Differential Privacy: An Overview

Salil Vadhan

Center for Research on Computation & Society

School of Engineering & Applied Sciences
Harvard University

“Privacy Tools for Sharing Research Data”
Summer 2014 Orientation

CRCS Center for Research on Computation and Society

Data Privacy: The Problem

Given a dataset with sensitive information, such as:

- Census data
- Health records
- Social network activity
- Telecommunications data

- Academic research
- Informing policy
- Identifying subjects for drug trial
- Searching for terrorists
- Market analysis
- …

How can we:

- enable "desirable uses" of the data
- while protecting the "privacy" of the data subjects?

????

Approach 1: Encrypt the Data

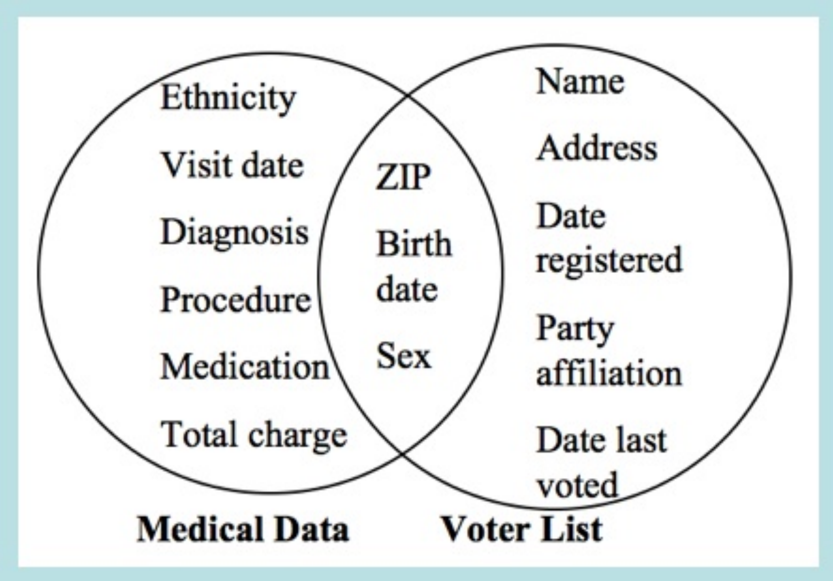
Name	Sex	Blood	⋯	HIV?
Chen	F	B	⋯	Y
Jones	M	A	⋯	N
Smith	M	O	⋯	N
Ross	M	O	⋯	Y
Lu	F	A	⋯	N
Shah	M	B	⋯	Y

Name	Sex	Blood	⋯	HIV?
100101	001001	110101	⋯	110111
101010	111010	111111	⋯	001001
001010	100100	011001	⋯	110101
001110	010010	110101	⋯	100001
110101	000000	111001	⋯	010010
111110	110010	000101	⋯	110101

Problems?

Approach 2: Anonymize the Data

~~Name~~	Sex	Blood	⋯	HIV?
~~Chen~~	F	B	⋯	Y
~~Jones~~	M	A	⋯	N
~~Smith~~	M	O	⋯	N
~~Ross~~	M	O	⋯	Y
~~Lu~~	F	A	⋯	N
~~Shah~~	M	B	⋯	Y



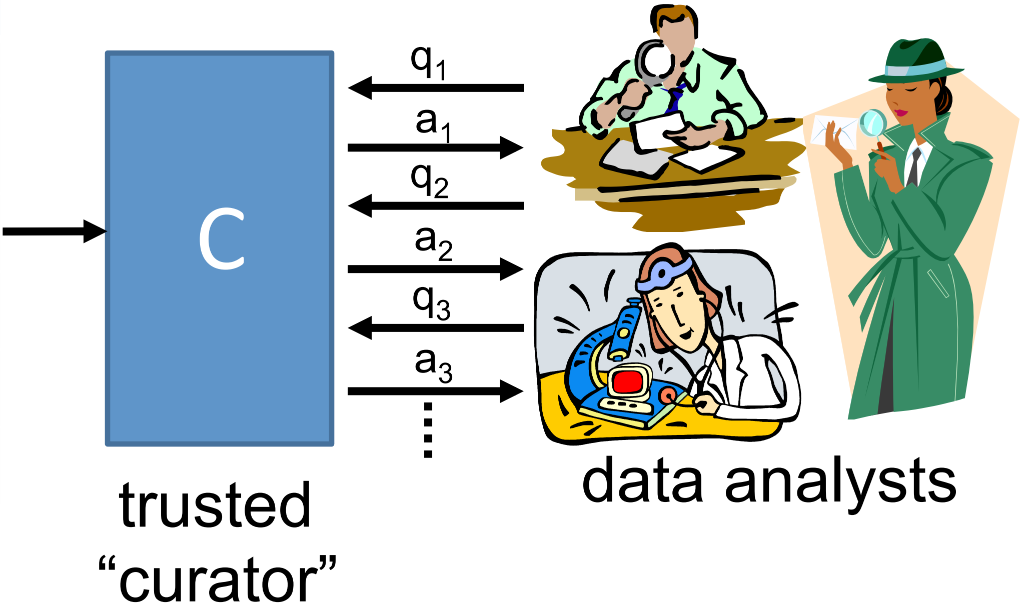
[Sweeney `97]

“re-identification” often easy

Problems?

Approach 3: Mediate Access

Name	Sex	Blood	⋯	HIV?
Chen	F	B	⋯	Y
Jones	M	A	⋯	N
Smith	M	O	⋯	N
Ross	M	O	⋯	Y
Lu	F	A	⋯	N
Shah	M	B	⋯	Y



C

q_1

a_1

q_2

a_2

q_3

a_3

⋮

trusted
“curator”

data analysts

Problems?

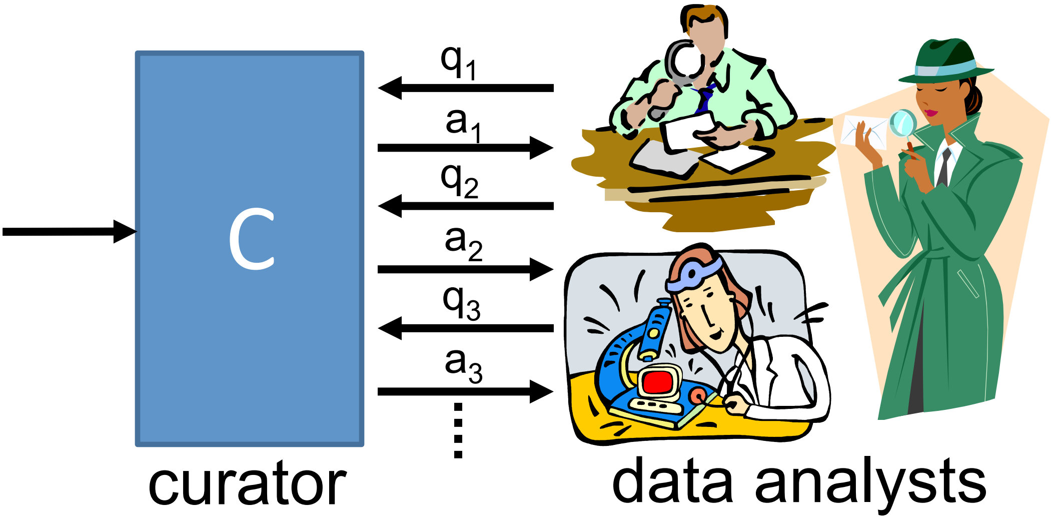
Privacy Models from CS

Model	Utility	Privacy	Who Holds Data?
Differential Privacy	statistical analysis of dataset	individual-specific info	trusted curator
Secure Function Evaluation	any query desired	everything other than result of query	original users (or semi-trusted delegates)
Fully Homomorphic (or Functional) Encryption	any query desired	everything (except possibly result of query)	untrusted server

Differential privacy

[Dinur-Nissim ’03+Dwork, Dwork-Nissim ’04, Blum-Dwork-McSherry-Nissim ’05, Dwork-McSherry-Nissim-Smith ’06]

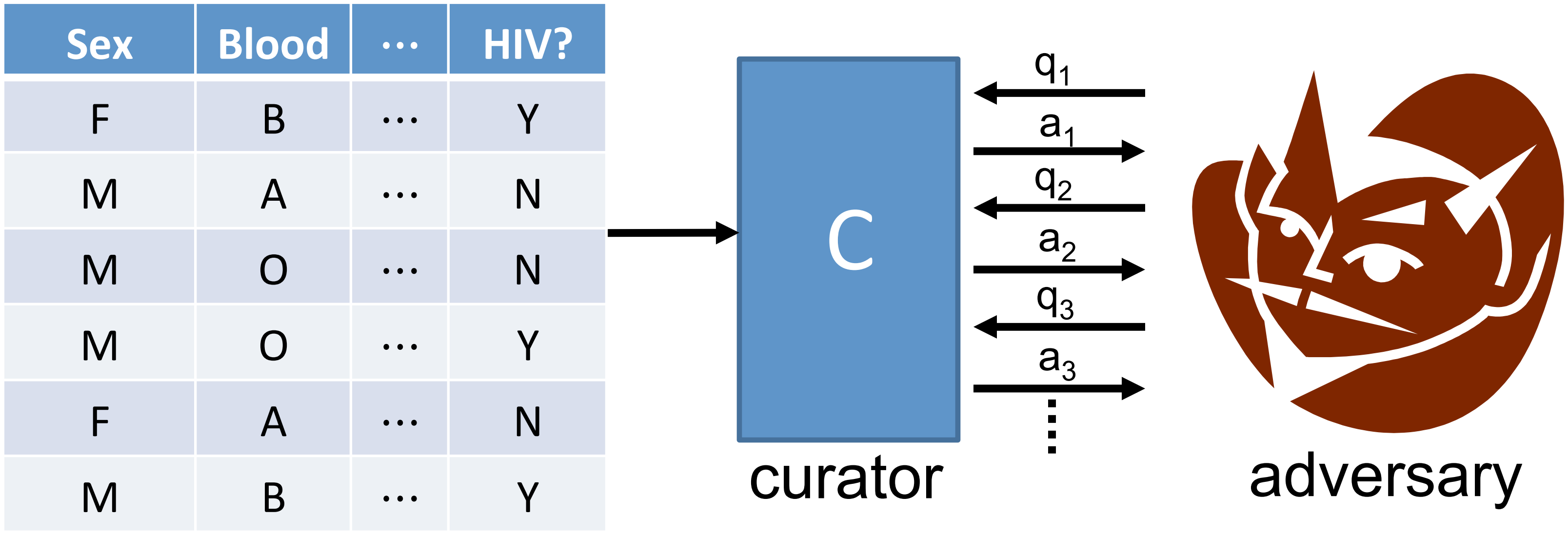
Sex	Blood	⋯	HIV?
F	B	⋯	Y
M	A	⋯	N
M	O	⋯	N
M	O	⋯	Y
F	A	⋯	N
M	B	⋯	Y



Requirement: effect of each individual should be “hidden”

Differential privacy

[Dinur-Nissim ’03+Dwork, Dwork-Nissim ’04, Blum-Dwork-McSherry-Nissim ’05, Dwork-McSherry-Nissim-Smith ’06]



Sex	Blood	⋯	HIV?
F	B	⋯	Y
M	A	⋯	N
M	O	⋯	N
M	O	⋯	Y
F	A	⋯	N
M	B	⋯	Y

C

curator

q_1

a_1

q_2

a_2

q_3

a_3

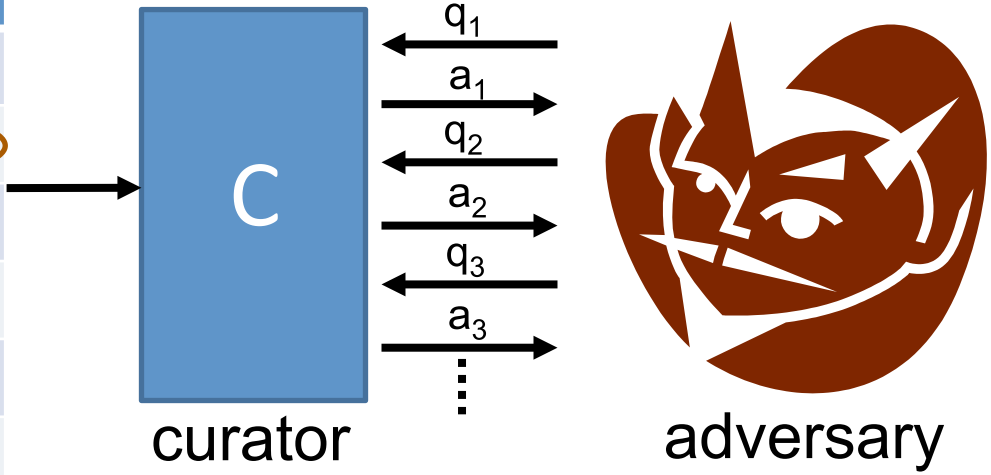
⋮

adversary

Differential privacy

[Dinur-Nissim ’03+Dwork, Dwork-Nissim ’04, Blum-Dwork-McSherry-Nissim ’05, Dwork-McSherry-Nissim-Smith ’06]

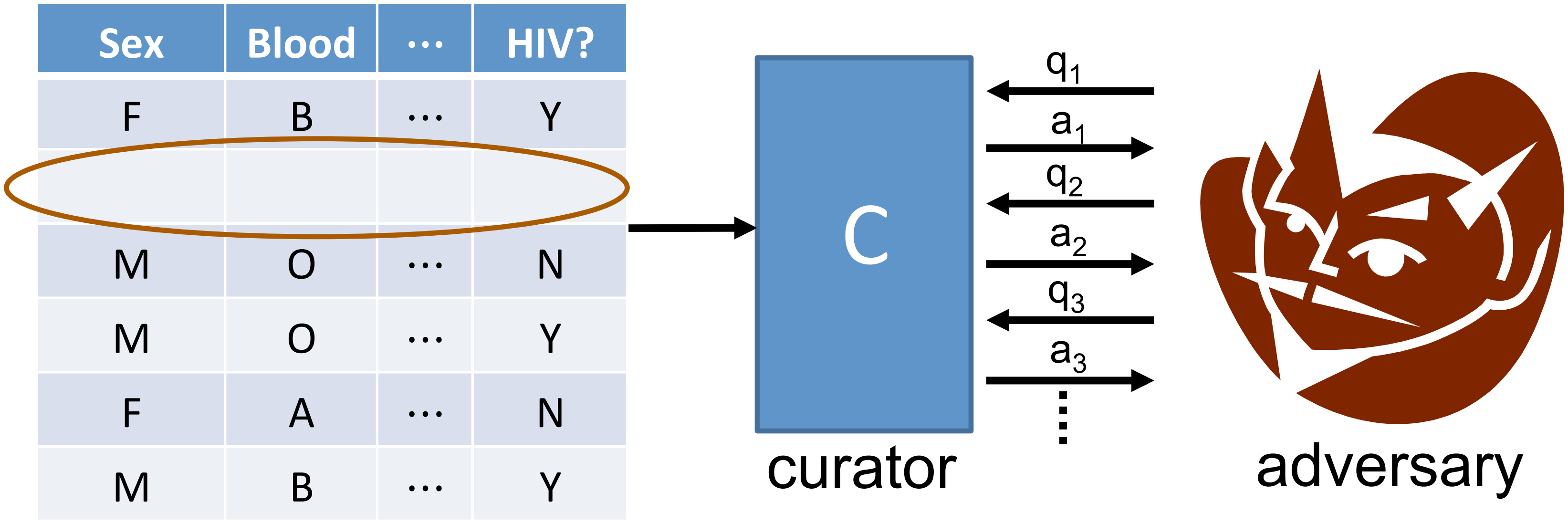
Sex	Blood	⋯	HIV?
F	B	⋯	Y
M	**A**	⋯	**N**
M	O	⋯	N
M	O	⋯	Y
F	A	⋯	N
M	B	⋯	Y



Requirement: an adversary shouldn’t be able to tell if any one person’s data were changed arbitrarily

Differential privacy

[Dinur-Nissim ’03+Dwork, Dwork-Nissim ’04, Blum-Dwork-McSherry-Nissim ’05, Dwork-McSherry-Nissim-Smith ’06]

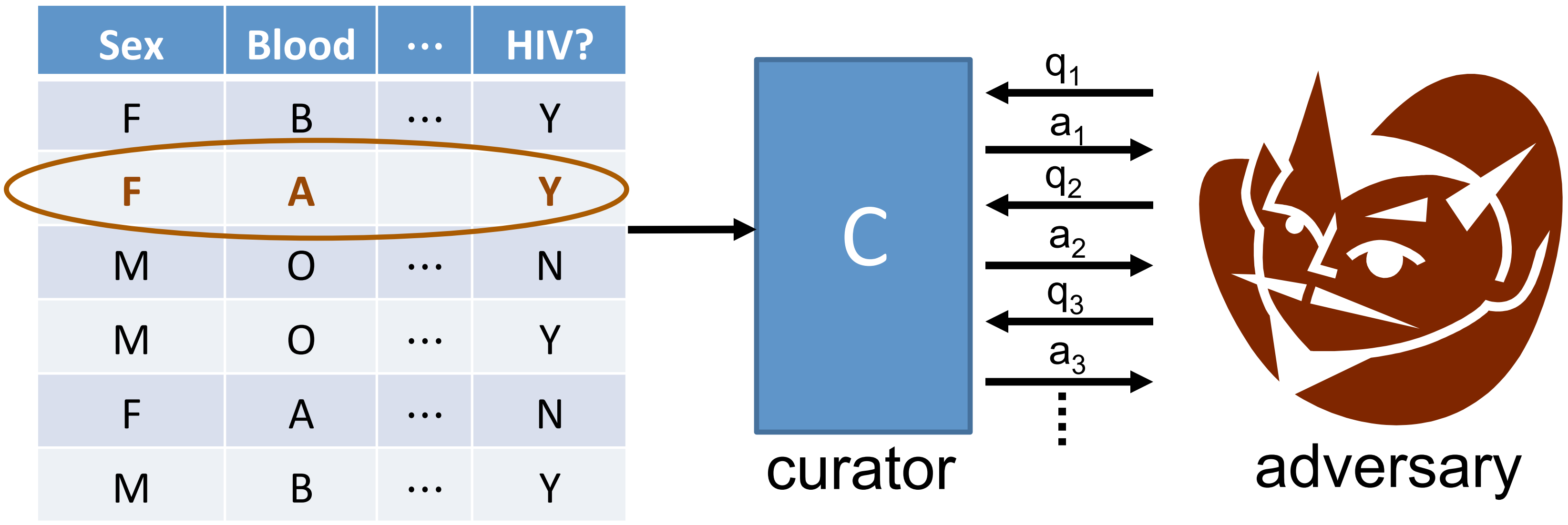


Sex	Blood	⋯	HIV?
F	B	⋯	Y
M	O	⋯	N
M	O	⋯	Y
F	A	⋯	N
M	B	⋯	Y

Requirement: an adversary shouldn’t be able to tell if any one person’s data were changed arbitrarily

Differential privacy

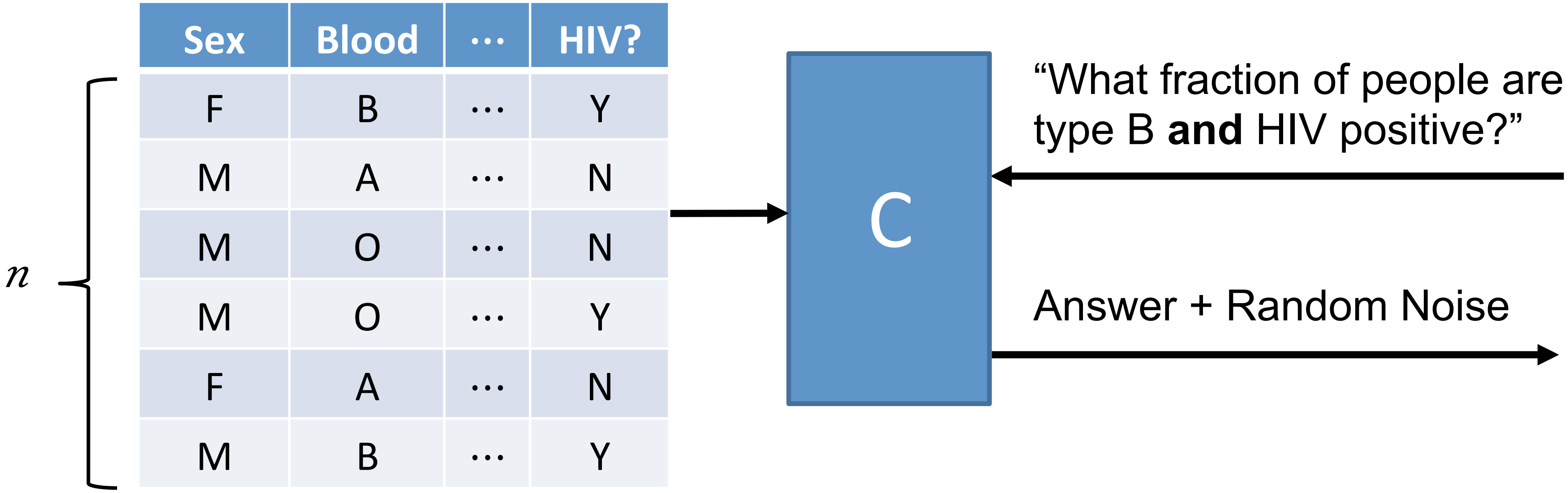
[Dinur-Nissim ’03+Dwork, Dwork-Nissim ’04, Blum-Dwork-McSherry-Nissim ’05, Dwork-McSherry-Nissim-Smith ’06]



Sex	Blood	⋯	HIV?
F	B	⋯	Y
F	**A**		**Y**
M	O	⋯	N
M	O	⋯	Y
F	A	⋯	N
M	B	⋯	Y

Requirement: an adversary shouldn’t be able to tell if any one person’s data were changed arbitrarily

Simple approach: random noise



Sex	Blood	⋯	HIV?
F	B	⋯	Y
M	A	⋯	N
M	O	⋯	N
M	O	⋯	Y
F	A	⋯	N
M	B	⋯	Y

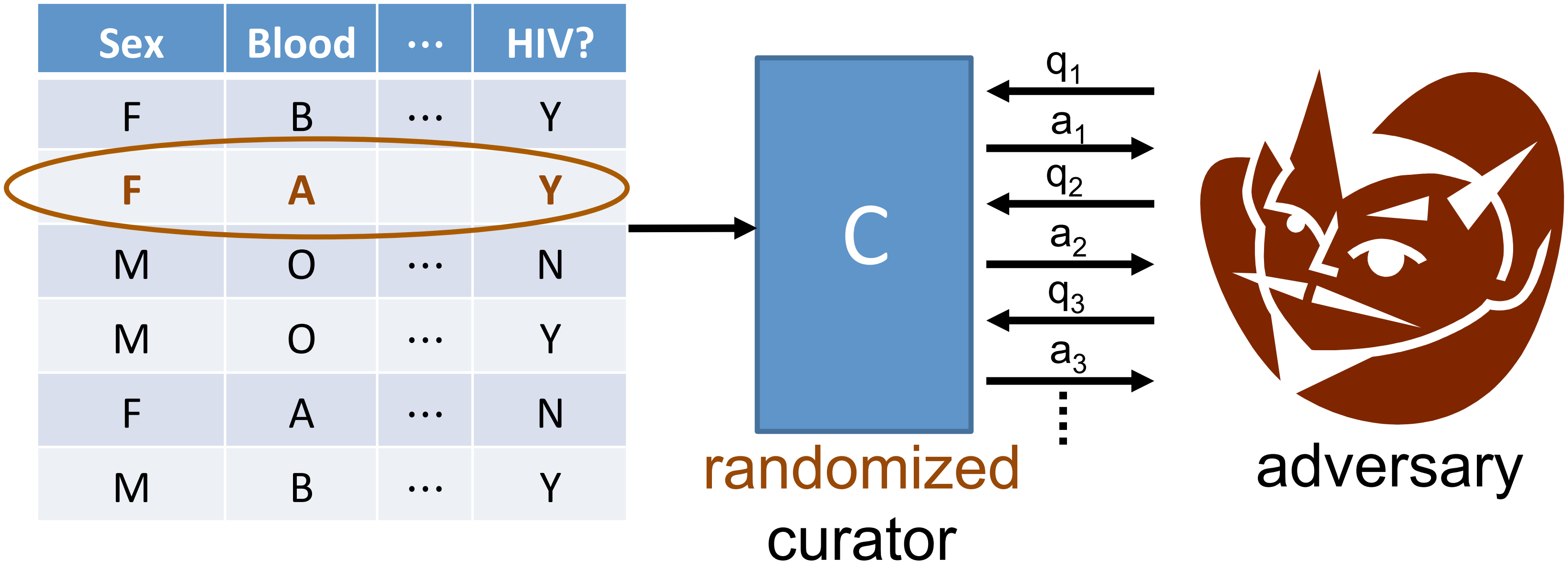
"What fraction of people are type B **and** HIV positive?"

Answer + Random Noise

- Very little noise needed to hide each person as $n \to \infty$.
- Limited to answering $\approx n{\uparrow}2$ queries [Dwork-Naor-Vadhan '12]

Differential privacy

[Dinur-Nissim ’03+Dwork, Dwork-Nissim ’04, Blum-Dwork-McSherry-Nissim ’05, Dwork-McSherry-Nissim-Smith ’06]



Sex	Blood	⋯	HIV?
F	B	⋯	Y
F	A		Y
M	O	⋯	N
M	O	⋯	Y
F	A	⋯	N
M	B	⋯	Y

Requirement: for all D, D’ differing on one row, and all $q_1,\ldots,q_t$

Distribution of $C(D,q_1,\ldots,q_t) \approx_{\varepsilon}$ Distribution of $C(D',q_1,\ldots,q_t)$

Some Differentially Private Algorithms

- histograms [DMNS06]
- contingency tables [BCDKMT07, GHRU11, TUV12, DNT14],
- machine learning [BDMN05,KLNRS08],
- regression & statistical estimation [CMS11,S11,KST11,ST12,JT13]
- clustering [BDMN05,NRS07]
- social network analysis [HLMJ09,GRU11,KRSY11,KNRS13,BBDS13]
- approximation algorithms [GLMRT10]
- singular value decomposition [HR12, HR13, KT13, DTTZ14]
- streaming algorithms [DNRY10,DNPR10,MMNW11]
- mechanism design [MT07,NST10,X11,NOS12,CCKMV12,HK12,KPRU12]
- …

See Simons Institute Workshop on Big Data & Differential Privacy 12/13

Differential Privacy: Interpretations

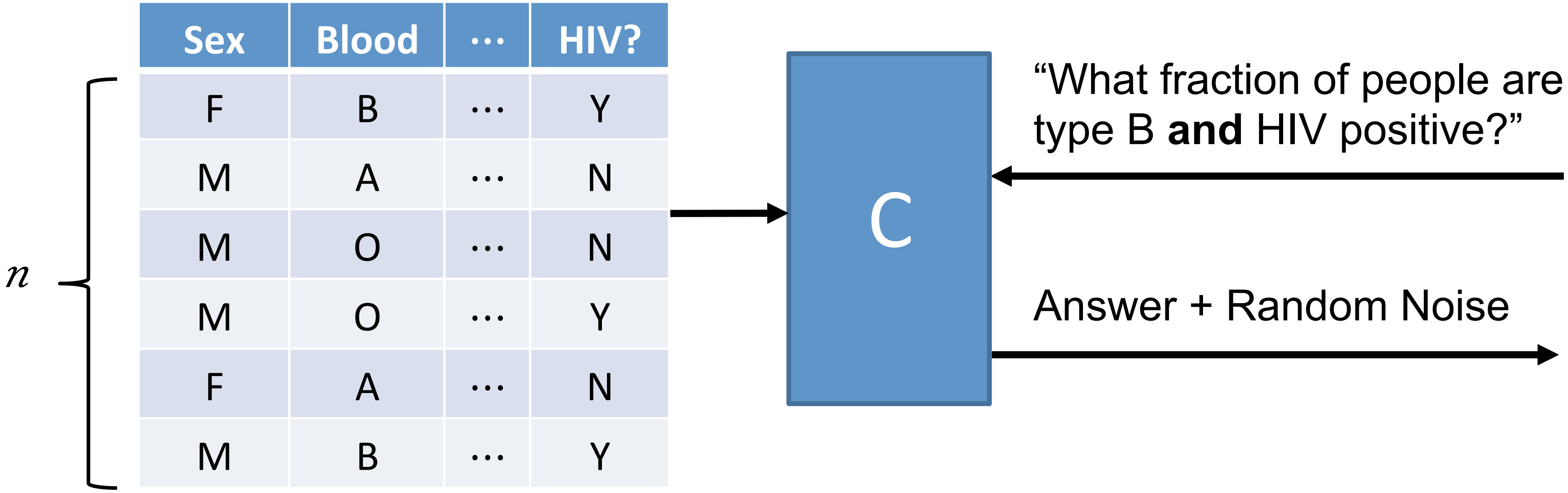
Distribution of $C(D,q_1,\ldots,q_t) \approx \downarrow \varepsilon$ Distribution of $C(D',q_1,\ldots,q_t)$

- Whatever an adversary learns about me, it could have learned from everyone else’s data.
- Mechanism cannot leak “individual-specific” information.
- Above interpretations hold regardless of adversary’s auxiliary information.
- Composes gracefully (k repetitions **)** kε differentially private)

But

- No protection for information that is not localized to a few rows.
- No guarantee that subjects won’t be “harmed” by results of analysis.

Simple approach: random noise



Sex	Blood	⋯	HIV?
F	B	⋯	Y
M	A	⋯	N
M	O	⋯	N
M	O	⋯	Y
F	A	⋯	N
M	B	⋯	Y

n

C

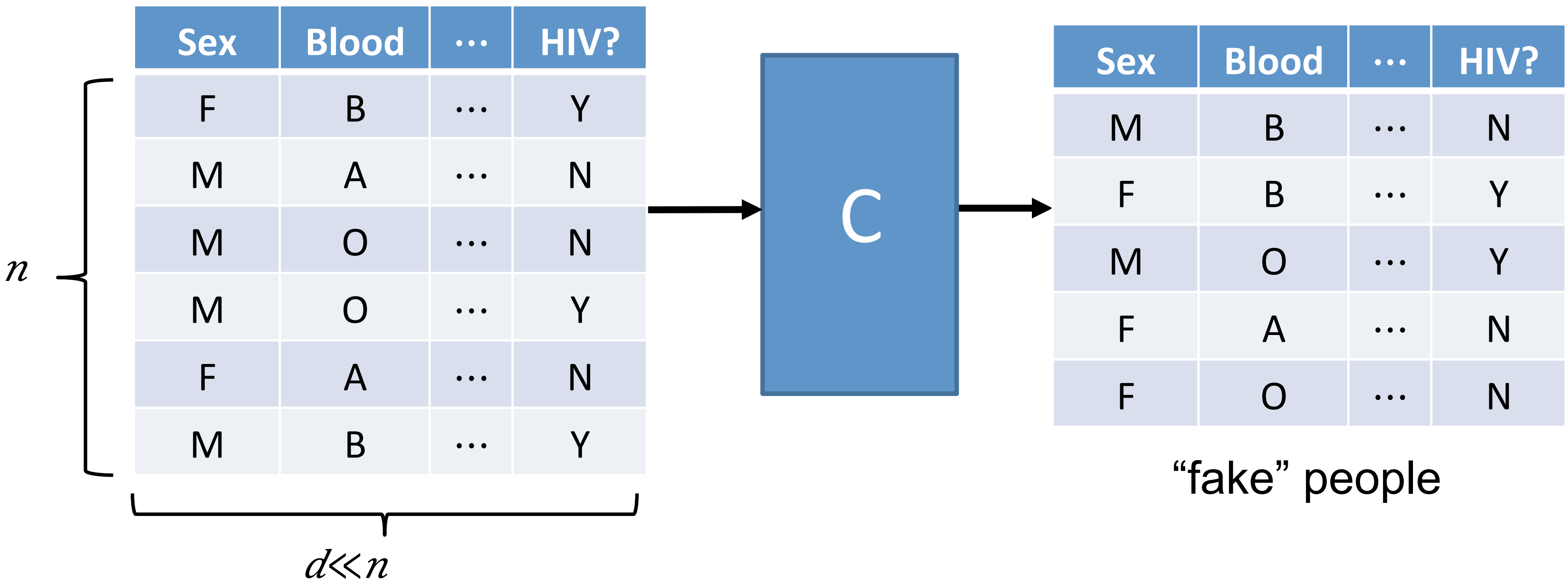
“What fraction of people are type B **and** HIV positive?”

Answer + Random Noise

- Very little noise needed to hide each person as $n \to \infty$.
- Limited to answering $\approx n^2$ queries [Dwork-Naor-Vadhan ’12]

Amazing possibility: synthetic data

[Blum-Ligett-Roth ’08]



Sex	Blood	⋯	HIV?
F	B	⋯	Y
M	A	⋯	N
M	O	⋯	N
M	O	⋯	Y
F	A	⋯	N
M	B	⋯	Y

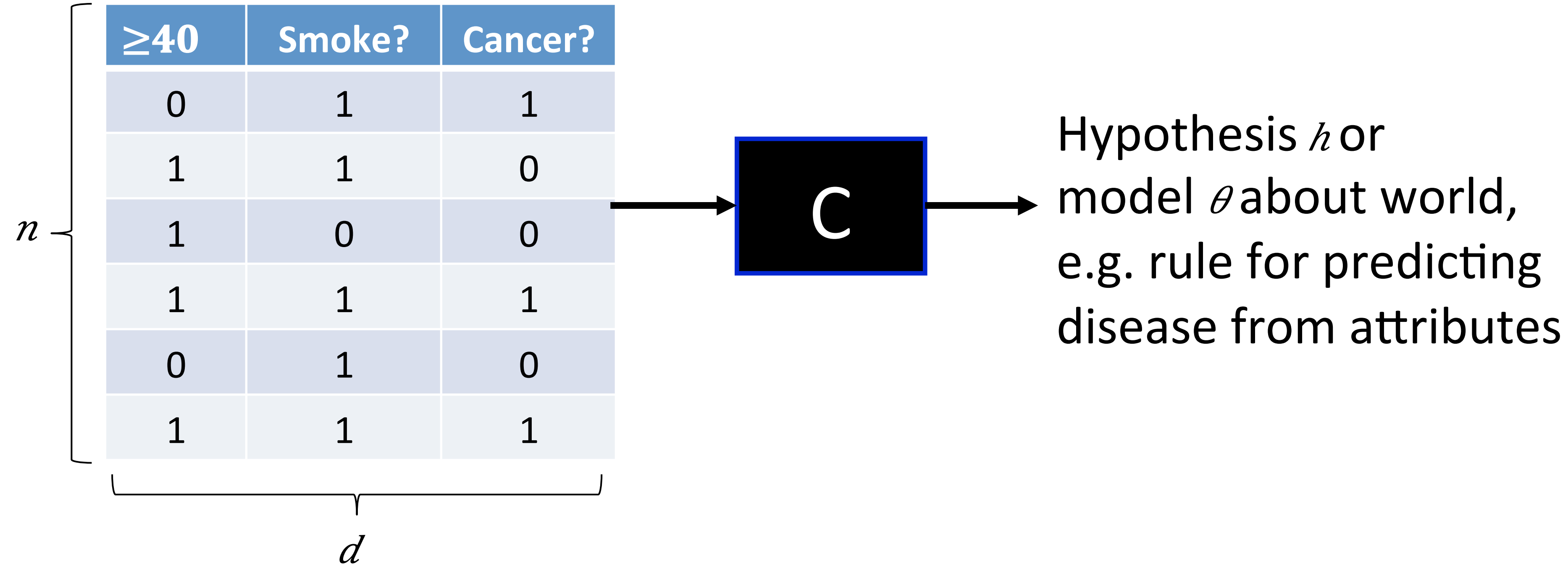
Sex	Blood	⋯	HIV?
M	B	⋯	N
F	B	⋯	Y
M	O	⋯	Y
F	A	⋯	N
F	O	⋯	N

“fake” people

Utility: preserves fraction of people with *every* set of attributes!

Challenge: make this computationally feasible for high-dimensional datasets

Amazing Possibility II:
Statistical Inference & Machine Learning



≥40	Smoke?	Cancer?
0	1	1
1	1	0
1	0	0
1	1	1
0	1	0
1	1	1

Hypothesis h or model θ about world, e.g. rule for predicting disease from attributes

Theorem [KLNRS08,S11]: Differential privacy for vast array of machine learning and statistical estimation problems with little loss in convergence rate as $n \to \infty$.

- Optimizations & practical implementations for logistic regression, ERM, LASSO, SVMs in [RBHT09,CMS11,ST13,JT14].

Challenges for DP in Practice

- Accuracy for “small data” (moderate values of n)

- Modelling & managing privacy loss over time
  - Especially over many different analysts & datasets

- Analysts used to working with raw data
  - One approach: “Tiered access”
  - DP for wide access, raw data only by approval with strict terms of use (cf. Census PUMS vs. RDCs)

- Cases where privacy concerns are not “local” (e.g. privacy for large groups) or utility is not “global” (e.g. targeting)

- Matching guarantees with privacy law & regulation

- …

Some Efforts to Bring DP to Practice

- CMU-Cornell-PennState “Integrating Statistical and Computational Approaches to Privacy”
  - See http://onthemap.ces.census.gov/
- UCSD “Integrating Data for Analysis, Anonymization, and Sharing” (iDash)
- UT Austin “Airavat: Security & Privacy for MapReduce”
- UPenn “Putting Differential Privacy to Work”
- Stanford-Berkeley-Microsoft “Towards Practicing Privacy”
- Duke-NISSS “Triangle Census Research Network”
- Harvard “Privacy Tools for Sharing Research Data”
- …

Privacy tools for sharing research data

http://privacytools.seas.harvard.edu/

NSF

Google

Computer Science, Law, Social Science, Statistics

CRCS Center for Research on Computation and Society

IQSS

The Institute for Quantitative Social Science

HARVARD UNIVERSITY

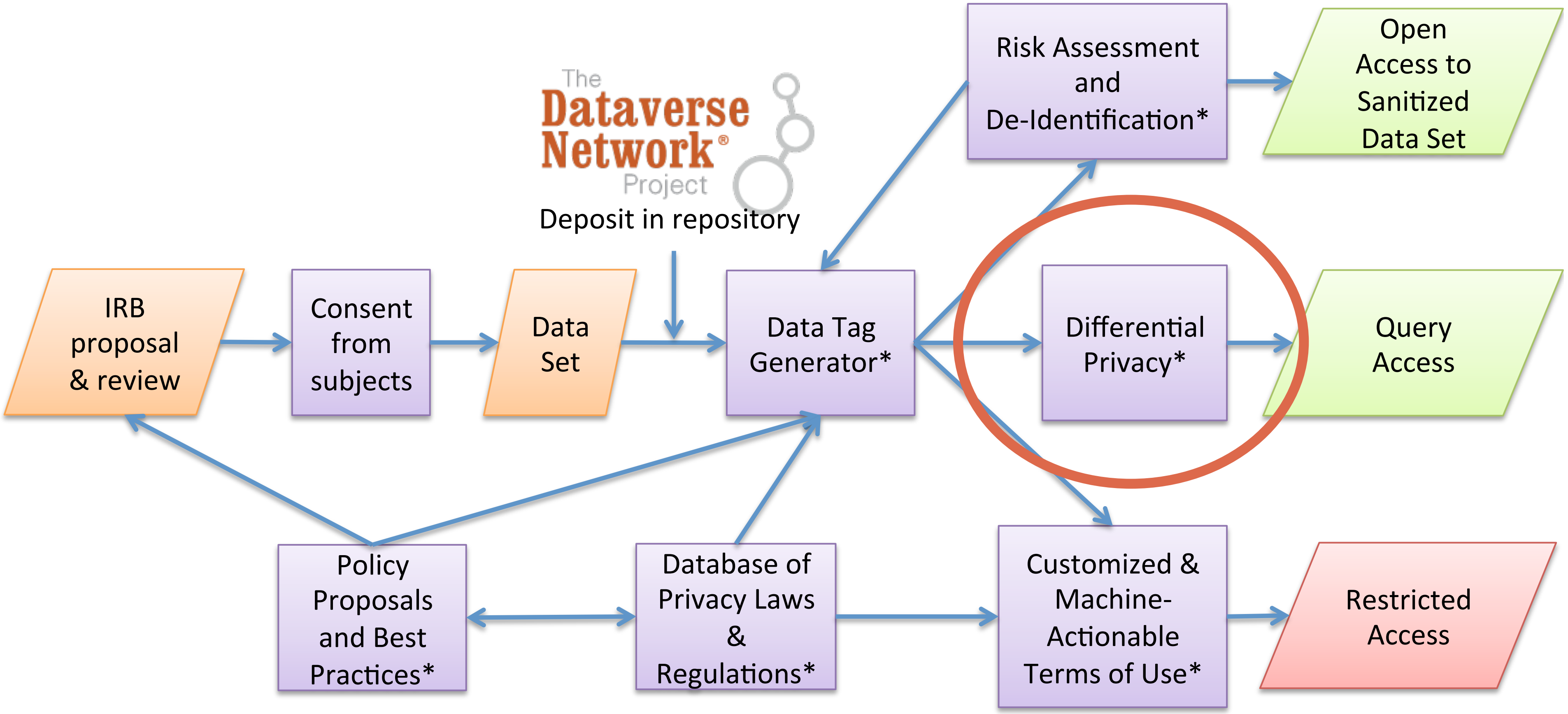
VE RI TAS

Berkman

The Berkman Center for Internet & Society at Harvard University

DATA PRIVACY LAB

Integrated Privacy Tools



The Dataverse Network® Project

Deposit in repository

IRB proposal & review

Consent from subjects

Data Set

Data Tag Generator*

Risk Assessment and De-Identification*

Differential Privacy*

Customized & Machine-Actionable Terms of Use*

Open Access to Sanitized Data Set

Query Access

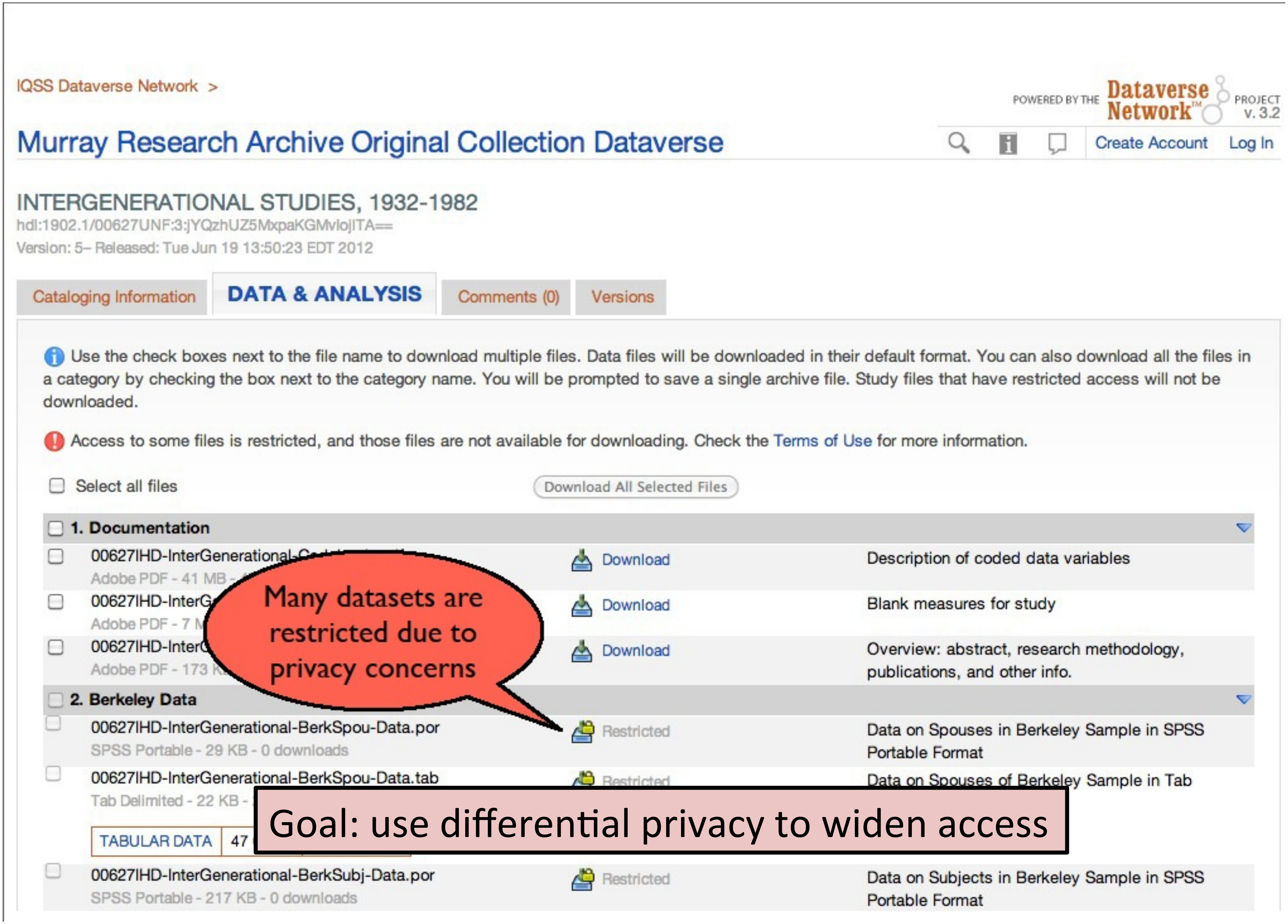
Restricted Access

Policy Proposals and Best Practices*

Database of Privacy Laws & Regulations*

Tools to be developed during project

* = Tools directly contributed to by Year 1 activities



IQSS Dataverse Network >

POWERED BY THE Dataverse Network™ PROJECT v. 3.2

Murray Research Archive Original Collection Dataverse

INTERGENERATIONAL STUDIES, 1932-1982

hdl:1902.1/00627UNF:3:jYQzhUZ5MxpaKGMvlojITA==

Version: 5– Released: Tue Jun 19 13:50:23 EDT 2012

Use the check boxes next to the file name to download multiple files. Data files will be downloaded in their default format. You can also download all the files in a category by checking the box next to the category name. You will be prompted to save a single archive file. Study files that have restricted access will not be downloaded.

Access to some files is restricted, and those files are not available for downloading. Check the Terms of Use for more information.

□ Select all files

□ 1. Documentation

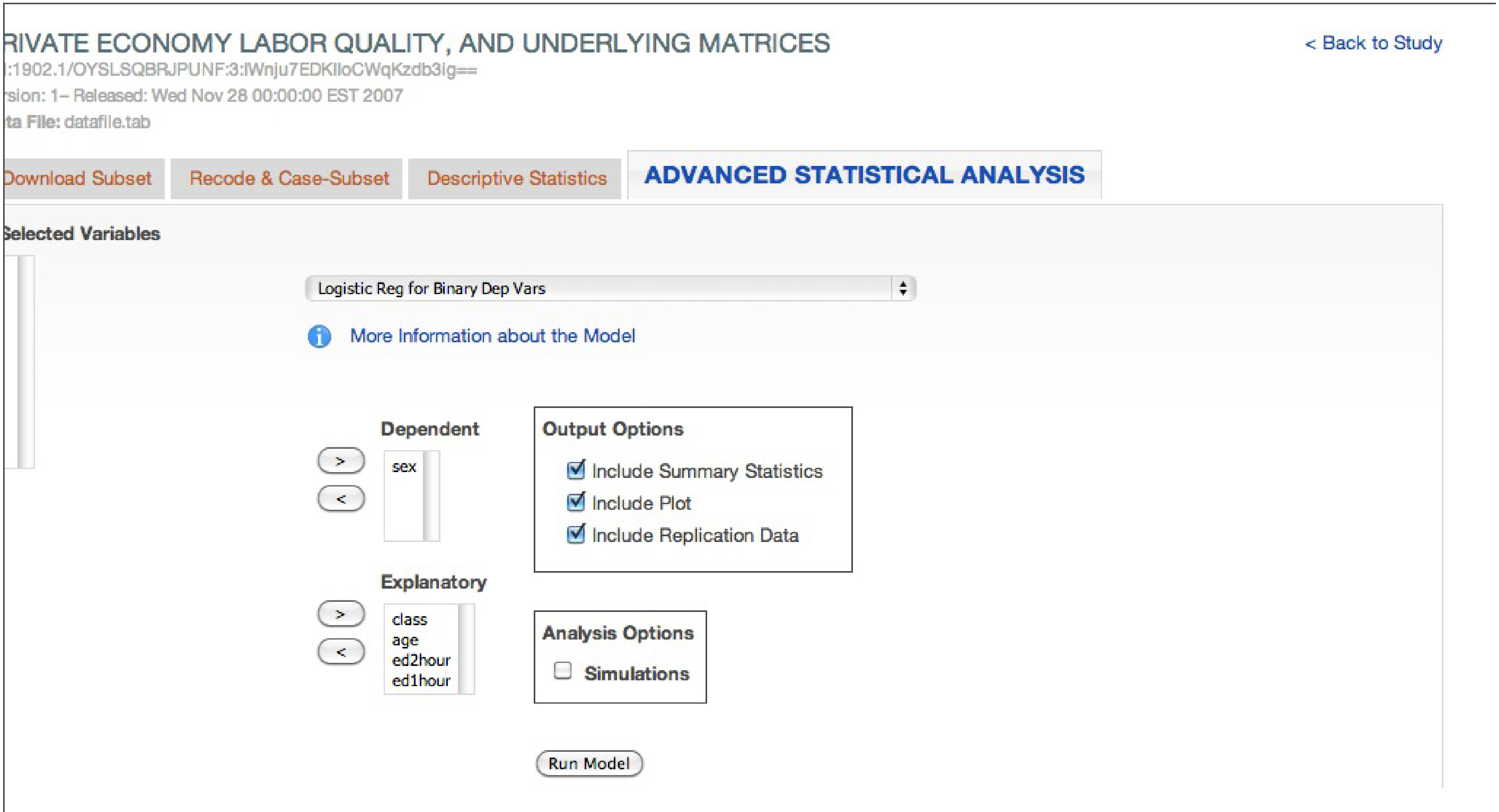
□	00627IHD-InterGenerational-C Adobe PDF - 41 MB -	Download	Description of coded data variables
□	00627IHD-InterG Adobe PDF - 7 M	Download	Blank measures for study
□	00627IHD-Inter Adobe PDF - 173 K	Download	Overview: abstract, research methodology, publications, and other info.

□ 2. Berkeley Data

□	00627IHD-InterGenerational-BerkSpou-Data.por SPSS Portable - 29 KB - 0 downloads	Restricted	Data on Spouses in Berkeley Sample in SPSS Portable Format
□	00627IHD-InterGenerational-BerkSpou-Data.tab Tab Delimited - 22 KB - TABULAR DATA 47	Restricted	Data on Spouses of Berkeley Sample in Tab
□	00627IHD-InterGenerational-BerkSubj-Data.por SPSS Portable - 217 KB - 0 downloads	Restricted	Data on Subjects in Berkeley Sample in SPSS Portable Format

Many datasets are restricted due to privacy concerns

Goal: use differential privacy to widen access



RIVATE ECONOMY LABOR QUALITY, AND UNDERLYING MATRICES

:1902.1/OYSLSQBRJPUNF:3:lWnju7EDKlloCWqKzdb3lg==

rsion: 1– Released: Wed Nov 28 00:00:00 EST 2007

ta File: datafile.tab

Logistic Reg for Binary Dep Vars

Dependent

sex

Explanatory

class
age
ed2hour
ed1hour

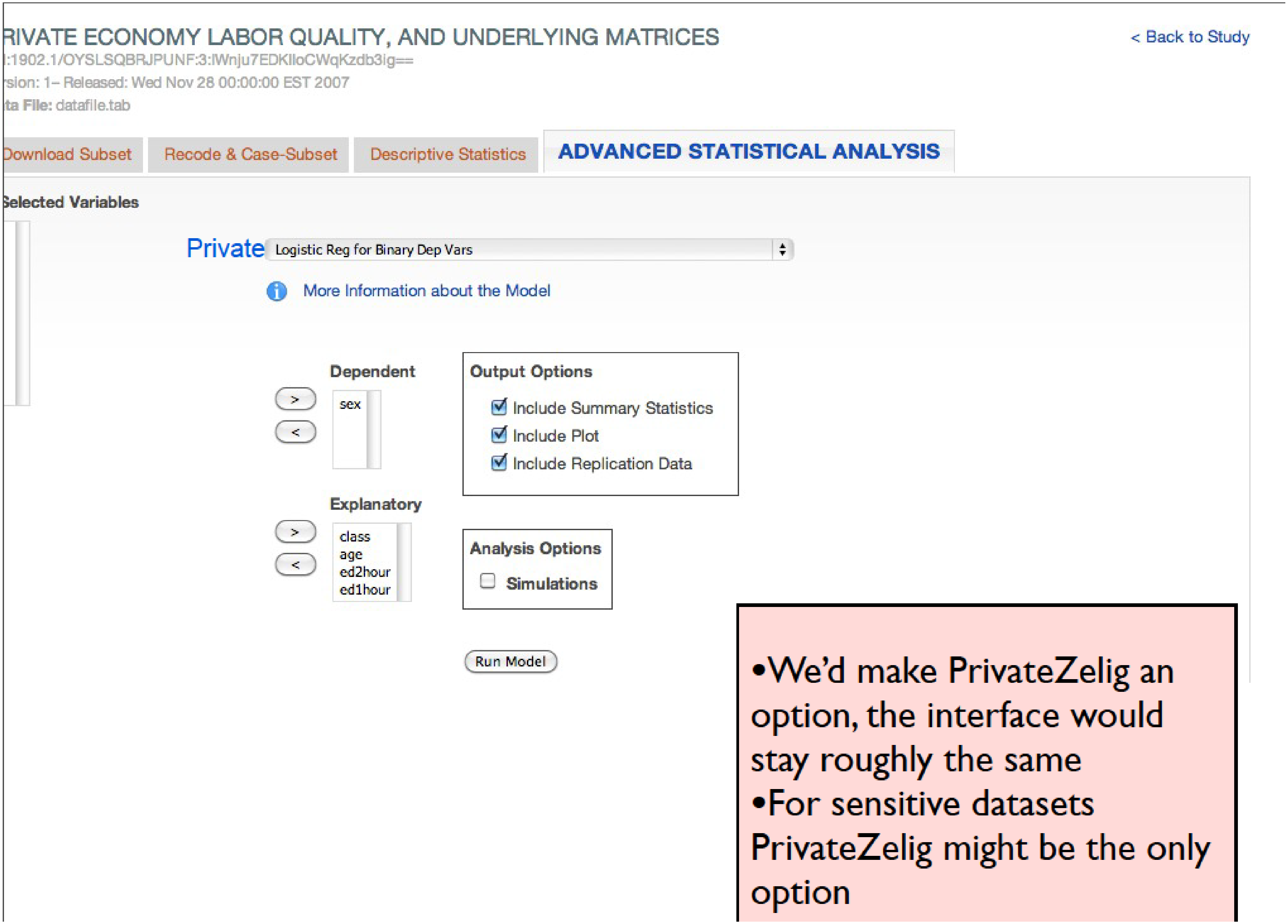
Output Options

☑ Include Summary Statistics
☑ Include Plot
☑ Include Replication Data

Analysis Options

☐ **Simulations**

For non-restricted datasets, can run many statistical analyses ("Zelig methods") through the Dataverse interface, without downloading data.



RIVATE ECONOMY LABOR QUALITY, AND UNDERLYING MATRICES

:1902.1/OYSLSQBRJPUNF:3:lWnju7EDKlloCWqKzdb3ig==

rsion: 1– Released: Wed Nov 28 00:00:00 EST 2007

ta File: datafile.tab

Private: Logistic Reg for Binary Dep Vars

Dependent: sex

Explanatory: class, age, ed2hour, ed1hour

Output Options

☑ Include Summary Statistics

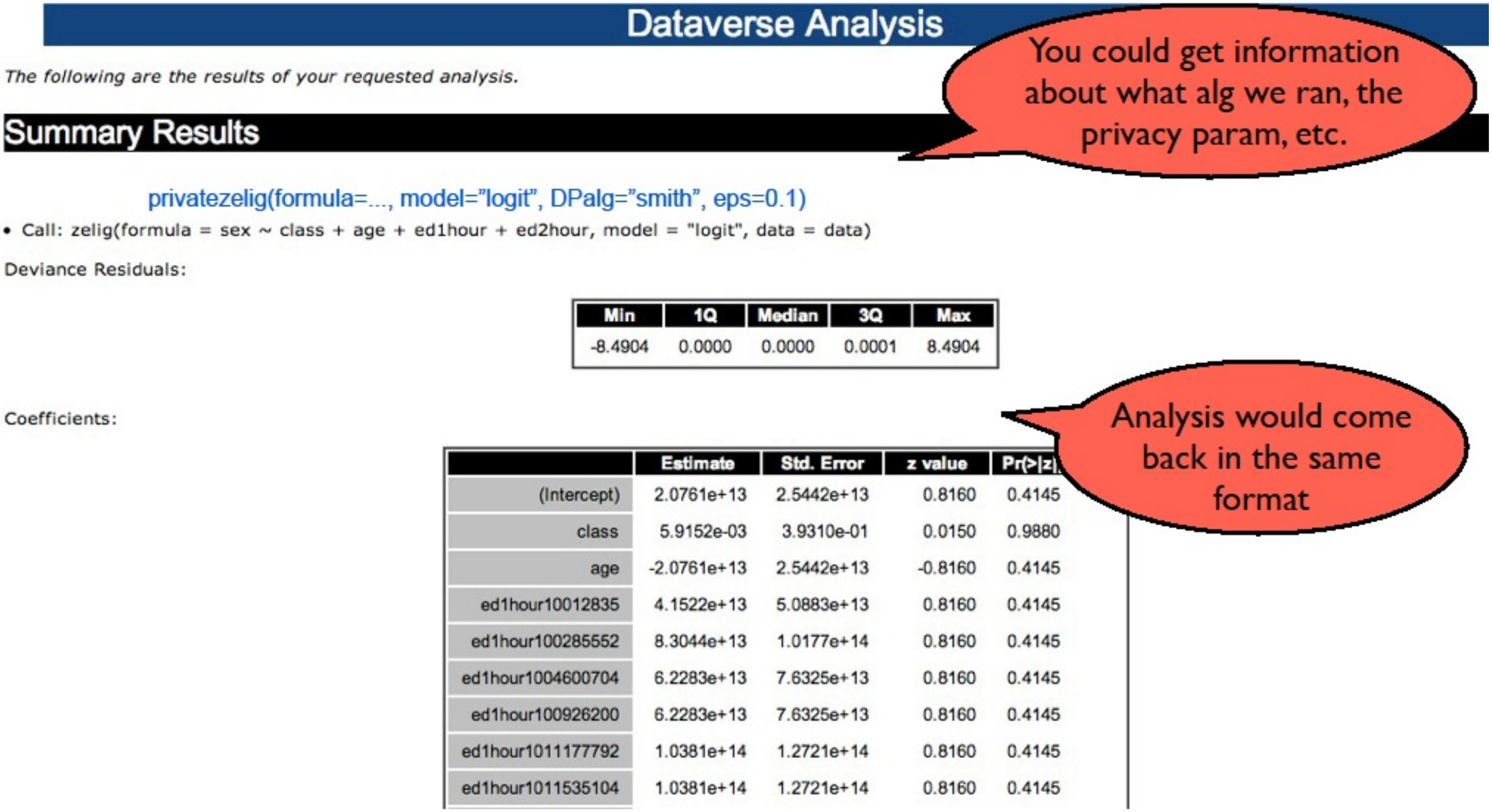
☑ Include Plot

☑ Include Replication Data

Analysis Options

☐ Simulations

- We'd make PrivateZelig an option, the interface would stay roughly the same
- For sensitive datasets PrivateZelig might be the only option



Dataverse Analysis

You could get information about what alg we ran, the privacy param, etc.

The following are the results of your requested analysis.

Summary Results

privatezelig(formula=..., model="logit", DPalg="smith", eps=0.1)

- Call: zelig(formula = sex ~ class + age + ed1hour + ed2hour, model = "logit", data = data)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-8.4904	0.0000	0.0000	0.0001	8.4904

Coefficients:

Analysis would come back in the same format

	Estimate	Std. Error	z value	Pr(>|z|
(Intercept)	2.0761e+13	2.5442e+13	0.8160	0.4145
class	5.9152e-03	3.9310e-01	0.0150	0.9880
age	-2.0761e+13	2.5442e+13	-0.8160	0.4145
ed1hour10012835	4.1522e+13	5.0883e+13	0.8160	0.4145
ed1hour100285552	8.3044e+13	1.0177e+14	0.8160	0.4145
ed1hour1004600704	6.2283e+13	7.6325e+13	0.8160	0.4145
ed1hour100926200	6.2283e+13	7.6325e+13	0.8160	0.4145
ed1hour1011177792	1.0381e+14	1.2721e+14	0.8160	0.4145
ed1hour1011535104	1.0381e+14	1.2721e+14	0.8160	0.4145

Our Implementation Goals

This summer: differentially private summary statistics

- Means, quantiles, histograms, (co)variances/PCA
- Computed at time of dataset deposit
- Depositor decides how to allocate “privacy budget”
- Enough to support interactive least-squares regressions

Future: interactive and/or more sophisticated statistics

- Synthetic data
- Contingency tables
- Other regressions
- Interactive queries

Privacy Models from CS

Model	Utility	Privacy	Who Holds Data?
Differential Privacy	statistical analysis of dataset	individual-specific info	trusted curator
Secure Function Evaluation	any query desired	everything other than result of query	original users (or semi-trusted delegates)
Fully Homomorphic (or Functional) Encryption	any query desired	everything (except possibly result of query)	untrusted server

For other two topics, see Shafi Goldwasser's talk
at White House-MIT Big Data Privacy Workshop 3/3/14

Differential Privacy: Summary

Differential Privacy offers

- Strong, scalable privacy guarantees
- Compatibility with many types of “big data” analyses
- Amazing possibilities for what can be achieved in principle

There are some challenges, but reasons for optimism

- Intensive research effort from many communities
- Some successful uses in practice already
- Differential privacy easier as $n \to \infty$

Schedule for Tomorrow (in MD323)

- 12-12:30 Lunch
- 12:30-1:30 Introduction to R (Vito)
- 1:30-2:00 Software Engineering, R, Zelig (James)
- 2:30-2:45 Break
- 2:45-4:15 More Differential Privacy (Sofya)

Future Weeks:

- every Mon 1:30-2:30: all-hands meeting
- 2x/week TBD: more tutorials & research mtgs on differential privacy, R, and statistics
- TBD: project-wide social activities (a hike?)