OpenDP

An open-source suite of tools for deploying differential privacy

Machine Learning in Science & Engineering 2020

December 14-15, Columbia University Data Science Institute

Mercè Crosas, Ph.D., Harvard University

scholar.harvard.edu/mercecrosas @mercecrosas

IQSS
The Institute for Quantitative Social Science

HARVARD
UNIVERSITY

What is Differential Privacy?

- A **differentially private (DP)** algorithm introduces a minimum amount of noise to released statistics to mathematically guarantee the privacy of any individual in a dataset
- Aims to [Dwork, McSherry, Nissim, Smith, '06]:
  - enable statistical analysis of datasets - **utility**
  - while protecting individual level data – **privacy**
- In the last years, DP has moved from theory to practice and starts to be deployed in products

What is OpenDP?

- OpenDP is a **community effort** to build a **trustworthy** and **open-source** suite of **differential privacy tools** to explore and analyze sensitive data
- Aims to:
  - Channel **collective advances** on the science and practice of DP
  - Enable wider **adoption** of DP
  - Address high-demand, compelling **use cases**
  - Provide the building blocks for deploying **custom DP solutions**
  - Identify important **research directions** for the field

The OpenDP Team

Executive Committee

Salil Vadhan
Faculty Co-Director

Gary King
Faculty Co-Director

Annie Wu
Program Director

James Honaker
Chief Privacy Engineer

Mercè Crosas
Chief Data Science & Technology Officer, IQSS

Development & Staff

Ellen Kraffmiller
Technical Lead

Ethan Cowan
Software Developer

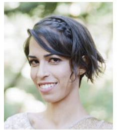
Christian Covington
Research Fellow

Ira Globus-Harris
Software Engineer

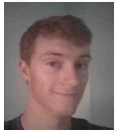
Jayshree Sarathy
Graduate Student

Lindsay Froess
Project Coordinator

Raman Prasad
Technical Lead for Research Software



Audrey Haque
UX Designer & Researcher



Michael Shoemate
Senior Software Developer

Robert Treacy
Senior Application Architect

Danny Brooke
Program Manager, Product Development, IQSS

Gustavo Durand
Dataverse Technical Lead and Architect

Tania Schlatter
UX & UI Lead, Product Development

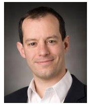
Michael Phelan
Senior Application Architect

Andy Vyrros
Senior Library Architect

Advisory Board

OpenDP

Gerome Miklau
(co-chair)
UMass, Amherst


Adam Smith
(co-chair)
Boston University

John Abowd
US Census Bureau,
Cornell University

Gilles Barthe
Max Planck Institute,
IMDEA Software
Institute

Barbara Bierer
Brigham & Women's Hospital
Harvard Medical School

Sarah Bird
Microsoft

danah boyd
Microsoft Research

Cynthia Dwork
Harvard University
Radcliffe Institute
Microsoft Research

Úlfar Erlingsson
Apple

John Friedman
Brown University

Jeff Gill
American University

Daniel Goroff
Alfred P. Sloan
Foundation
(ex officio)

Frauke Kreuter
University of Mannheim
Institute for Employment
Research, Germany

Orran Krieger
PI Mass Open Cloud
Boston University

David Lazer
Northeastern University

Katrina Ligett
Hebrew University

Margaret
Levenstein
University of Michigan

Carlos Maltzahn
UC Santa Cruz
CROSS

Kenneth Mandl
Harvard Medical School
Boston Children's Hospital

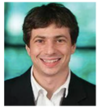
Ilya Mironov
Facebook

Helen Nissenbaum
Cornell University

Kobbi Nissim
Georgetown University

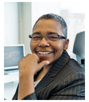
Dina N. Paltoo
National Heart, Lung
and Blood Institute

Jules Polonetsky
Future of Privacy Forum

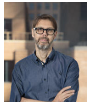

Aaron Roth
University of
Pennsylvania

Aleksandra Slavkovic
Pennsylvania State
University

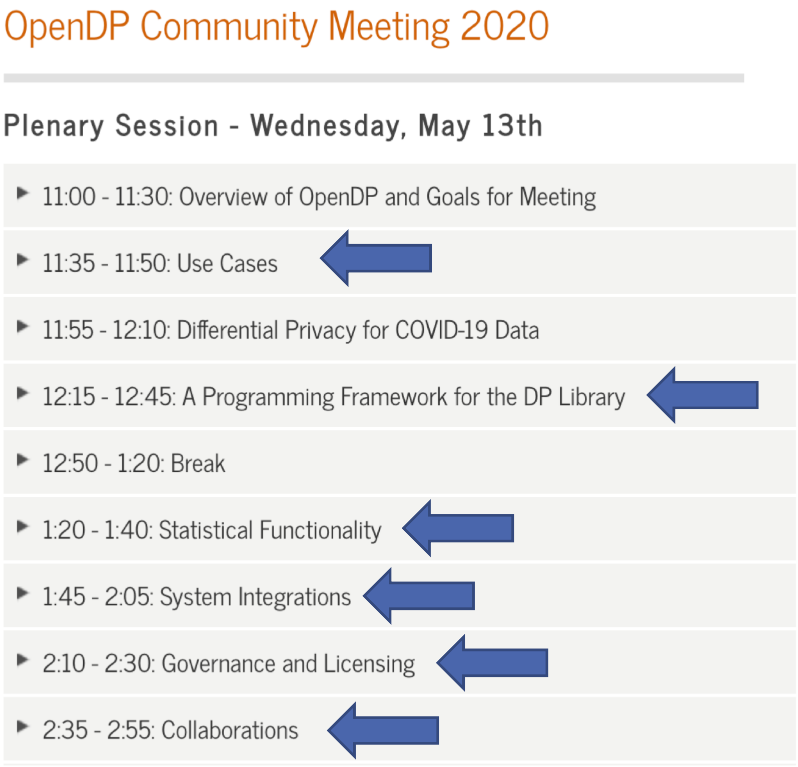
Dawn Song
University of California,
Berkeley


Latanya Sweeney
Harvard University

Omer Tene
International
Association of Privacy
Professionals


Stefaan G Verhulst
New York University

The OpenDP initiative launched in 2020



OpenDP Community Meeting 2020

Plenary Session - Wednesday, May 13th

- 11:00 - 11:30: Overview of OpenDP and Goals for Meeting
- 11:35 - 11:50: Use Cases
- 11:55 - 12:10: Differential Privacy for COVID-19 Data
- 12:15 - 12:45: A Programming Framework for the DP Library
- 12:50 - 1:20: Break
- 1:20 - 1:40: Statistical Functionality
- 1:45 - 2:05: System Integrations
- 2:10 - 2:30: Governance and Licensing
- 2:35 - 2:55: Collaborations

- **OpenDP Community Meeting in May 2020**
- **Key elements in whitepaper:** use cases, governance, programming framework, statistical functionality, system integrations, collaboration & community
- Previous work with the **Harvard Privacy Tools Project**, funded by NSF, US Census Bureau, the Sloan Foundation, and Google
- Current grants from the **Sloan Foundation**
- Current collaboration with **Microsoft**

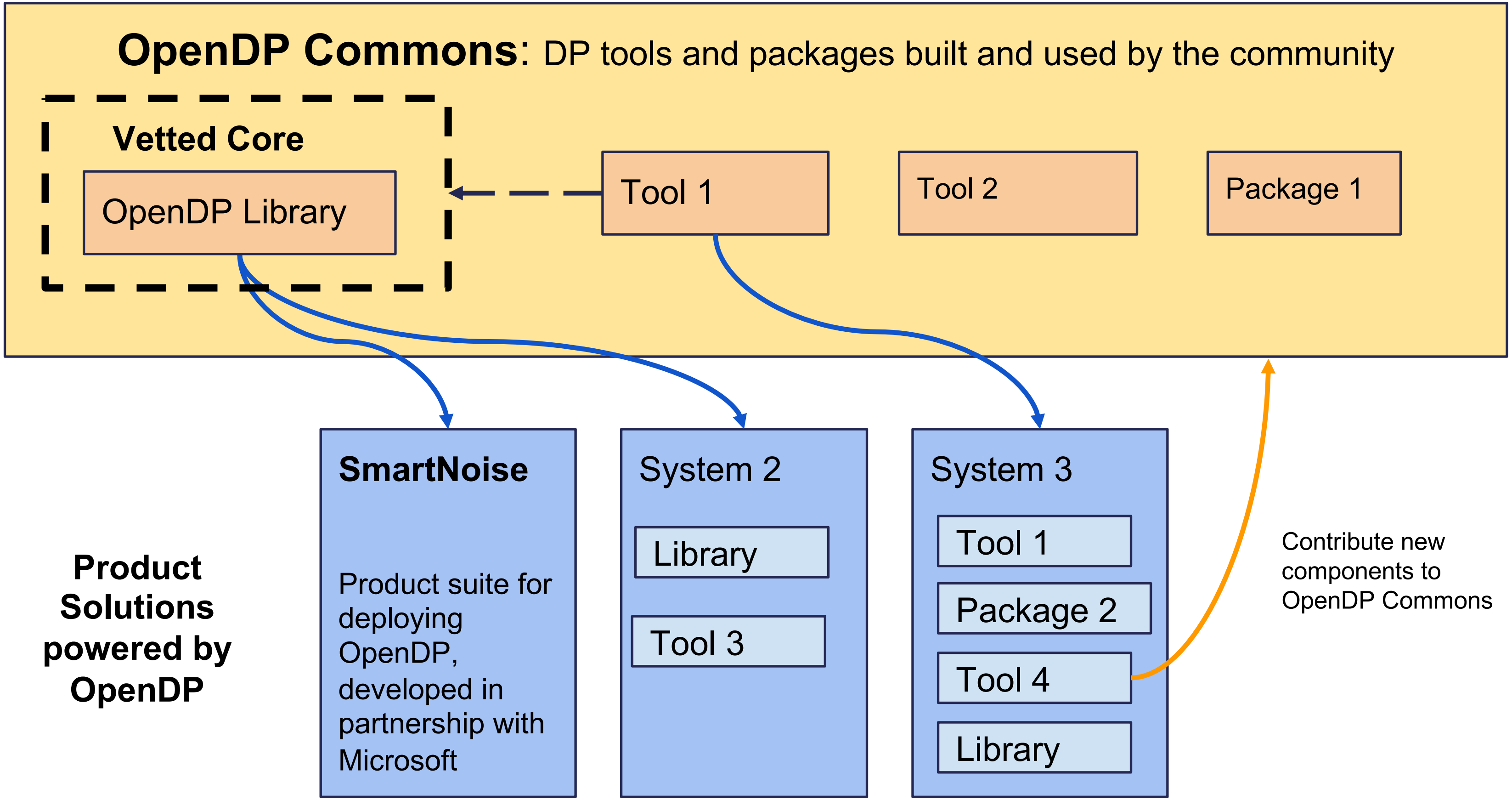
Alfred P. Sloan FOUNDATION

Microsoft

NSF

Google

Open-source Governance and Components



OpenDP Commons: DP tools and packages built and used by the community

Vetted Core

OpenDP Library

Tool 1

Tool 2

Package 1

Product Solutions powered by OpenDP

SmartNoise

Product suite for deploying OpenDP, developed in partnership with Microsoft

System 2

Library

Tool 3

System 3

Tool 1

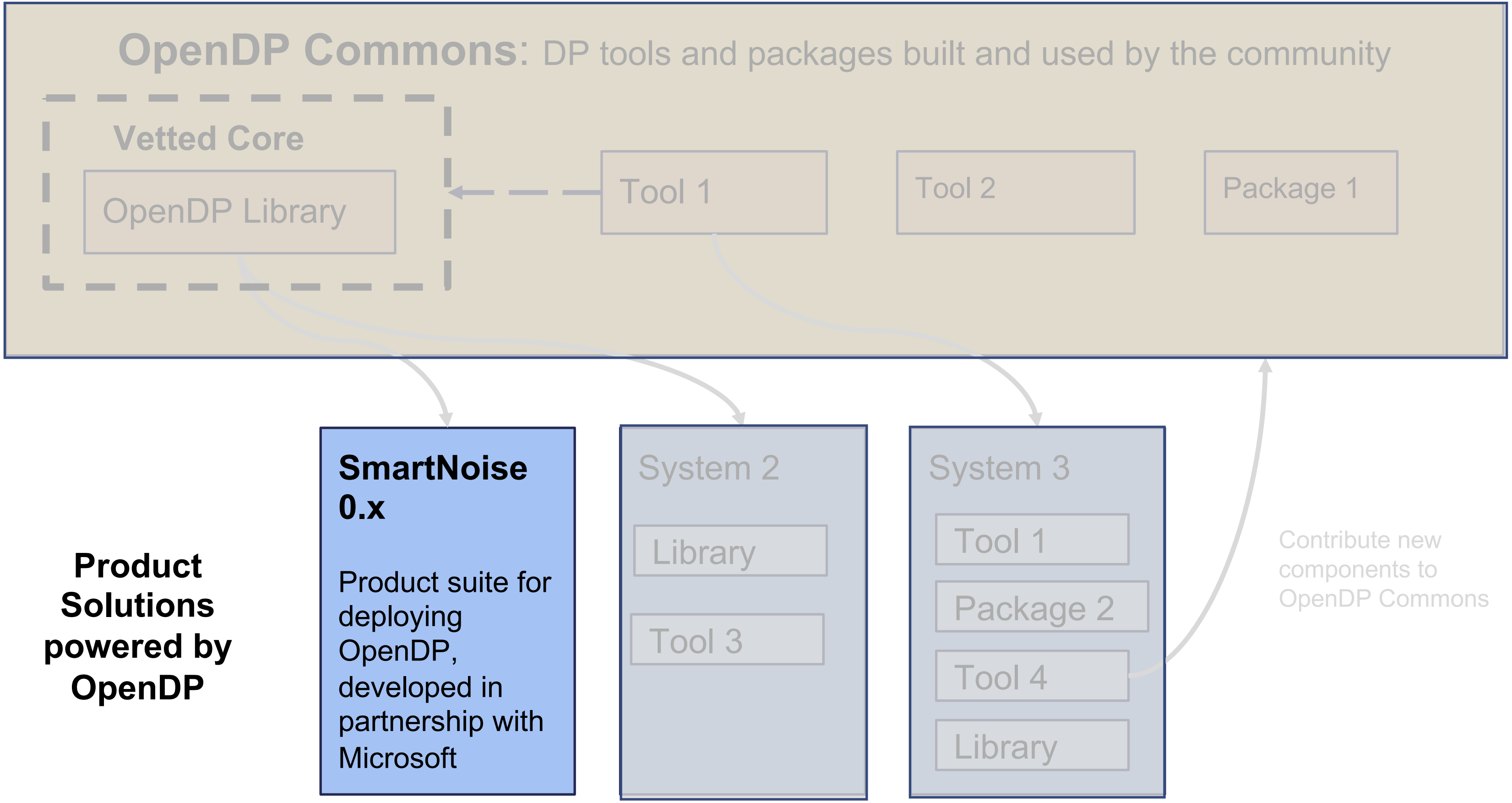
Package 2

Tool 4

Library

Contribute new components to OpenDP Commons

Open-source Governance and Components



OpenDP Commons: DP tools and packages built and used by the community

Vetted Core

OpenDP Library

Tool 1

Tool 2

Package 1

Product Solutions powered by OpenDP

SmartNoise 0.x

Product suite for deploying OpenDP, developed in partnership with Microsoft

System 2

Library

Tool 3

System 3

Tool 1

Package 2

Tool 4

Library

Contribute new components to OpenDP Commons

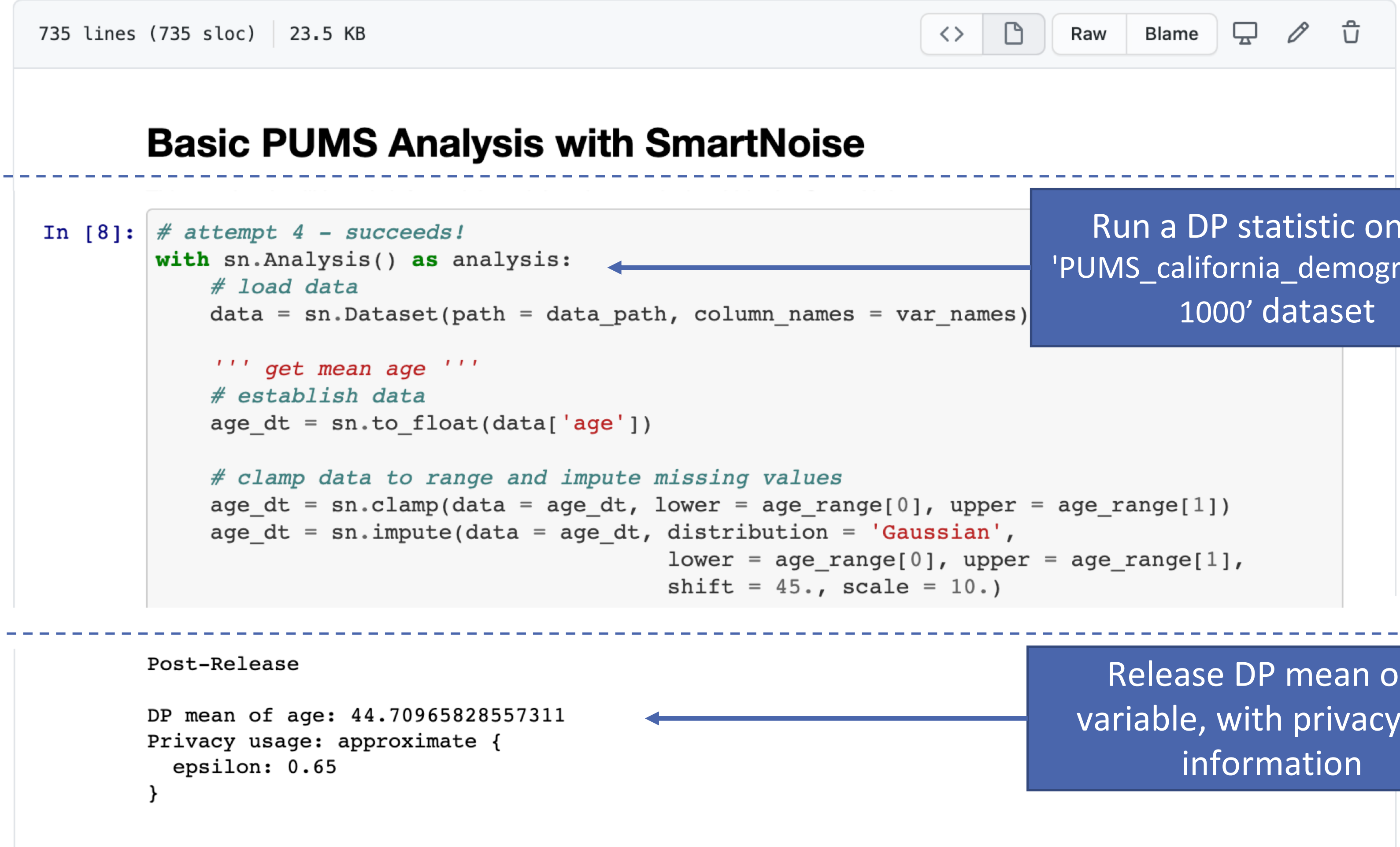
SmartNoise 0.x: First end-to-end OpenDP system

Use Data.
Preserve Privacy.
A differential privacy toolkit for analytics and machine learning

- **Released in June 2020**
- A collaboration with Microsoft: **Sarah Bird, John Kahan, Joshua Allen, Eddie de Leon, Kevin White**
- An OpenDP end-to-end system **proof-of-concept**
- **SmartNoise Core v0.2.0:**
  - Rust Library and Python bindings
  - Statistics: mean, variance, count, histogram, general quantiles, sum, covariance, simple linear regression
  - Mechanisms: Laplace, Gaussian, snapping, geometric, exponential
- **SmartNoise Tools/SDK v0.1.0**

Microsoft IQSS The Institute for Quantitative Social Science Harvard John A. Paulson School of Engineering and Applied Sciences

SmartNoise Core 0.2.0



Basic PUMS Analysis with SmartNoise

```
In [8]: # attempt 4 - succeeds!
        with sn.Analysis() as analysis:
            # load data
            data = sn.Dataset(path = data_path, column_names = var_names)

            ''' get mean age '''
            # establish data
            age_dt = sn.to_float(data['age'])

            # clamp data to range and impute missing values
            age_dt = sn.clamp(data = age_dt, lower = age_range[0], upper = age_range[1])
            age_dt = sn.impute(data = age_dt, distribution = 'Gaussian',
                                              lower = age_range[0], upper = age_range[1],
                                              shift = 45., scale = 10.)
```

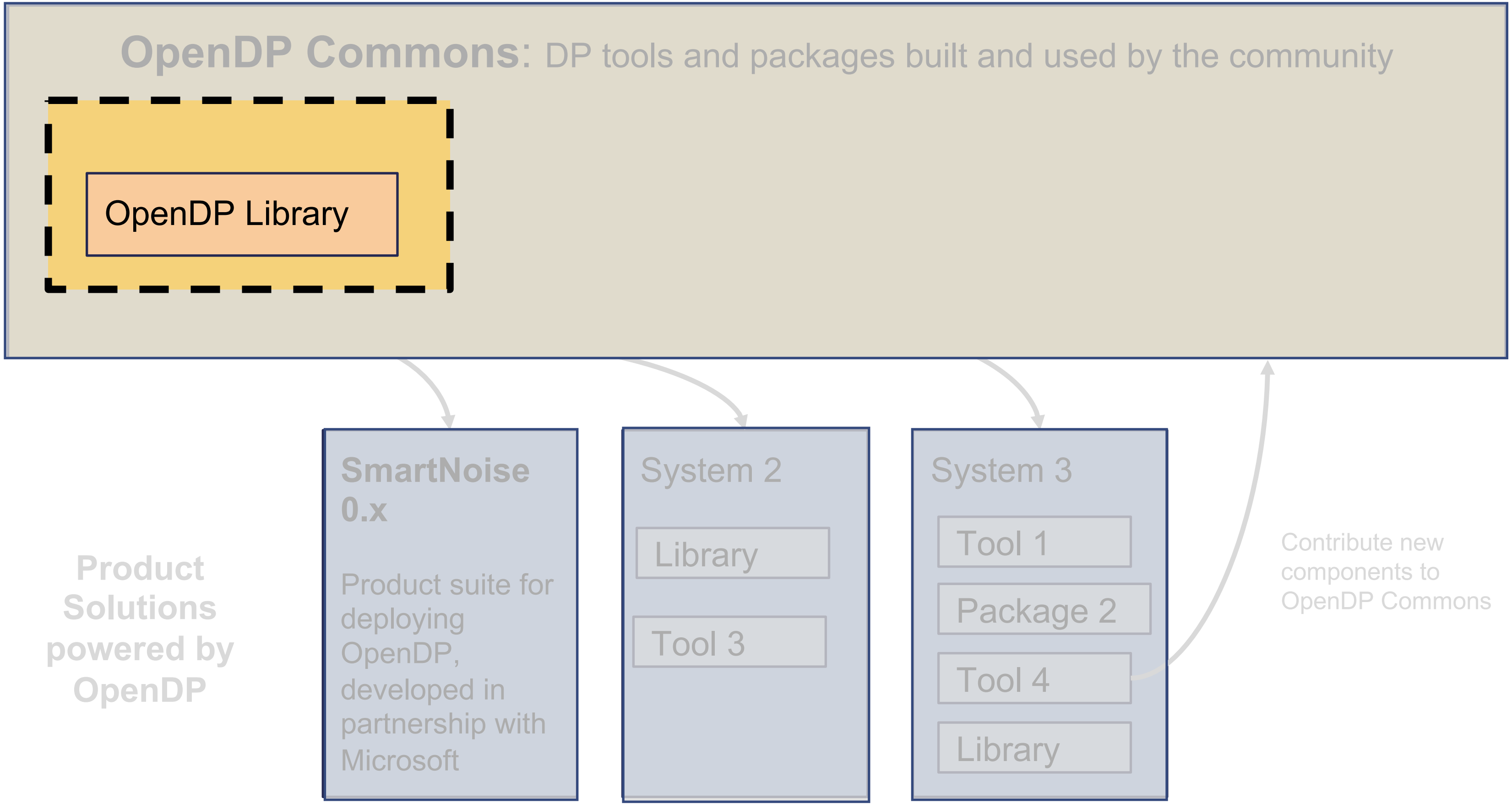
Run a DP statistic on the 'PUMS_california_demographics_1000' dataset

```
Post-Release

DP mean of age: 44.70965828557311
Privacy usage: approximate {
  epsilon: 0.65
}
```

Release DP mean of a variable, with privacy loss information

Open-source Governance and Components



OpenDP Commons: DP tools and packages built and used by the community

OpenDP Library

Product Solutions powered by OpenDP

SmartNoise 0.x

Product suite for deploying OpenDP, developed in partnership with Microsoft

System 2

Library

Tool 3

System 3

Tool 1

Package 2

Tool 4

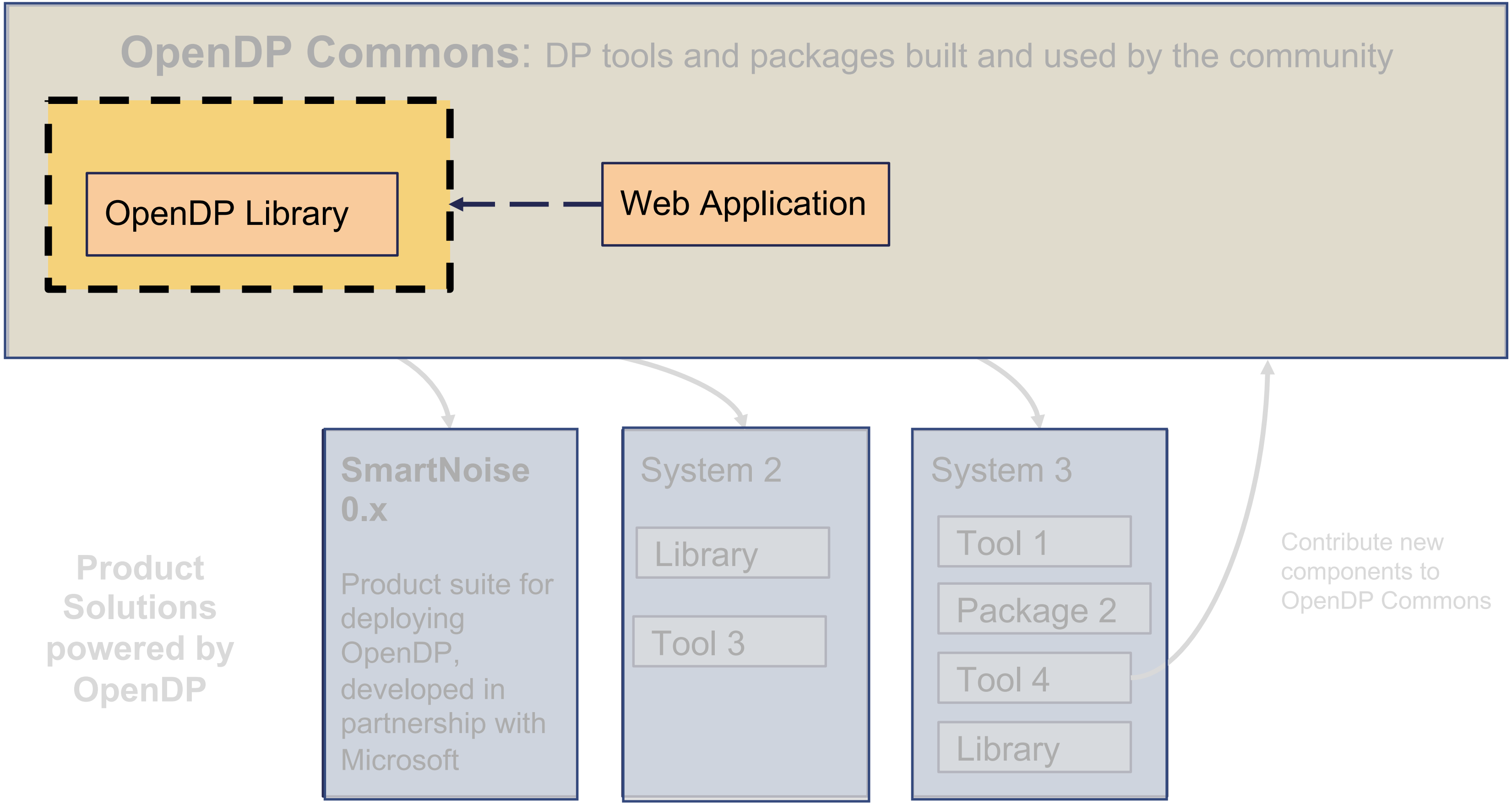
Library

Contribute new components to OpenDP Commons

OpenDP Library

- A **novel programming framework** [Marco Gaboardi, Michael Hay, Salil Vadhan]
- Goes beyond simply solving a fixed set of DP models
- **Abstracts** the treatment of privacy mechanisms
- Allows building of **new, complex privacy** systems
- Facilitates and verifies **contributions**
- **First release to the community: Early 2021**

Open-source Governance and Components



OpenDP Commons: DP tools and packages built and used by the community

OpenDP Library

Web Application

SmartNoise 0.x

Product suite for deploying OpenDP, developed in partnership with Microsoft

System 2

Library

Tool 3

System 3

Tool 1

Package 2

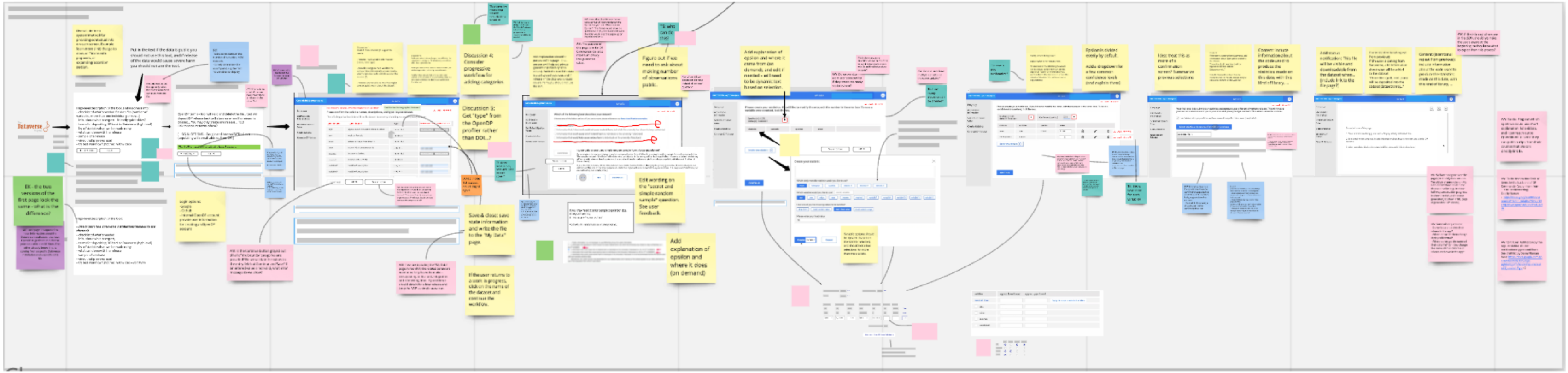
Tool 4

Library

Product Solutions powered by OpenDP

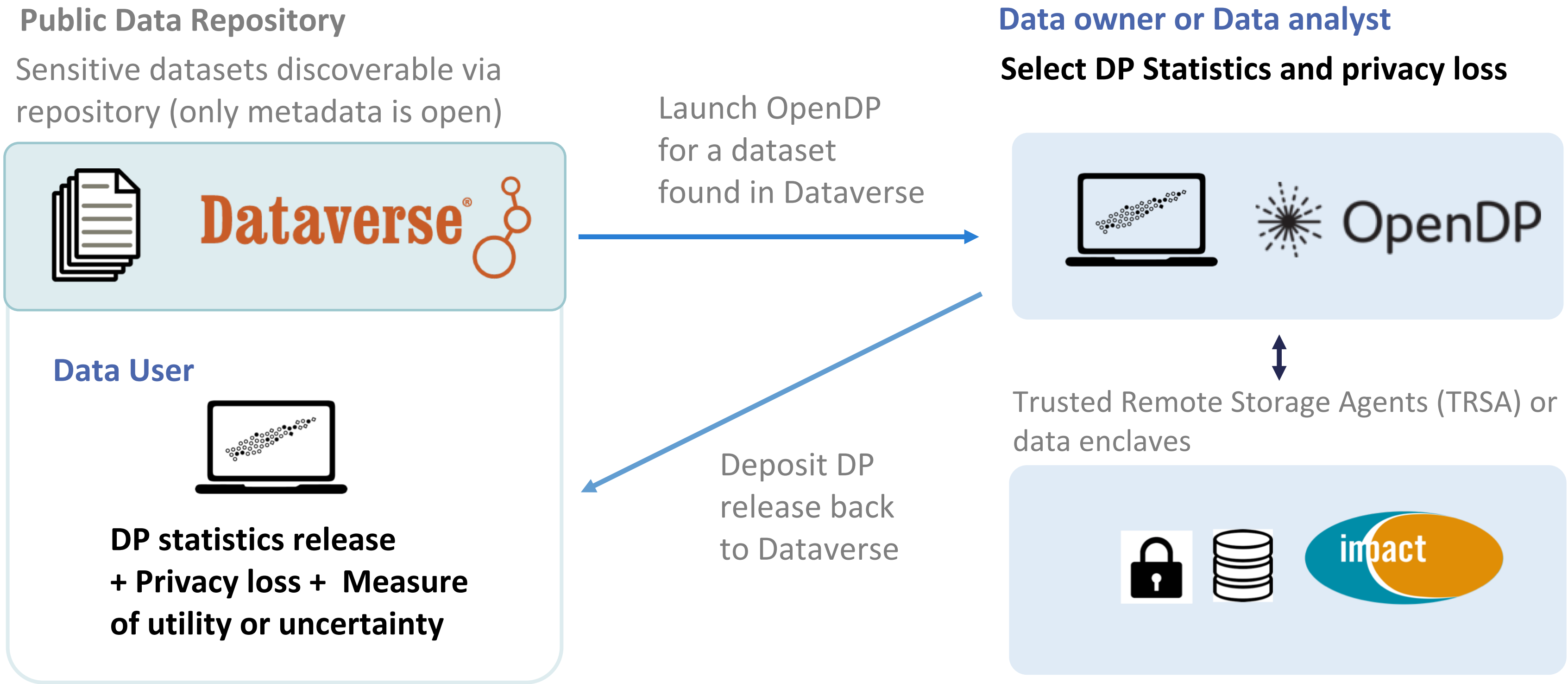
Contribute new components to OpenDP Commons

OpenDP Web Application


Dataverse
EK - the two versions of the first page look the same - what is the difference?
Discussion 4: Consider progressive workflow for adding categories
Discussion 5: Get "type" from the OpenDP profiler rather than DDI...?
TS: who can do this?
Figure out if we need to ask about making number of observations public.
Edit wording on the "secret and simple random sample" question. See user feedback.
Save & close: save state information and write the file to the "My Data" page.
If the user returns to a work in progress, click on the name of the dataset and continue the workflow.
Add explanation of epsilon and where it does (on demand)

- Uses the OpenDP library and focuses on UI/UX for **user-friendly, wide adoption**
- First **use case** aims to support archival **data repositories**:
  - Enables secondary reuse of data and reproducibility of published results
  - Integrates with > 60 **Dataverse repositories** world-wide
- DP released statistics must expose measures of **utility and uncertainty**
- **First release to the community: Early 2021**

OpenDP + Dataverse



Public Data Repository

Sensitive datasets discoverable via repository (only metadata is open)

Launch OpenDP for a dataset found in Dataverse

Data owner or Data analyst

Select DP Statistics and privacy loss

Trusted Remote Storage Agents (TRSA) or data enclaves

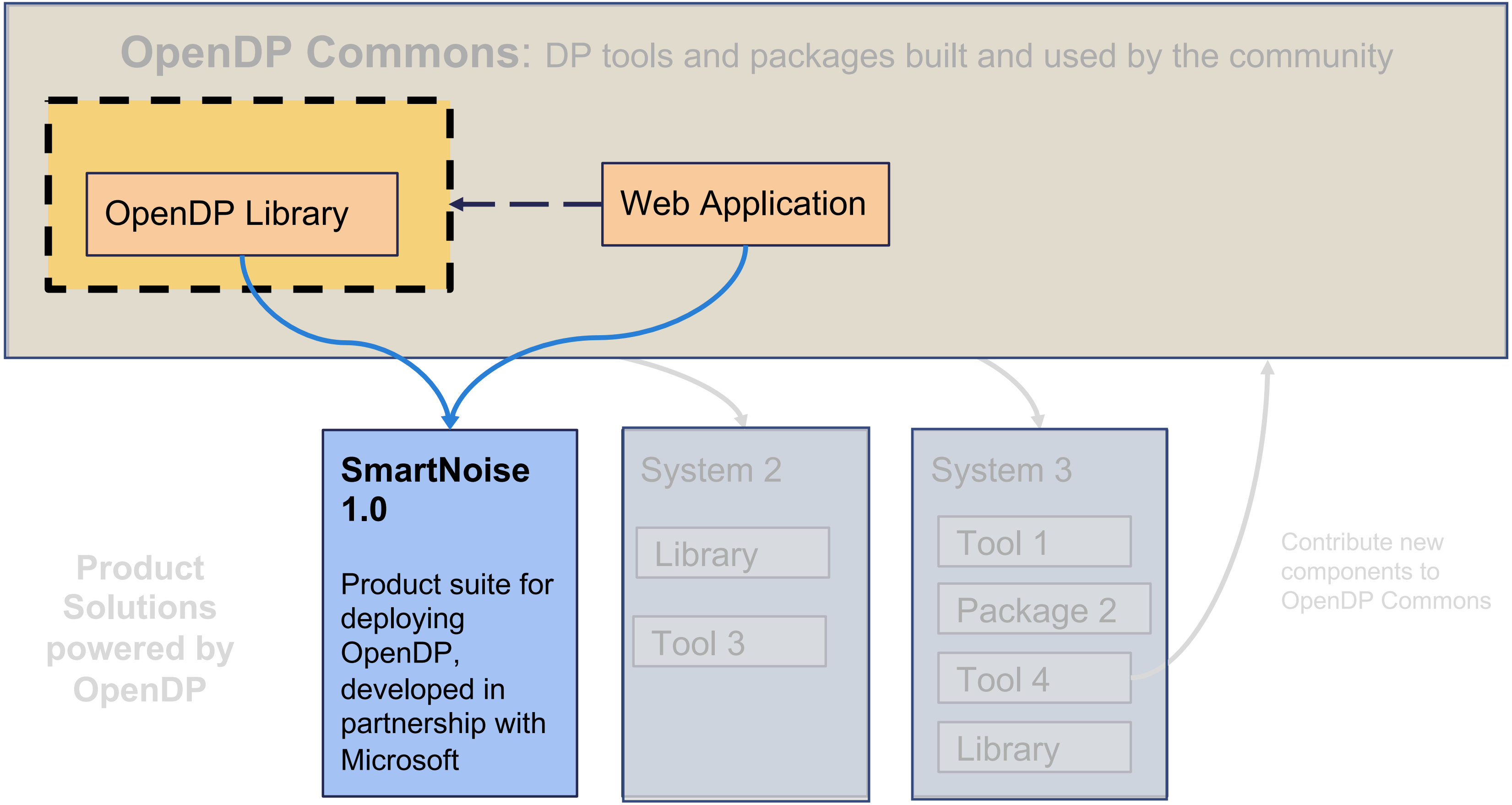
Deposit DP release back to Dataverse

Data User

DP statistics release + Privacy loss + Measure of utility or uncertainty

Impact: Infrastructure for Privacy-Assured Computation
https://cyberimpact.us/

Open-source Governance and Components



OpenDP Commons: DP tools and packages built and used by the community

OpenDP Library

Web Application

SmartNoise 1.0

Product suite for deploying OpenDP, developed in partnership with Microsoft

System 2

Library

Tool 3

System 3

Tool 1

Package 2

Tool 4

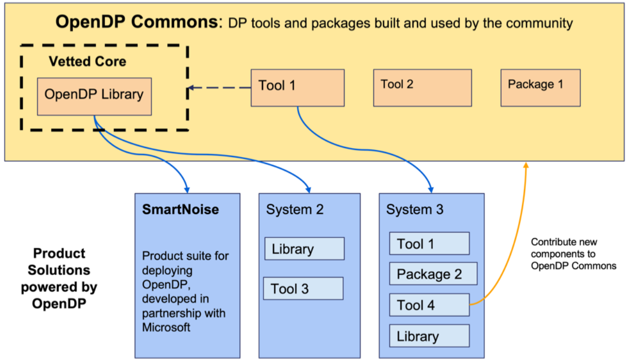
Library

Product Solutions powered by OpenDP

Contribute new components to OpenDP Commons

Spring 2021

- **Feedback** on OpenDP Library
- Review and vetting process for **new contributions**
- OpenDP Community Meeting
- New OpenDP fellows program
- Launch working groups
- Build the community to help deploy **new systems**



OpenDP Commons: DP tools and packages built and used by the community
Vetted Core
OpenDP Library
Tool 1
Tool 2
Package 1
SmartNoise
Product suite for deploying OpenDP, developed in partnership with Microsoft
System 2
Library
Tool 3
System 3
Tool 1
Package 2
Tool 4
Library
Product Solutions powered by OpenDP
Contribute new components to OpenDP Commons

OpenDP

Get Involved

https://opendp.org

Mailing list: https://opendp.org/join

https://github.com/opendifferentialprivacy/

@opendp_io