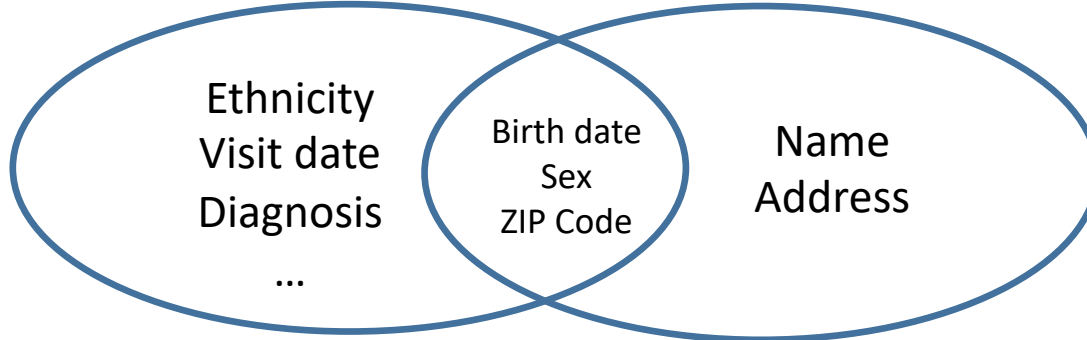




There is privacy risk in de-identified data — what can we do about it?

Sweeney, 2000

Field Suppression, Swapping, Generalization



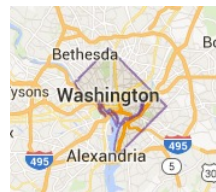
Medical Data

Voter List

HIPAA “Safe Harbor” standard.
k-anonymity, *ℓ*-diversity, *t*-closeness
Differential Privacy

Race	Birthdate	Sex	Zip	Med
Black	9/20/85	M	37203	
Black	2/14/85	M	37203	
Black	10/23/85	F	37215	
Black	8/24/85	F	37215	
Black	11/7/84	F	37215	
Black	12/1/84	F	37215	
White	10/23/84	M	37215	
White	3/15/84	F	37217	
White	8/13/84	M	37217	
White	5/5/84	M	37217	
White	2/13/87	M	37215	
White	3/21/87	M	37215	

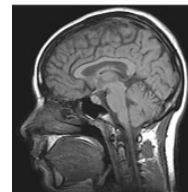
Tabular



Geospatial



Social Media



Medical



DNA



Bodycams/Surveillance

Today:

How do we balance
Utility vs. Risk?

How does risk
change over time?

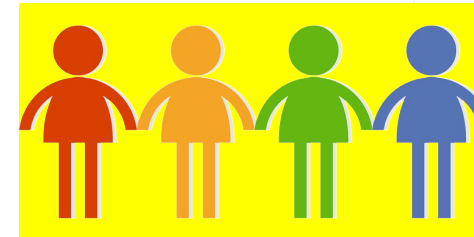
New data types
to de-identify.



NYC Taxi drivers
Re-identified



Bikeshare riders re-identified?



Community harms
without re-identification

Options moving forward:

- Better de-identification
- Synthetic data
- Data use agreements
- Privacy remediation

NIST FY2016 GOALS:

- Develop guidance
- Gov De-ID Stakeholder meeting
- Pilot de-identification evaluation

Simson L. Garfinkel, Ph.D., Senior Advisor, Information Access Division, simson.garfinkel@nist.gov

DISCLAIMER: Specific products and organizations identified in this report were used in order to perform the evaluations described. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that identified are necessarily the best available for the purpose.