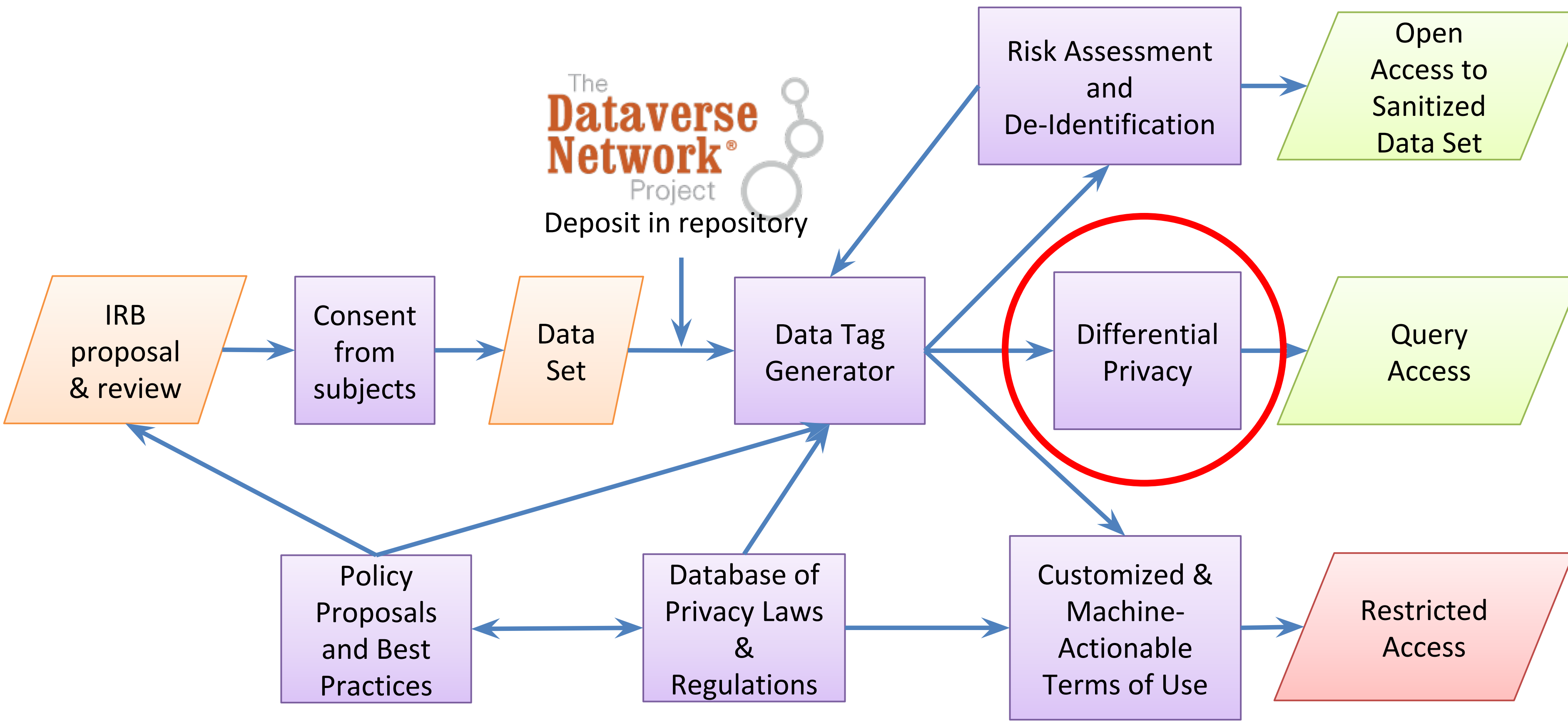
PRIVACY BUDGETING INTERFACE

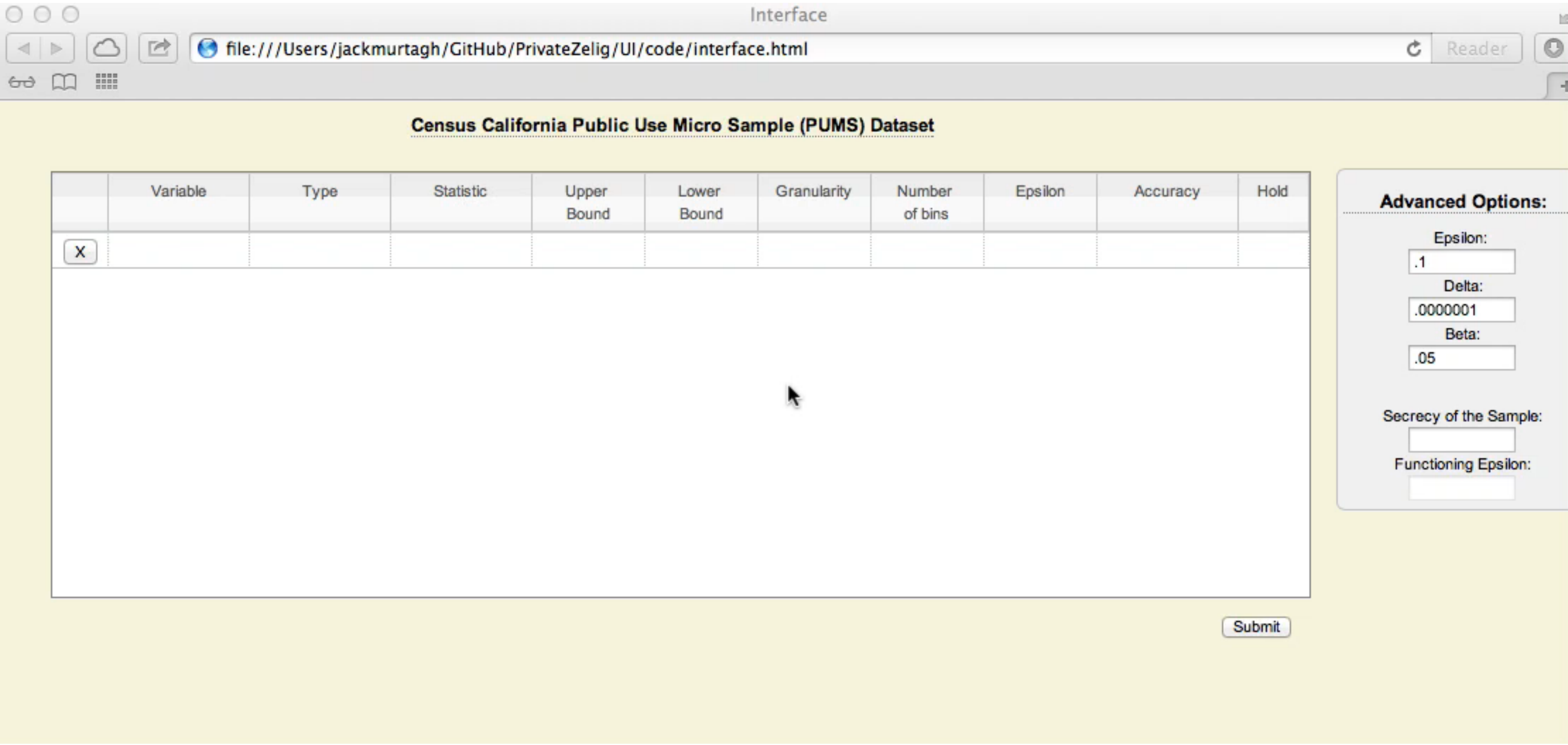


The Dataverse Network® Project
Deposit in repository
IRB proposal & review
Consent from subjects
Data Set
Data Tag Generator
Risk Assessment and De-Identification
Open Access to Sanitized Data Set
Differential Privacy
Query Access
Policy Proposals and Best Practices
Database of Privacy Laws & Regulations
Customized & Machine-Actionable Terms of Use
Restricted Access

Interface

- Select Statistics (mean, quantiles, histogram)
- Provide metadata
  - Range
  - Granularity
  - Number of bins
- Control apportioning of privacy budget

Demo


Interface
file:///Users/jackmurtagh/GitHub/PrivateZelig/UI/code/interface.html
Reader
Census California Public Use Micro Sample (PUMS) Dataset
Variable
Type
Statistic
Upper Bound
Lower Bound
Granularity
Number of bins
Epsilon
Accuracy
Hold
X
Advanced Options:
Epsilon:
.1
Delta:
.0000001
Beta:
.05
Secrecy of the Sample:
Functioning Epsilon:
Submit

API

- Statistic.getAccuracy(eps, del, metadata, beta)
  - Returns: Upper bound on distance from truth with probability 1-beta
- Statistic.getParameter(acc, del, metadata, beta)
  - Returns: Minimum epsilon required to achieve accuracy of acc with probability 1-beta
- Statistic.Compute(eps, del, data, metadata)
  - Returns: (eps, del)-DP statistic on data

Goals

- More advanced DP algorithms
- Interactivity
- Utility testing
- User testing
- Many more…

Composition

Theorem 0.1 (Basic Composition). *For every $\epsilon \geq 0$, $\delta \in [0,1]$, and (ϵ, δ)-differentially private algorithms $M_1, \ldots, M_k$, the k-fold composition $(M_1, M_2, \ldots, M_k)$ satisfies $(k\epsilon, k\delta)$-differential privacy.*

Theorem 0.2 (Advanced Composition [DRV10]). *For every $\epsilon > 0, \delta, \delta' > 0$, $k \in \mathbb{N}$, and (ϵ, δ)-differentially private algorithms $M_1, \ldots, M_k$, the k-fold composition $(M_1, M_2, \ldots, M_k)$ satisfies $(\epsilon_g, k\delta + \delta')$-differential privacy for*

$$\epsilon_g = \sqrt{2k \ln(1/\delta')} \cdot \epsilon + k \cdot \epsilon(e^\epsilon - 1)$$

Secrecy of the Sample

M: $X^m \rightarrow R$ is (ε, δ)-DP

M': $X^n \rightarrow R$ where $n > m$

M'(x):

1. Construct x' composed of a random sample of m rows of x
2. Output M(x')

Privacy guarantee on M':

$$\left(\ln\left(1 + \frac{(e^{\varepsilon} - 1)m}{n}\right), \frac{\delta m}{n}\right)$$