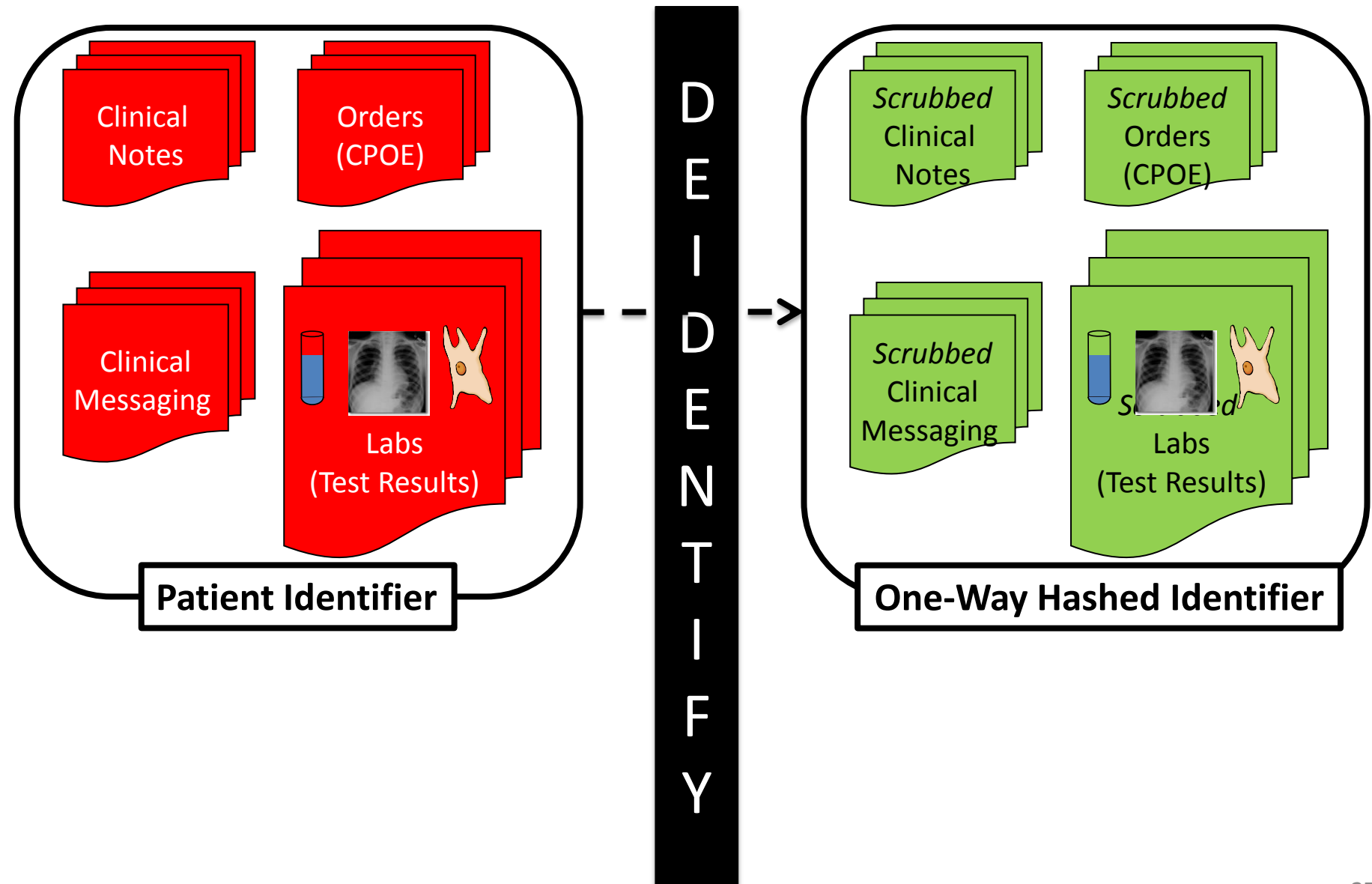
HIPAA
Limited Dataset

HIPAA
Safe Harbor

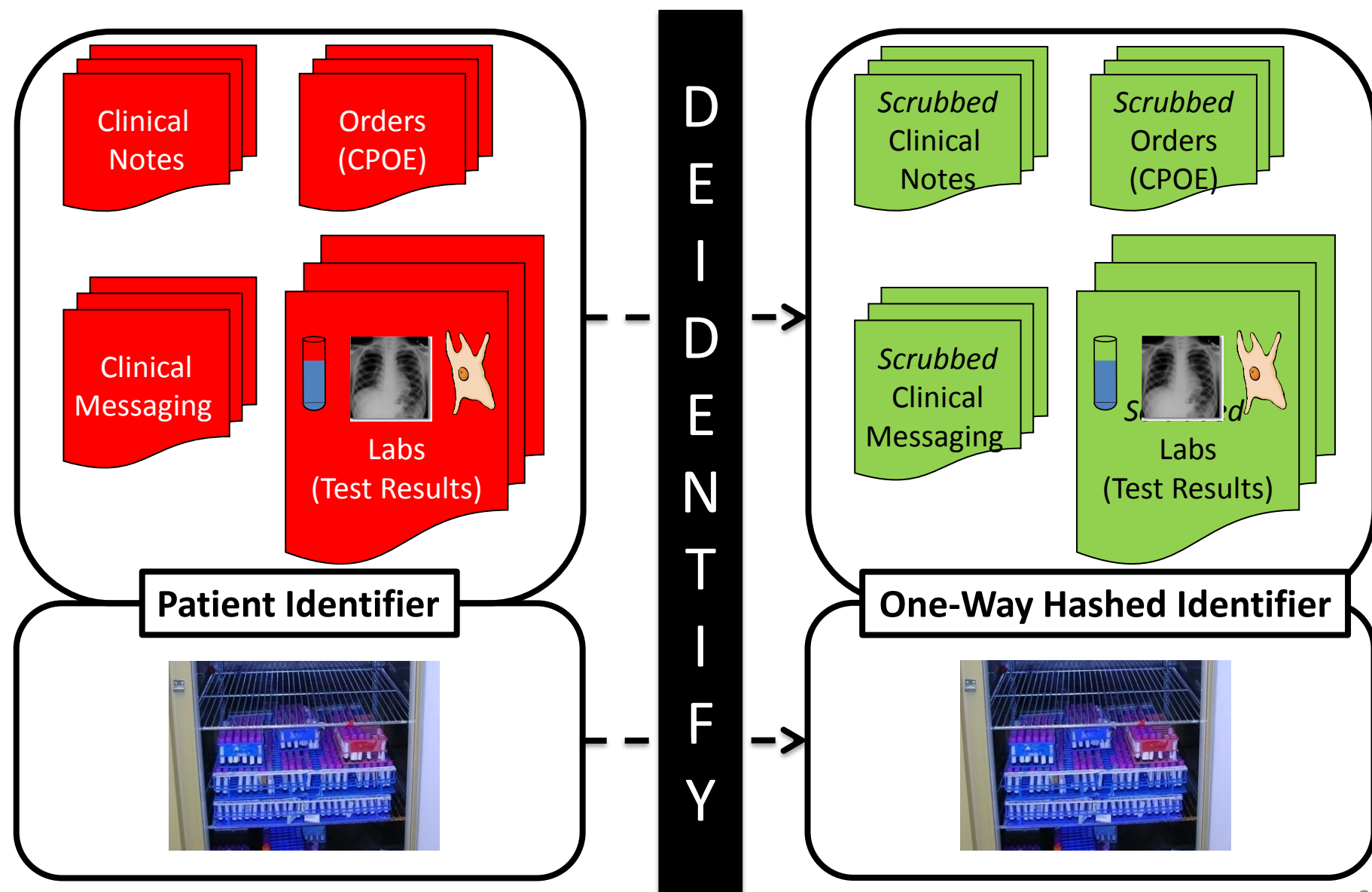
Vanderbilt's Research Derivative



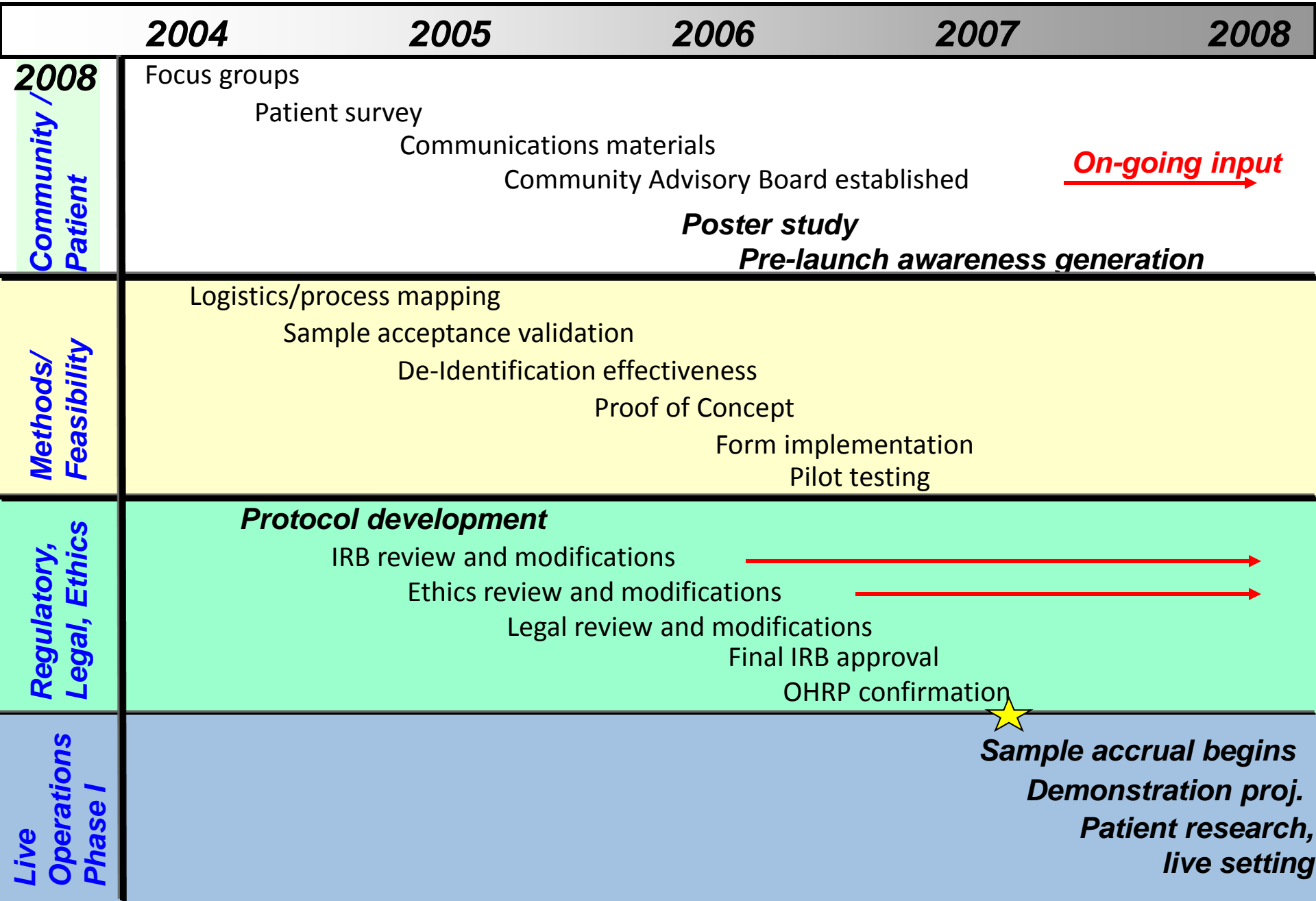
Vanderbilt's Synthetic Derivative



Vanderbilt's BioVU



Vanderbilt De-identified EMR + DNA



A Two Tier Access Model

① Initial Hypothesis Generation

Juvenile Diabetes = Positive
Rx = anti-psychotic
Adverse event < 3 months of Rx

Aggregate Results

Interface

Honest Broker

Scrubbed
Clinical
Notes

Scrubbed
Orders
(CPOE)

Scrubbed
Clinical
Messaging

Scrubbed
Labs
(Test Results)

One-Way Hashed Identifier



A Two Tier Access Model

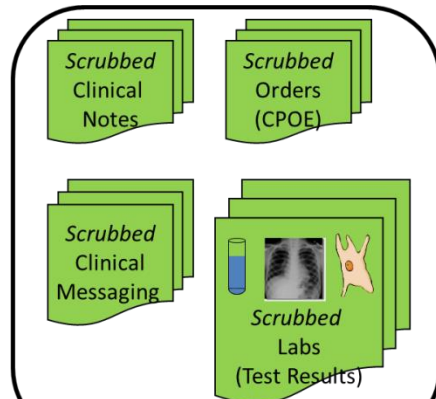
②

Hypothesis Testing

Juvenile Diabetes = Positive
Rx = anti-psychotic
Adverse event < 3 months of Rx

Approved IRB
Protocol

Data Use
Agreement

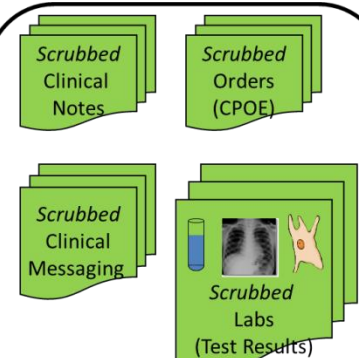


Investigator-Specific IDs



Raw Data
Download
&
Specimen
Collection

Honest Broker



One-Way Hashed Identifier



Redaction in Natural Language

Original PHI

Smith, 61 yo ...
daughter, Lynn, to ...
oncologist Dr. White ...
5/13/10 to consider ...
SWOG protocol 1811, ...
was randomized 5/10 ...
to call Mr. Smith on ...
PLAN:Dr White and I ...

Redaction in Natural Language

Original PHI

Redacted PHI

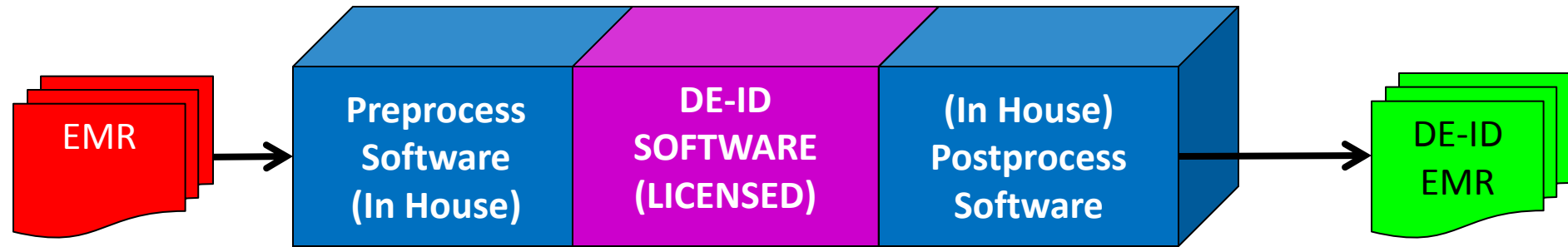
<u>Smith</u> , <u>61</u> yo ... daughter, <u>Lynn</u> , to ... oncologist <u>Dr. White</u> ... <u>5/13/10</u> to consider ... SWOG protocol <u>1811</u> , ... was randomized <u>5/10</u> ... to call <u>Mr. Smith</u> on ... PLAN: <u>Dr White</u> and I ...	<u>**pt_name<A></u> , <u>**age<60s></u> yo ... daughter, <u>**pt_name</u> , to ... oncologist Dr. <u>**MD_name<C></u> ... <u>**date<5/28/10></u> to consider ... SWOG protocol <u>**other_id</u> , ... was randomized <u>5/10</u> ... to call Mr. <u>**pt_name<A></u> on ... PLAN: <u>**MD_name<C></u> and I ...
---	--

Redaction in Natural Language

Rules

- Dictionaries
- Regular Expressions
- Lexicons
- Exclusions
- Note-specific rules
- Site-specific rules

Vanderbilt Scrubbing Process (Simplified)



Redact all Safe Harbor features ... shift dates random offset in $[-365, 0)$

Recall = 0.999

Redaction in Natural Language

Rules

- Dictionaries
- Regular Expressions
- Lexicons
- Exclusions
- Note-specific rules
- Site-specific rules

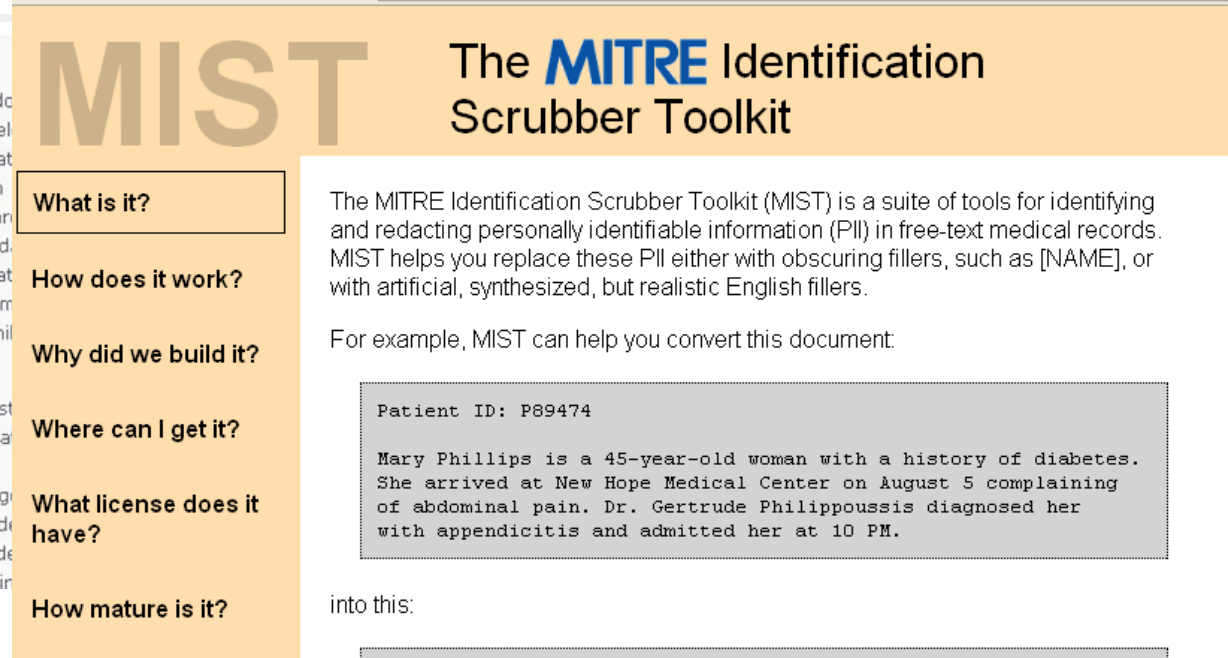
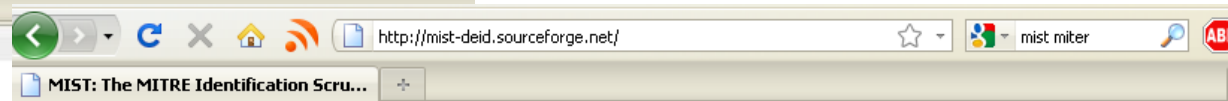
Machine Learned

- Naïve Bayes
- Decision Trees / Stumps
- Support Vector Machines
- ...
- Conditional Random Fields

Software: From Theory to Practice

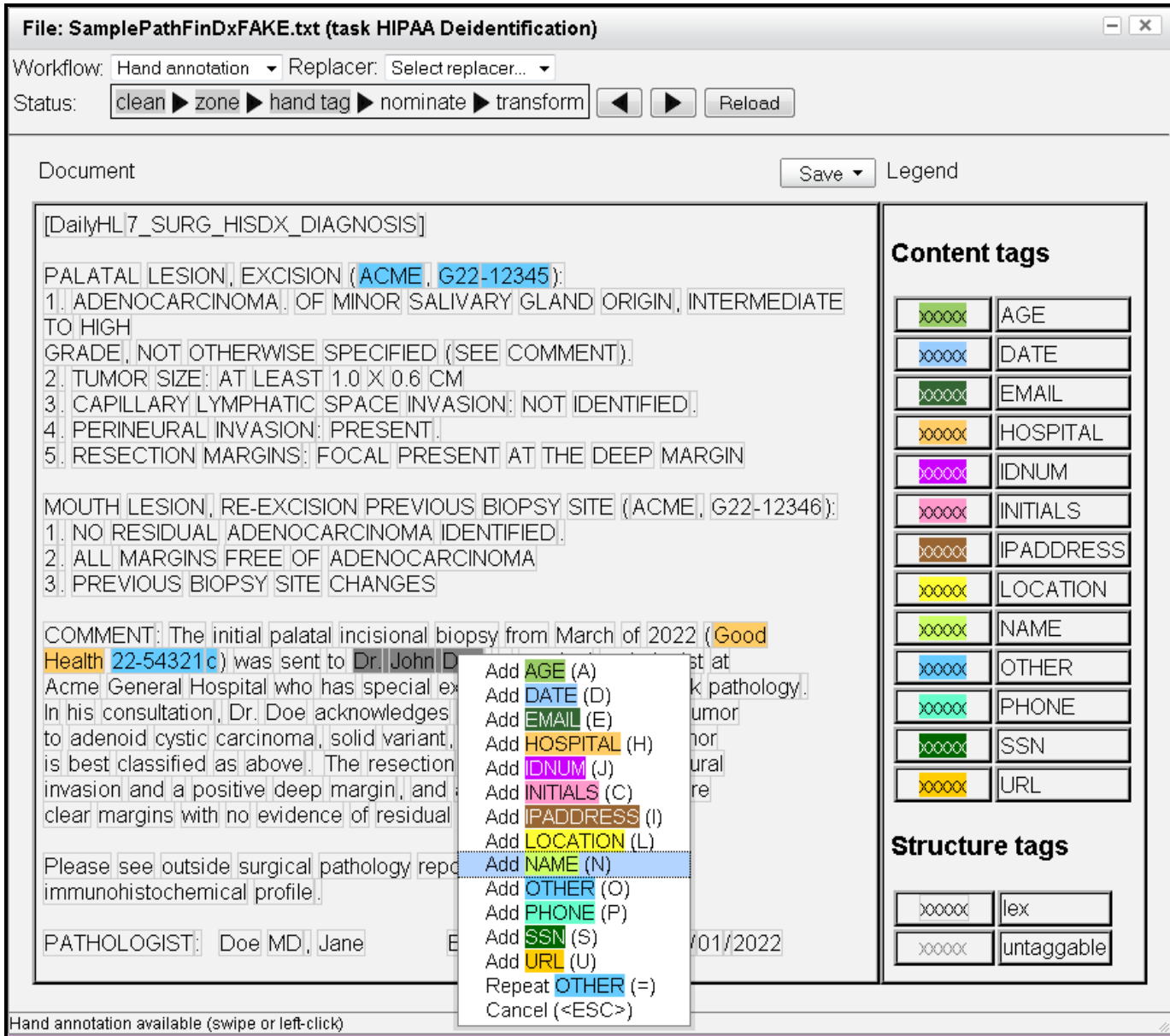
HIDE (Gardner & Xiong 2009)

MIST (Aberdeen et al 2010)



into this:

MIST Installation & Training



Does MIST Work? (Vanderbilt EMR)

	Discharge	Laboratory	Letter	Order	All
Train	200	400	200	400	1200
Test	50	100	50	100	300
Precision	0.946	0.905	0.931	0.993	0.943
Recall	0.986	0.966	0.956	0.999	0.978

Precision: 0.91 – 0.99

Recall: 0.95 – 0.99

Redaction Has its Limits

Original PHI

****Redacted PHI & Leaked PHI**

<u>Smith</u> , <u>61</u> yo ... daughter, <u>Lynn</u> , to ... oncologist <u>Dr. White</u> ... <u>5/13/10</u> to consider ... SWOG protocol <u>1811</u> , ... was randomized <u>5/10</u> ... to call <u>Mr. Smith</u> on ... PLAN: <u>Dr White</u> and I ...	**pt_name<A> , **age<60s> yo ... daughter, <u>Lynn</u> , to ... oncologist Dr. **MD_name<C> ... **date<5/28/10> to consider ... SWOG protocol **other_id , ... was randomized <u>5/10</u> ... to call Mr. **pt_name<A> on ... PLAN: <u>Dr White</u> and I ...
---	--

Redaction Has its Limits

Original PHI

****Redacted PHI & Leaked PHI**

Unknown residual re-identification potential (e.g. “the Senator’s wife”)

was randomized 8/18 ... to call <u>Mr. Smith</u> on ... PLAN: <u>Dr White</u> and I ...	was randomized 8/18 ... to call Mr. **pt_name<A> on ... PLAN: <u>Dr White</u> and I ...
---	--

Policy:
Data Use Agreements

Redaction Has its Limits...

but it Isn't the Only Option

Original PHI

Smith, 61 yo ...
daughter, Lynn, to ...
oncologist Dr. White ...
5/13/10 to consider ...
SWOG protocol 1811, ...
was randomized 5/10 ...
to call Mr. Smith on ...
PLAN:Dr White and I ...

****Redacted PHI & Leaked PHI**

****pt_name<A>**, ****age<60s>** yo ...
daughter, Lynn, to ...
oncologist Dr. ****MD_name<C>** ...
****date<5/28/10>** to consider ...
SWOG protocol ****other_id**, ...
was randomized 5/10 ...
to call Mr. ****pt_name<A>** on ...
PLAN:Dr White and I ...

Surrogate PHI & Hidden PHI

Jones, a 64 yo ...
daughter, Lynn, for ...
oncologist Dr. Howe ...
5/28/10 to consider ...
SWOG protocol 1798, ...
was randomized 5/10 ...
to call Mr. Jones on ...
PLAN:Dr White and I ...

Idea: Inject surrogated information to hide the leaks!

Hiding in Plain Sight [HIPS]

- Added a surrogation component to MIST*
- ~130 oncology notes from Group Health Coop of Puget Sound

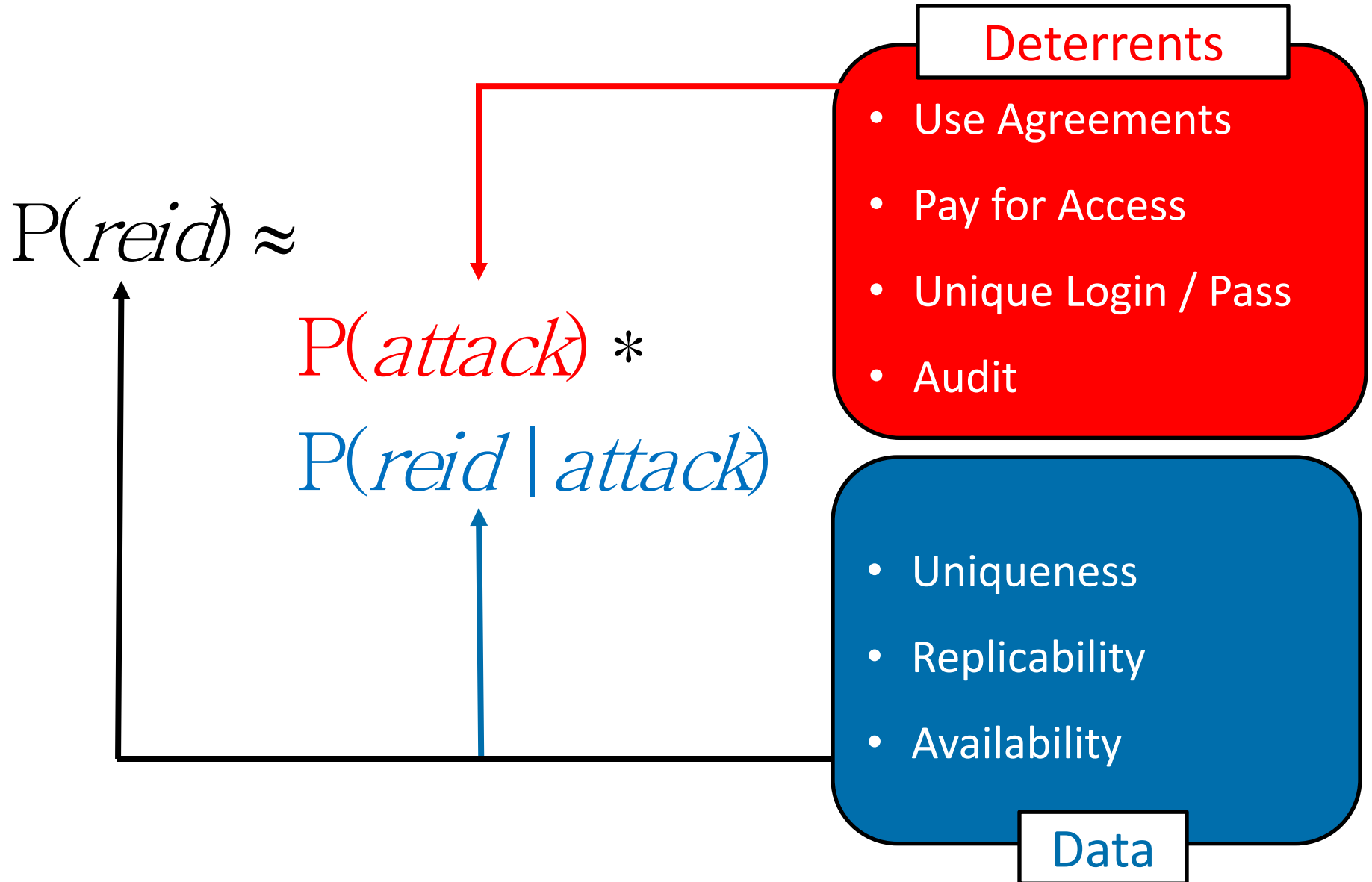
**MIST forced into a dumbed-down state for assessment*

	Can effectively raise de-identification performance from 0.15 to 0.99			Extractor)
Identifier type				
HIPAA				
Pat. name				
Age				
Phone #				
Address				
Date	100	17	0.85	
MRN	3	3	1.00	
Acct. #	1	1	1.00	
Other ID #s	10	9	0.90	
ALL	323	47	0.15	
OTHER				
Prac name	82	9	0.11	
Org. name	27	20	0.74	
ALL	109	29	0.27	

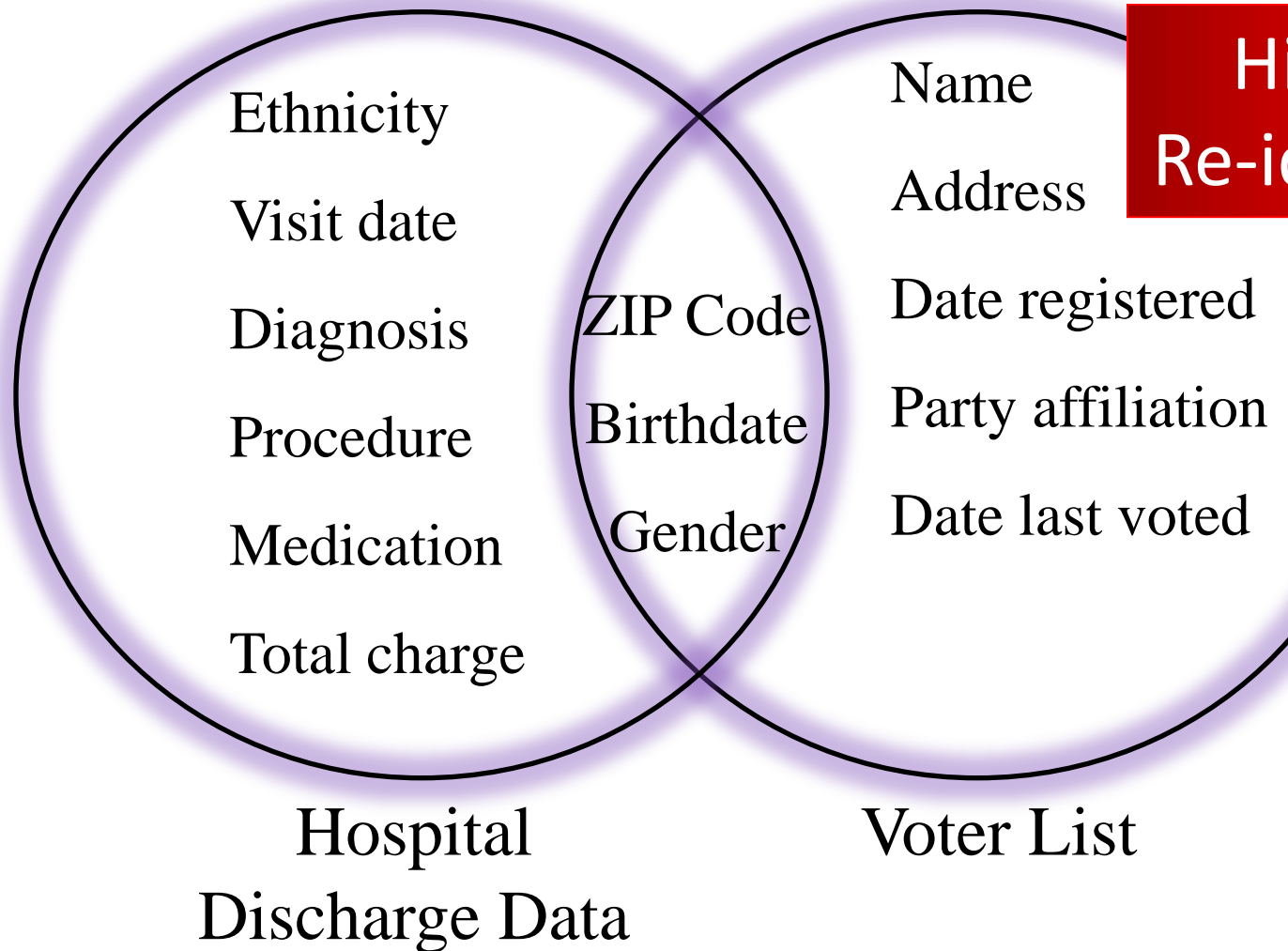
HIPAA Expert Determination (abridged)

Certify via “generally accepted statistical and scientific principles & methods, that the **risk is very small** that the information could be used, alone or in combination with other **reasonably available information**, by the **anticipated recipient** to identify the subject of the information.”

A Simplified View on Risk



A “Quasi-identifier” Conundrum



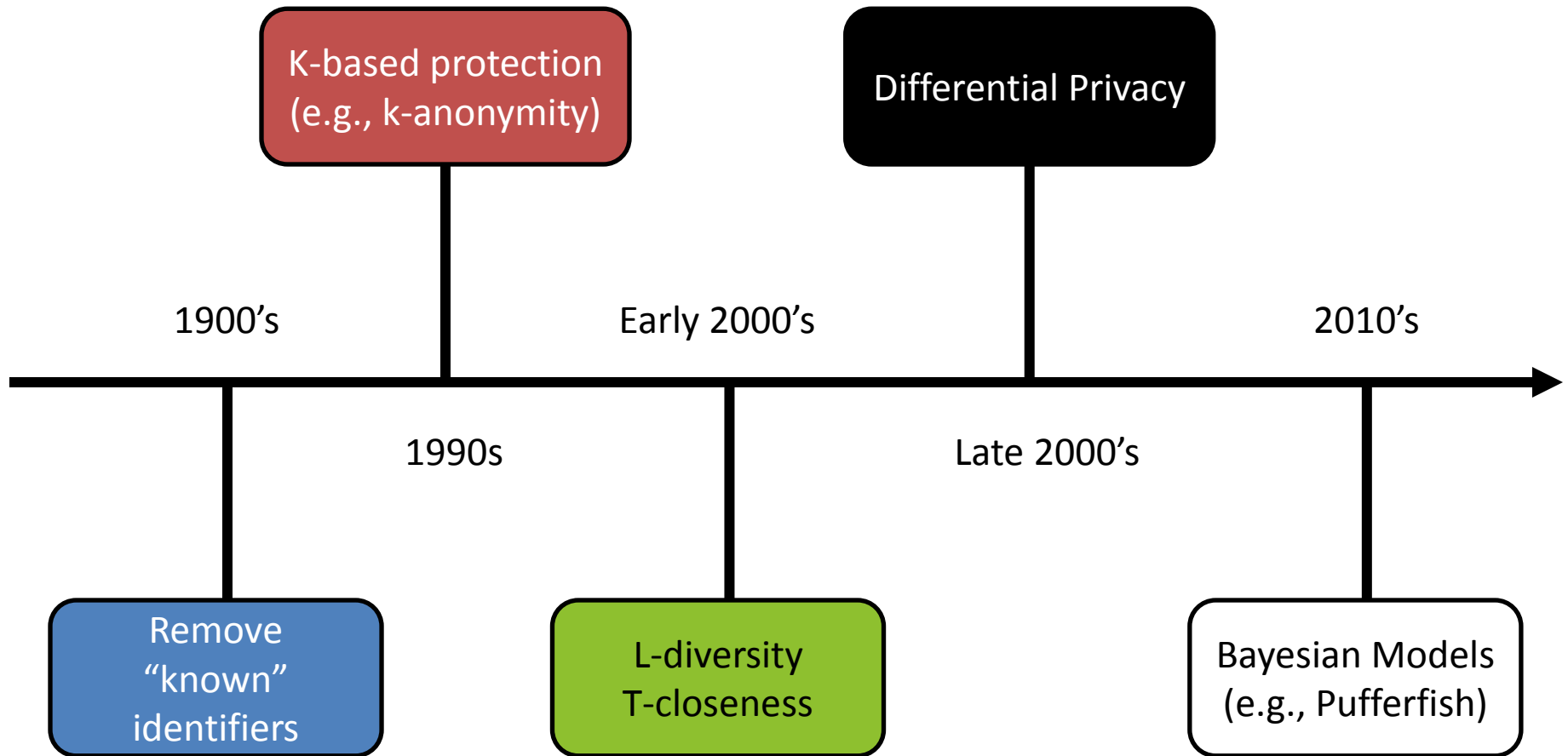
**High Profile
Re-identification**



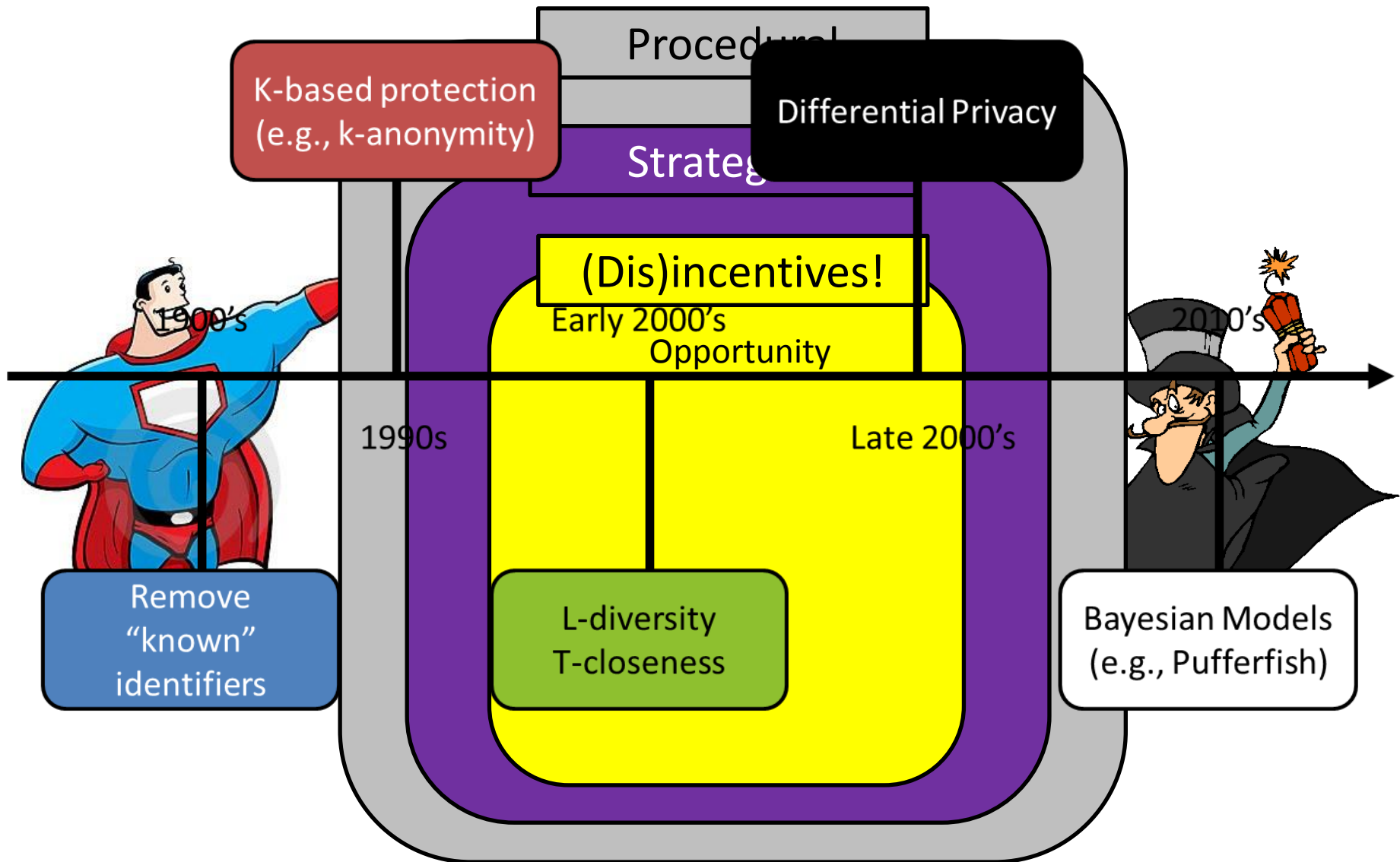
Availability of Demographics Varies...

	Illinois	Minnesota	Tennessee	Washington	Wisconsin
WHO	Registered Political Committees (ANYONE – In Person)	MN Voters	Anyone	Anyone	Anyone
Format	Disk	Disk	Disk	Disk	Disk
Cost	\$500	\$46; “use ONLY for elections, political activities, or law enforcement”	\$2500	\$30	\$12,500
Name	●	●	●	●	●
Address	●	●	●	●	●
Date of Birth	●	○	●	●	
Sex	●		●	●	
Race			●		
Phone Number	●	●			

Technical Approaches to Privacy



An Augmented View of Data Privacy



Stackelberg Game

Sharing Strategy 1
<i>Utility 1</i>
<i>Risk ???</i>

Attack Strategy A
<i>Utility A</i>
<i>Risk A</i>

Strategies:

- Generalize Demographics
- Perturb Statistics
- Apply Data Use Agreement
- ...
- Charge for Access

Attack Strategy B
<i>Utility B</i>
<i>Risk B</i>

⋮

Attack Strategy C
<i>Utility C</i>
<i>Risk C</i>

Publisher

Recipient

Stackelberg Game

Sharing Strategy 1
<i>Utility 1</i>
<i>Risk ???</i>

Attack Strategy A
<i>Utility A</i>
<i>Risk A</i>

Attack Strategy B
<i>Utility B</i>
<i>Risk B</i>

**Recipient's
Best Strategy**

⋮

Attack Strategy C
<i>Utility C</i>
<i>Risk C</i>

Publisher

Recipient

Stackelberg Game

Sharing Strategy 1
<i>Utility 1</i>
<i>Risk B</i>

Attack Strategy A
<i>Utility A</i>
<i>Risk A</i>

Attack Strategy B
<i>Utility B</i>
<i>Risk B</i>

Attack Strategy C
<i>Utility C</i>
<i>Risk C</i>

Recipient's
Best Strategy

Publisher

Recipient

Stackelberg Game

Sharing Strategy 1
<i>Utility 1</i>
<i>Risk B</i>

Sharing Strategy 2
<i>Utility 2</i>
<i>Risk ???</i>

Publisher

Attack Strategy A
<i>Utility A</i>
<i>Risk A</i>

Attack Strategy B
<i>Utility B</i>
<i>Risk B</i>

⋮

Attack Strategy C
<i>Utility C</i>
<i>Risk C</i>

Recipient

Stackelberg Game

Sharing Strategy 1
<i>Utility 1</i>
<i>Risk B</i>

Sharing Strategy 2
<i>Utility 2</i>
<i>Risk A</i>

**Recipient's
Best Strategy**

Attack Strategy A
<i>Utility A</i>
<i>Risk A</i>

Attack Strategy B
<i>Utility B</i>
<i>Risk B</i>

⋮

Attack Strategy C
<i>Utility C</i>
<i>Risk C</i>

Publisher

Recipient

Stackelberg Game

Sharing Strategy 1
<i>Utility 1</i>
<i>Risk B</i>

Sharing Strategy 2
<i>Utility 2</i>
<i>Risk A</i>

⋮

Sharing Strategy Z
<i>Utility Z</i>
<i>Risk Z</i>

Publisher

Stackelberg Game

Sharing Strategy 1
<i>Utility 1</i>
<i>Risk B</i>

Sharing Strategy 2
<i>Utility 2</i>
<i>Risk A</i>

⋮

Sharing Strategy Z
<i>Utility Z</i>
<i>Risk Z</i>

Publisher

Choose strategy that maximizes overall benefit

Optimizes the Risk-Utility tradeoff


Demographic Case Study

{Date of Birth, Gender, Geocode, Race}

- ~30,000 Census records
- Average Payoff Per Record
- \$1200: Benefit per record
- \$300: Cost per violation
- \$4: Access cost per record


People Search - Updated Daily, Accurate and Fast!

People Search



First Name	M.I.	Last Name required	City and/or State	Search
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	

Reverse Phone Lookup



Phone Number	Search
(<input type="text"/>) <input type="text"/> - <input type="text"/>	

More ways to get info you need:

- ✦ Perform a Background Check
- ✦ Run a Background Check by SSN
- ✦ Perform an Address Lookup
- ✦ Do a Reverse Phone Lookup

What is People Search?

It's a confidential way to find people so you can reconnect or just get more info on a person. People Search reports can include phone numbers, address history, age & date of birth, relatives, and more. Find a person you're curious about – search today!



What is Reverse Phone Lookup?

It's a confidential way to find out who a phone number belongs to. Reverse phone search works for landline, unlisted & non-published numbers, and cell phone lookups. Reports can include phone type, owner name, address & more. Curious? Do a phone number lookup!



Game Variations

- Safe Harbor (SH) Game
 - Defender shares data according to federal policy
- Basic Game
 - Defender shares data to maximize overall payoff
- SH-Friendly
 - Defender constrains strategy space to disclose no greater detail than SH
- No Attack
 - Defender constrains strategy space to disclose no greater detail than SH

Demographic Case Study

- ~30,000 Census records
- Average Payoff Per Record
- \$1200: Benefit per record
- \$300: Cost per violation
- \$4: Access cost per record



Demographic Case Study

- ~30,000 Census records
- Average Payoff Per Record

- \$1200: Benefit per record
- \$300: Cost per violation
- \$4: Access cost per record



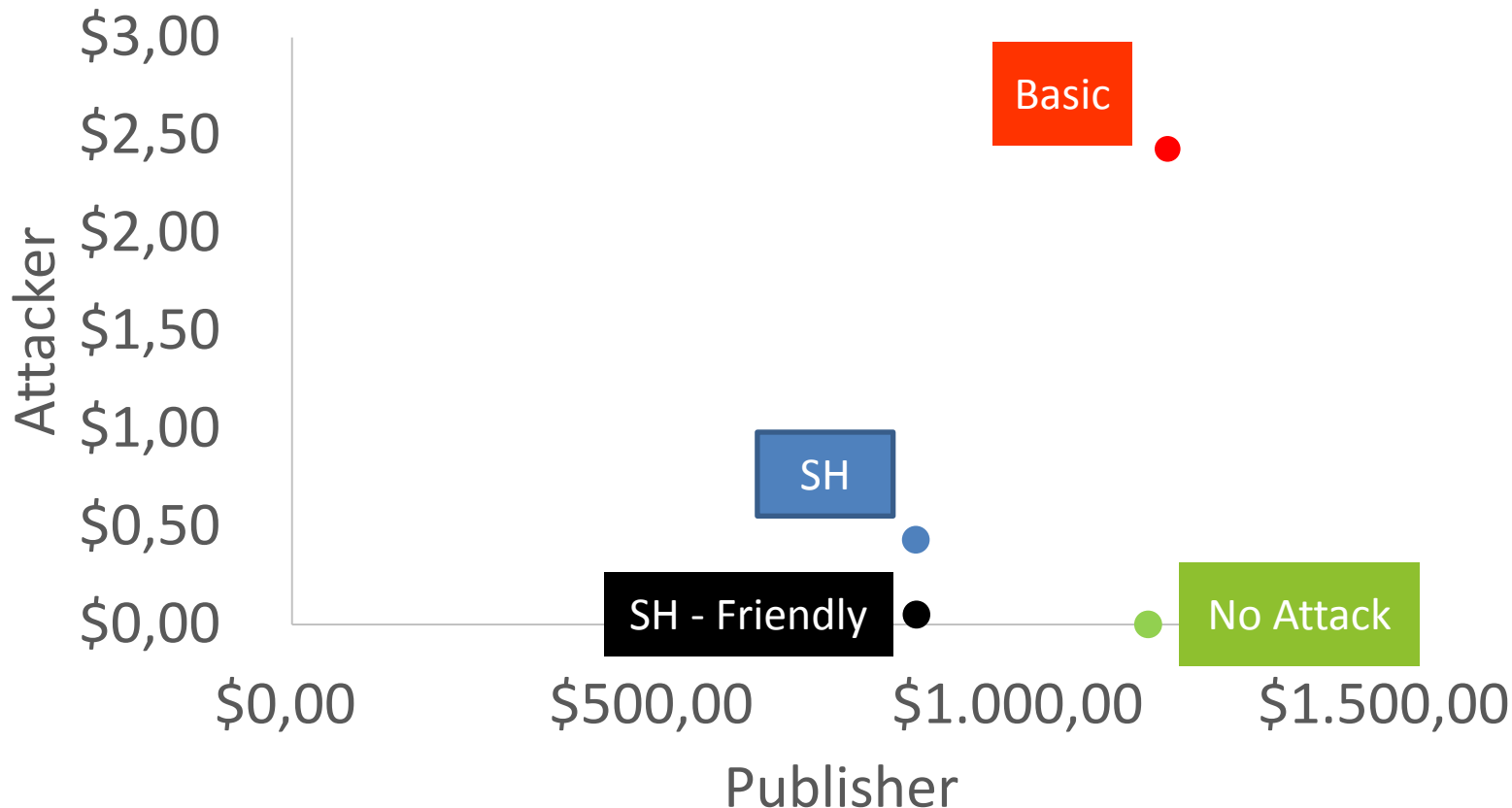
Demographic Case Study

- ~30,000 Census records
- Average Payoff Per Record
- \$1200: Benefit per record
- \$300: Cost per violation
- \$4: Access cost per record



Demographic Case Study

- ~30,000 Census records
- Average Payoff Per Record
- \$1200: Benefit per record
- \$300: Cost per violation
- \$4: Access cost per record





ARX – Powerful Data Anonymization

A comprehensive software for risk- and utility-based privacy-preserving microdata publishing

[Home](#)[Overview](#)[Anonymization tool](#)[API](#)[Development](#)[Downloads](#)[Imprint](#)

Sign up for our newsletter to receive notifications about new releases, presentations and publications.

ARX

Data Anonymization Tool

ARX is a comprehensive open source software for anonymizing sensitive personal data. It supports:

- Risk-based anonymization using super-population models, strict-average risk and k-map
- Syntactic privacy models, such as k-anonymity, ℓ -diversity, t-closeness, δ -disclosure privacy and δ -presence
- Semantic privacy models, such as (ϵ, δ) -differential privacy
- Methods for optimizing the profitability of data publishing based on monetary cost-benefit analyses
- Data transformation with generalization, suppression, microaggregation and top/bottom-coding as well as global and local recoding
- Methods for analyzing data utility
- Methods for analyzing re-identification risks

The software is able to handle very large datasets on commodity hardware and it features an intuitive cross-platform graphical user interface. You can find further information [here](#), or directly proceed to our [downloads](#) section.

<http://arx.deidentifier.org/>



Anonymisation is NOT a Panacea

- There is *always* a risk of re-identification
- But risk exists in any security setting
- The challenges are
 - Determine an appropriate level of risk
 - Ensure accountability
- Combine with data use agreements
- Risk is proportional to anticipated recipient trustworthiness (public vs. vetted investigator)

Acknowledgements

- Kathleen Benitez
- Ellen Wright Clayton, M.D., J.D.
- Josh Denny, M.D.
- Xiaofeng Ding, Ph.D. (U. So. Australia)
- Khaled El Emam, Ph.D. (U. Ottawa)
- James Gaupp, M.S.
- Aris Gkoulalas-Divanis, Ph.D. (IBM)
- Jonathan Haines, Ph.D.
- Paul Harris, Ph.D.
- Raymond Heatherly, Ph.D.
- Murat Kantarcioglu, Ph.D. (U. of Texas)
- Klaus Kuhn, M.D. (TUM)
- Jiuyong Li, Ph.D. (U. So. Australia)
- Grigorios Loukides, Ph.D. (Kings College London)
- Dan Masys, M.D. (U. Washington)
- Steve Nyemba, M.S.
- Fabian Prasser (TUM)
- Dan Roden, M.D.
- Latanya Sweeney, Ph.D. (Harvard U.)
- Acar Tamersoy (Georgia Tech)
- Weiyi Xia, Ph.D.
- Yevgeniy Vorobeychik, Ph.D. (Washington Uni. St. Louis)
- Zhiyu Wan, M.S.

eMERGE Teams

- Boston Children's Hospital
- Children's Hospital of Philadelphia
- Cincinnati Children's Hospital
- Geisinger Health System
- Group Health Research Institute / U. Washington
- Marshfield Clinic
- Mayo Clinic
- Mt. Sinai Medical Center / Columbia University
- Northwestern University
- Vanderbilt University

Funding

- NHGRI @ NIH
 - U01 HG006385 (eMERGE)
 - U01 HG006378 (VGER)
 - RM1 HG009034 (GetPreCiSe)
- NLM @ NIH
 - R01 LM009989

Questions?

b.malin@vanderbilt.edu

Center for Genetic Privacy and Identity in Community Settings

<http://www.vumc.org/getprecise/>

Vanderbilt Health Data Science Center

<http://www.vumc.org/heads/>

Vanderbilt Health Information Privacy Laboratory

<http://www.hiplab.org/>