





Combined Committee Meeting
Sep 9, 2019

Privacy Preserving Interactive Record Linkage (PPIRL)
via Information Suppression



8/31/2019 1

1



Agenda

- Short Introductions (10 min)
- Project Overview (5 min)
 - Project ends in 6 months
- Results (15 min) so far
- Plan for large scale privacy survey (1h)
 - We need your input
- Open Discussion (30 min)
- REMINDER: Last advisory meeting will be mid Jan.

8/31/2019 2

2

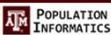


Short Introductions: Committee Members

- User Committee
 - Jeffrey Curtis, Consultant, UAB, Clinical, Research Data Network PI, CER, PCOR, ELSI
 - Michael Morrissey, Texas A&M Univ., Linking claims data
 - Ben Nowell, PPRN, Great Healthy Living Foundation; Arthritis Power Patient Powered Research Network (PPRN)
 - Alison Fraser, U of Utah, Linking data for cancer outcomes
- Methods Committee
 - Jeff Baumes, Kitware, Open Source health application. HCI
 - Peter Yu, Texas A&M Univ., HIPAA Privacy Officer
 - Daniel Basile, co-Investigator, Texas A&M Uni., Patient (chronic illness), Security Expert, IT support research
 - [Late] Li Xiong, Emory Uni., PI of PCORI project on Privacy

8/31/2019 3

3



Our team


- Hye-Chung Kum, Principal Investigator, Texas A&M Univ., Computer Science (information privacy), secondary data analysis (user)
- Alva Ferdinand, Aim 3 lead, Texas A&M Univ., Public Health and Law, secondary data analysis (user)
- Cason Schmit, Aim 3 co-lead, Texas A&M Univ., Public Health and Law, Information Privacy, IRB, DUA
- Eric Ragan, Aim 1 lead, Univ. of Florida, CHI (computer human interaction)
- GARs:
 - Theo & Kobi (public health)
 - Mahin, Qinbo & Guru (computer science)

8/31/2019 5

5

Project Overview

Only FYI. Will skim very quickly in the meeting to remind everyone.



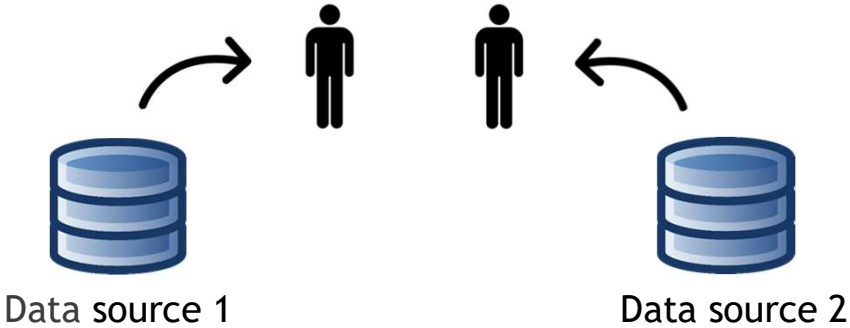
8/31/2019 6

6

Record Linkage for Person-Level Data Privacy Enhanced System using Privacy-by-Design

POPULATION INFORMATICS

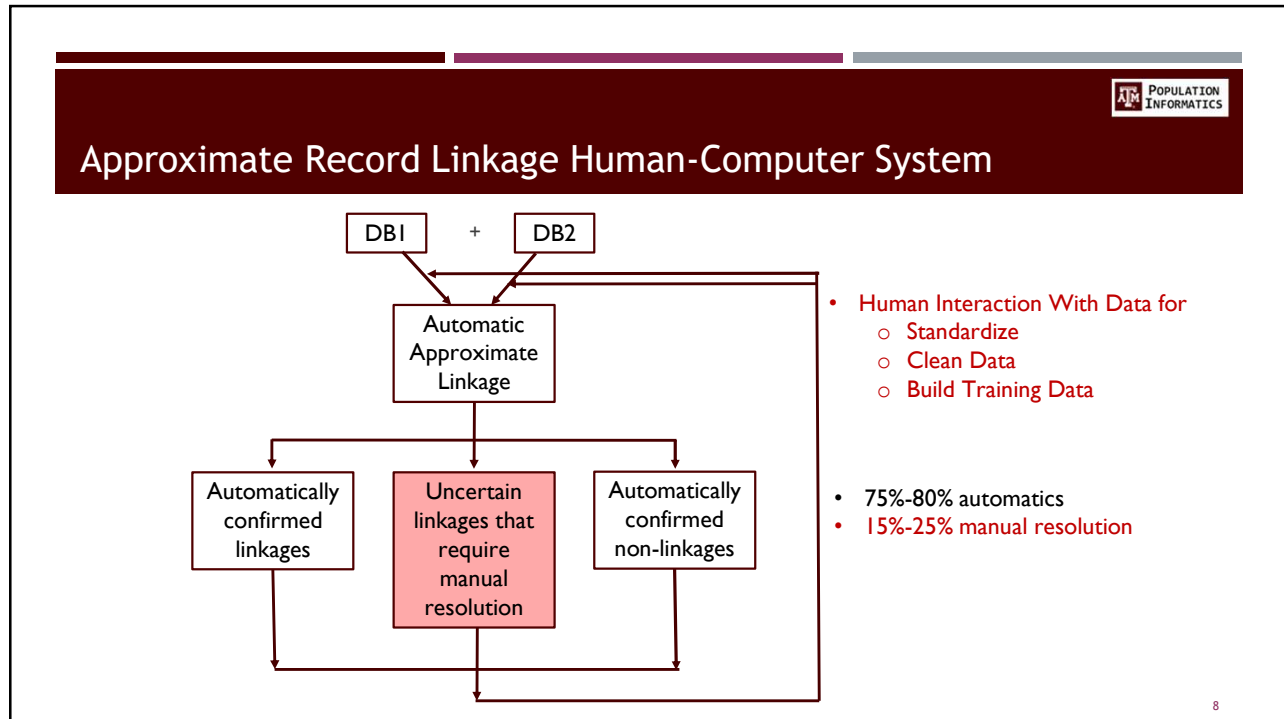
Same person?
(How many emergency department visits last year?)



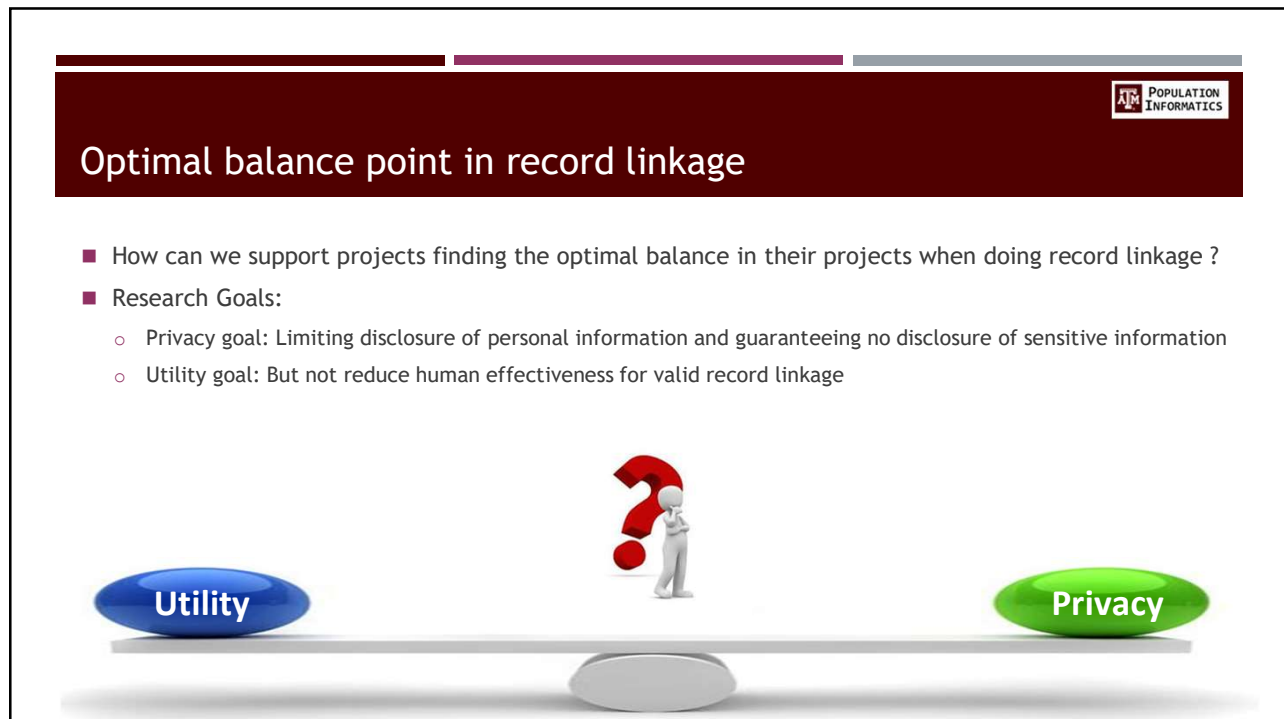
Data source 1 Data source 2

7

7



8



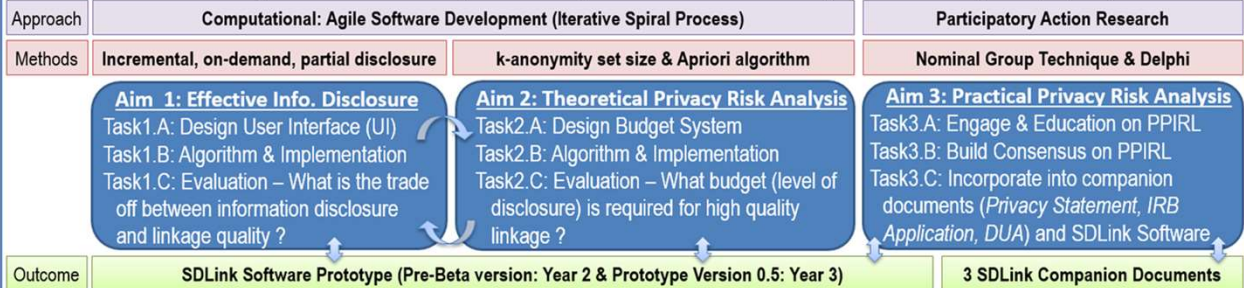
9

Aims & Outcomes Prototype software & companion documents



Phase 1 – Completed Framework on Privacy Preserving Interactive Record Linkage (PPIRL): Privacy & Utility Objective

Phase 2 – Research Needed: Algorithm & Methods Development for Design of SDLink Software and Companion Documents (PCORI proposal)



Phase 3 – After Project is Completed: Hardening Code – SDLink Software Development & Release (Collaboration with Kitware Inc.)

10

10

Summary of Results <https://pinformatics.org/ppirl/>



Aims 1 & 2: open source software

- MINDFIRL: Develop and release open source prototype software in git
- SIG CHI 2018 Best paper award
 - User study of static design
- SOUPS 2019
 - User study of over all system
 - Expert user study
- Papers in progress
 - KAPR score: quantifying the risk (ArXiv; Poster at AMIA 2019)
 - UT Houston & UAB formative study
- **TODO in 6 months**
 - UAB PCORnet formative study
 - Last updates to the software to include optional fields
 - Final release of MINDFIRL

Aim 3: accompanying documents

- Privacy Statement: FAQ
 - <http://mindfirl-uth.herokuapp.com/faq>
 - NGT & Delphi with patients
 - **TODO: large scale survey**
- Template IRB applications
 - NGT & Delphi with ELSI experts
- Template DUA
 - Draft completed
 - **TODO: Expert review**

8/31/2019

11

11

Aims 1 & 2: Real Question

■ Can we find the “sweet spot” between accessing PII for legitimate use while providing the maximum privacy protection as possible through the privacy by design approach by

YES!!

Privacy by Design Works

Significantly improved privacy for same quality of results
no extra time

PRIVACY RISK

Approach	Privacy Risk
FULL ACCESS	100%
STATIC DESIGN	30%
ON-DEMAND DESIGN	7.80%

12

Three Design Elements for Implementing the Minimum Necessary Standard

2

Privacy risk: 38.3% + 1.56%

3

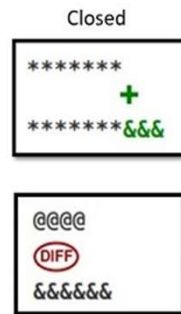
Pair	ID	FFreq	First Name	Last Name	LFreq	DoB(M/D/Y)	Sex	Race	Choice Panel
1	1995553862	...	WILLIAM	KING JR	...	01/25/1968	F	W	Our Proposed Key Design Elements
	?	...	WILLIAM	KING	...	01/25/1968	M	W	
2	1000563341	∞	***MY	**W***	...	07/03/****	✓	✓	<ol style="list-style-type: none"> 1. Minimum Disclosure via Interactive Just-in-Time Interface <ul style="list-style-type: none"> • Hide data values (when possible) • Add visual meta-data to help decision making without seeing raw data 2. Accountability via Quantified Privacy Risk 3. Limiting Privacy Risk via Budget
	(DIFF)	∞	+	X	...	X	✓	✓	
3	1000391562	∞	***	**R***	...	03/07/****	✓	✓	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;"> <p>****@**** (1) @@@@@@</p> <p>****@**** (2.5) @@@@@@</p> </div> <div style="font-size: 2em; color: blue;">X</div> <div style="text-align: center;"> <p>@@@@@</p> <p>@@@@@ (1)</p> </div> </div>
	****@**** (1) @@@@@@	∞	**/**/****@	✓	✓	****@**** (2.5) @@@@@@	**/**/****@	✓	

13

Our proposed approach 1: Interactive Interfaces Dynamic On-demand Incremental Disclosure



- Dynamic: Click to see more
- On-demand: When needed
 - Just-in-time decision
- Incremental: As needed
 - Not all at once
- Allow for easy accountability in information Use



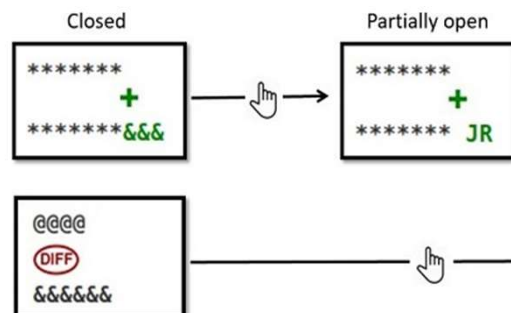
14

14

Our proposed approach 1: Interactive Interfaces Dynamic On-demand Incremental Disclosure



- Dynamic: Click to see more
- On-demand: When needed
 - Just-in-time decision
- Incremental: As needed
 - Not all at once
- Allow for easy accountability in information Use



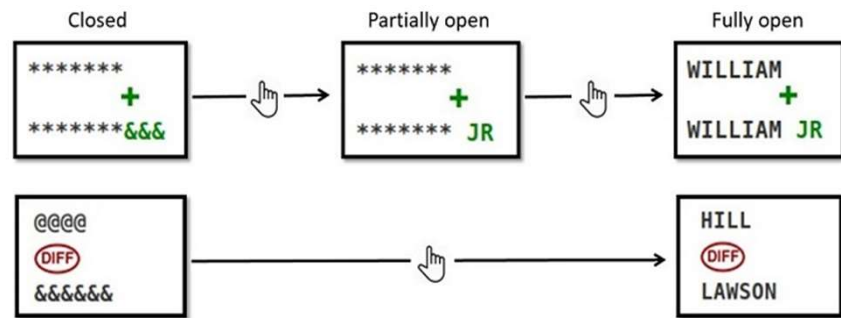
15

15

Our proposed approach 1: Interactive Interfaces Dynamic On-demand Incremental Disclosure



- Dynamic: Click to see more
- On-demand: When needed
 - Just-in-time decision
- Incremental: As needed
 - Not all at once
- Allow for easy accountability in information Use



16

16

Aims 1 & 2: Expert Study Results Compared to Full access to PII



- Five of the experts normally conducted record linkage with full access to PII
- They perceived that this system
 - offered more privacy protection
 - with little to no impact on accuracy in the linkage
 - but may take more time
- Evidence for improving linkage (i.e., more consistent linkage decisions) by providing better processed information for decision making in place of raw data



“Once I got used to the coding, allowing partial disclosure helped in decision making”

17

17

Aims 1 & 2: Expert Study Results Compared to Encryption Based No Access to PII



- One expert had prior experience using encryption-based methods of data hiding for private record linkage with no access to PII.
- Compared to the encryption-based method, this participant perceived our system
 - to have less protection
 - and require more time
 - but to also allow for much better accuracy



“I never know how well the hashing worked, or how accurate it is. It would be helpful to use this method to spot check a random sample (e.g., 5%)”

- This seems to agree with our goal of providing a level of access between the all or nothing that provides better accuracy than no access, but more protection than full access.

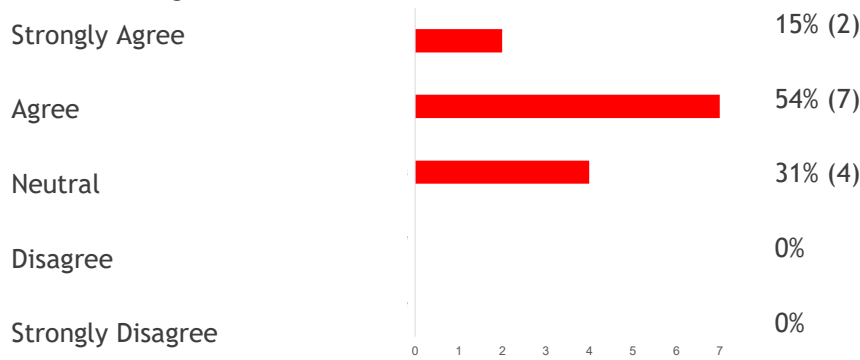
18

18

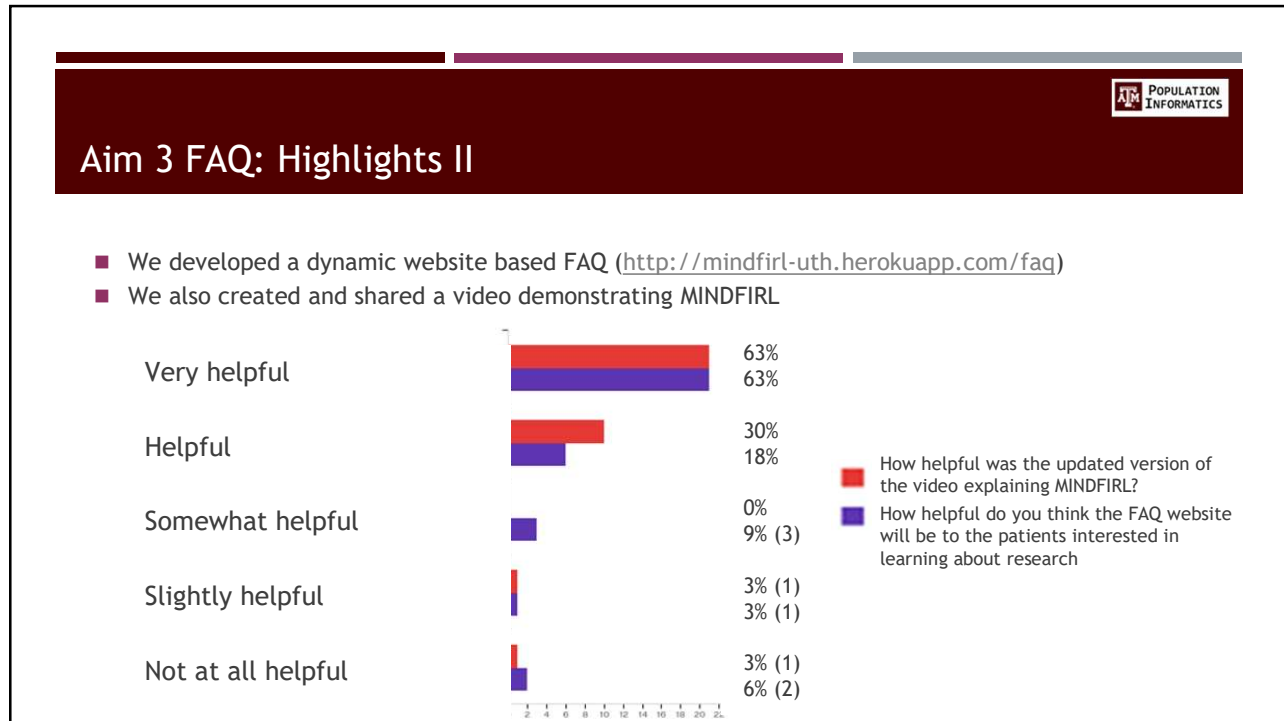
Aim 3 IRB template Highlights (N=13)



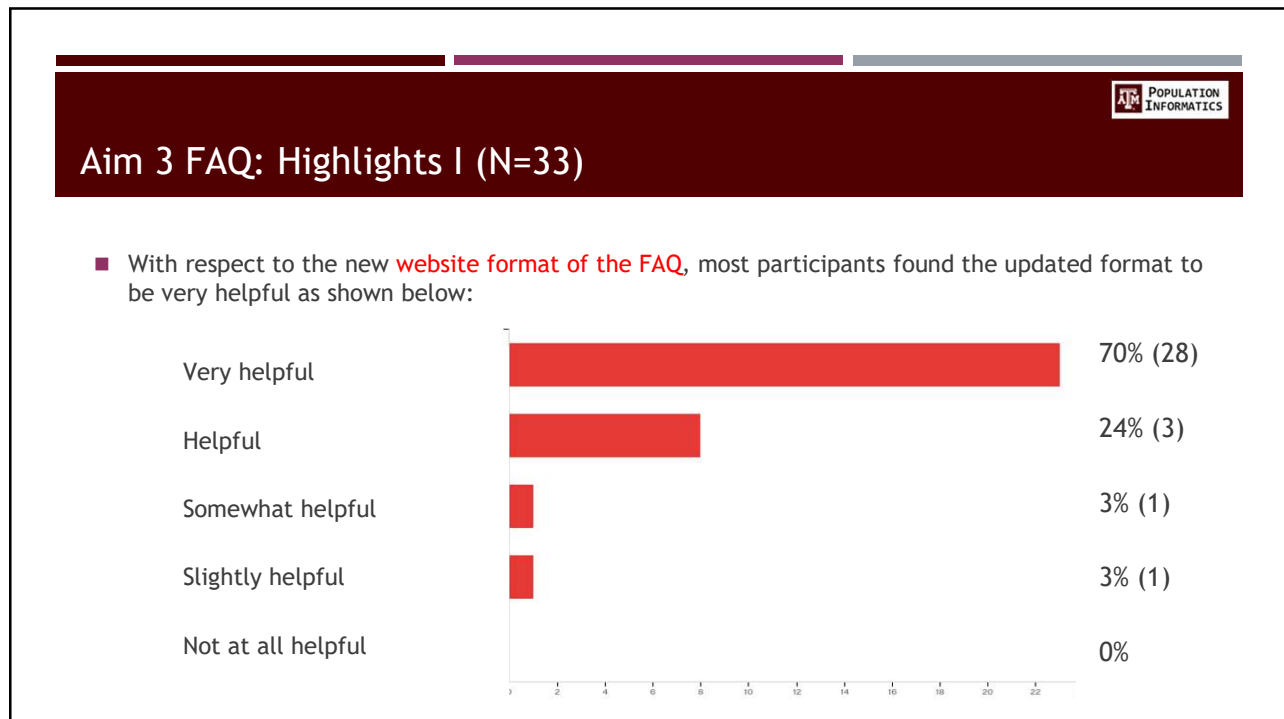
- We asked ELSI experts about their opinion on risk reduction to minimum when using MINDFIRL
- “The use of the MINDFIRL software will further reduce risk to the minimum necessary to conduct reliable record linkage.”



19



20




21

Privacy Survey

22

22



Studies of Public Perceptions of Privacy and Data Use

- Surveys on privacy exist. Still there is an opportunity to contribute
 - Perceptions change
 - Relevance to current policy debate
 - Changes to IRB laws (i.e., Common Rule) raise new questions
 - Forcing hard decisions

Individuals' concerns about the privacy and security of both paper and electronic medical records declined significantly between 2013 and 2014

Figure 1: Proportion of individuals who expressed concerns regarding the privacy and security of their medical record and withheld information from their healthcare provider due to those concerns, 2012-2014

Year	Very or somewhat concerned with privacy of medical records	Very or somewhat concerned about security of medical records	Withholding information from health care provider due to privacy or security concerns
2012	77%	72%	7%
2013	75%	69%	8%
2014	58%*	52%*	5%

NOTE: *Significantly different from 2013 and 2012 (p<.05)
SOURCE: 2012 - 2014 Consumer Survey of Attitudes Toward the Privacy and Security Aspects of Electronic Health Records and Health Information Exchange

<https://dashboard.healthit.gov/evaluations/data-briefs/trends-individual-perceptions-privacy-security-ehrs-hie.php>

2/25/2020

23

23

What types of privacy (or data use) questions do you think would be interesting to ask the public?

Start the presentation to see live content. Still no live content? Install the app or get help at PollEv.com/app

2/25/2020

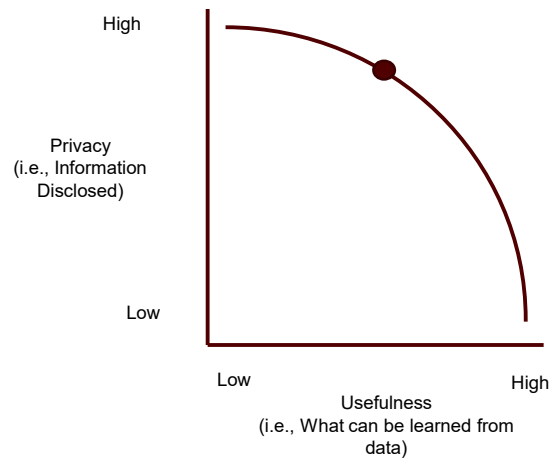
24

24

Privacy Survey



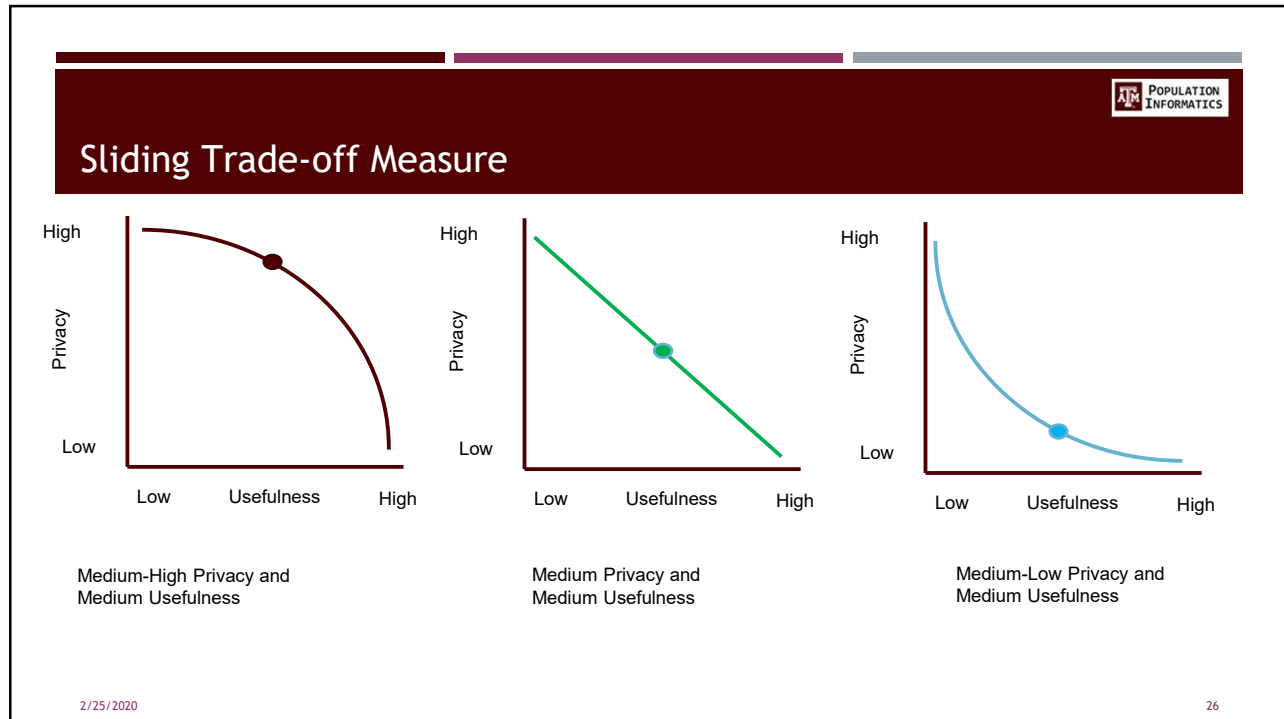
- Privacy Utility Trade-off
 - Restrict data element/dataset disclosures → limits the utility of data.
 - People like privacy and research independently
 - Unclear how people will score one if it means giving up the other.
- Curved sliding-scale measure
 - Forces people to choose the appropriate balance between competing concepts



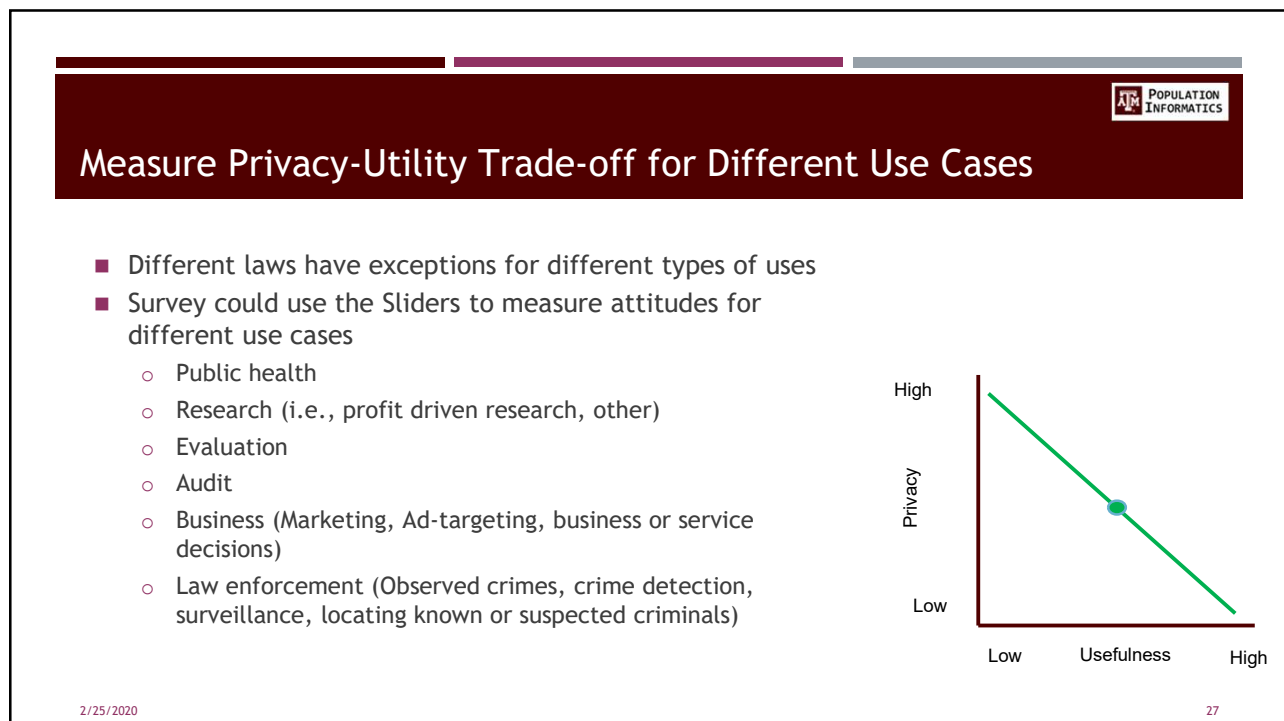
2/25/2020

25


25



26



27




Ethics Relating to Data Use

- Bioethics (i.e., Common Rule)
 - Respect for Persons (e.g., informed consent)
 - Beneficence (maximize benefits; minimize risks)
 - Justice
- Public health ethics (e.g., WHO guidelines for ethical issues in surveillance)
 - Common Good
 - Equity
 - Respect for Persons
 - Good Governance

Since ethical issues/concerns frequently inhibit data projects, are there questions that we can ask that will help navigate these issues?

2/25/2020
28

28



Potential Issues to Explore

- Willingness to have own data used for various activities
- Risk perceptions for certain activities
- Benefit or value perceptions for certain activities
- Fairness perceptions relating to data use for certain activities
- Data governance priorities
- Right of notice v. de-identification
- Right of consent v. de-identification
- Risk tolerance and “de-identification”
- Broad consent
- Perceptions of security (including as it relates to comfort with specified activities)
- Perceptions of ethical oversight (including as it relates to comfort with specified activities)
- Factors that increase willingness or comfort with the use of personal data
- Privacy v. Cost (e.g., difficulty linking records)

Legal implications: disclosure exceptions

Relate to various ethical principles used for evaluating data use

2/25/2020
29

29

What types of privacy (or data use) questions do you think would be interesting to ask the public?

2/25/2020

Start the presentation to see live content. Still no live content? Install the app or get help at PollEv.com/app

30

30

Thank You!!



Hye-Chung Kum (kum@tamu.edu)

Population Informatics Lab (<https://pinformatics.org/>)

Privacy is a BUDGET constrained problem

The goal is to achieve the maximum utility under a fixed privacy budget

Utility

Privacy

31

31