OpenDP: an Overview

Salil Vadhan
Harvard University
salil_vadhan@harvard.edu

OpenDP Community Meeting
13-15 May 2020

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of our funders.

CC BY 4.0
Goals of Differential Privacy
[... Dwork-McSherry-Nissim-Smith `06]

- **Utility:** enable “statistical analysis” of datasets

- **Privacy:** protect “individual-level” data

[See appendix of OpenDP whitepaper for a brief primer on DP]
### Statistical Releases

<table>
<thead>
<tr>
<th>Name</th>
<th>Sex</th>
<th>Blood</th>
<th>HIV?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chen</td>
<td>F</td>
<td>B</td>
<td>Y</td>
</tr>
<tr>
<td>Jones</td>
<td>M</td>
<td>A</td>
<td>N</td>
</tr>
<tr>
<td>Smith</td>
<td>M</td>
<td>O</td>
<td>N</td>
</tr>
<tr>
<td>Ross</td>
<td>M</td>
<td>O</td>
<td>Y</td>
</tr>
<tr>
<td>Lu</td>
<td>F</td>
<td>A</td>
<td>N</td>
</tr>
<tr>
<td>Shah</td>
<td>M</td>
<td>B</td>
<td>Y</td>
</tr>
</tbody>
</table>

C is a trusted “curator” of statistical tables, trained ML model, and synthetic data etc.
Statistical Query Systems

<table>
<thead>
<tr>
<th>Name</th>
<th>Sex</th>
<th>Blood</th>
<th>HIV?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chen</td>
<td>F</td>
<td>B</td>
<td>Y</td>
</tr>
<tr>
<td>Jones</td>
<td>M</td>
<td>A</td>
<td>N</td>
</tr>
<tr>
<td>Smith</td>
<td>M</td>
<td>O</td>
<td>N</td>
</tr>
<tr>
<td>Ross</td>
<td>M</td>
<td>O</td>
<td>Y</td>
</tr>
<tr>
<td>Lu</td>
<td>F</td>
<td>A</td>
<td>N</td>
</tr>
<tr>
<td>Shah</td>
<td>M</td>
<td>B</td>
<td>Y</td>
</tr>
</tbody>
</table>

trusted “curator”

data analysts
Existing Query Interfaces
Why DP? Attacks on Privacy

- **Re-identification:** determining who is who even after “PII” removed
  - Applied to medical data [Sweeney `97], Netflix challenge [Narayanan-Shmatikov `08], …

- **Database Reconstruction:** reconstructing almost the entire underlying dataset [Dinur-Nissim `03,…]
  - Applied to Census releases [Garfinkel et al. `18] and Diffix [Cohen-Nissim `19].

- **Membership Inference:** determining whether a target individual is in the dataset [Dwork-Smith-Steinke-Ullman-V. `15]
  - Applied to genomic data [Homer et al. `08,…] and ML as a service [Shokri et al. `17,…].
Goals of Differential Privacy
[... , Dwork-McSherry-Nissim-Smith `06]

• **Utility:** enable “statistical analysis” of datasets
  - e.g. inference about population, ML training, descriptive statistics, synthetic data

• **Privacy:** protect “individual-level” data
  - against “all” attack strategies, background info.
  - now accepted as a “gold standard” for protection

How to achieve?
- Inject “small” amount of random noise into statistical computations

[See appendix of OpenDP whitepaper for a brief primer on DP]
Differentially Private Algorithms circa 2014

- 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]
- synthetic data [BLR08, HR10, GGHRW14]
- …
Differential Privacy Deployed

U.S. Census Bureau
• “OnTheMap” commuter data [Machanavajjhala et al. `06]
• Planned: all public-use products from 2020 Decennial Census [Abowd `18]

Tech Industry
• RAPPOR for Chrome Statistics [Erlingsson et al. `14]
• Tensorflow Privacy [Abadi et al. `16,…]
• iOS10 and Safari [2016]
• Windows 10 [Ding et al. `17]
• …

Research Community
• Numerous prototypes from individual groups
OpenDP

A community effort to build a trustworthy and open-source suite of differential privacy tools that can be easily adopted by custodians of sensitive data to make it available for research and exploration in the public interest.

Why?

• Channel our collective advances on science & practice of DP
• Enable wider adoption of DP
• Address high-demand, compelling use cases
• Provide a starting point for custom DP solutions
• Identify important research directions for the field
Planned Structure

**OpenDP: An Open-Source platform for Differential Privacy**

**OpenDP Commons:** DP Library, Tools, Packages designed and built by the community

- OpenDP Library
- Privacy Budgeting Tool
- Common Documentation and Templates
- Testing package
- ...

**OpenDP Systems:** End-to-end differential privacy systems, usually designed and built in a partnership to address a particular use case

- System 1: End-to-end solution for deploying OpenDP in partnership with Microsoft
- System 2: Another end-to-end solution for a different use case
- System 3: ...
  - Package
  - Tool

Library and other common components used by the Systems

New components developed by OpenDP systems contributed back to OpenDP Commons
Key Elements

• Use Cases
• Governance
• Programming Framework
• Statistical Functionality
• System Integrations
• Collaborations
• Community!

More details in plenaries, breakouts, and the whitepaper.
How we got here

Spring/Summer 2019
• Pitch to DP community @ Simons Institute
• Proposal to Sloan Foundation
• Funding received
• Microsoft collaboration starts

Fall/Winter 2019-2020
• Ad Hoc Design Committee meetings & workshop
• OpenDP staff hired
• Software development advances with Microsoft

Spring 2020
• Programming Framework & other elements fleshed out
• First version of system with Microsoft near completion
• Advisory Board formed
• OpenDP Community Meeting!
Where we’re going

**Summer 2020:**
- Absorb community feedback
- Implement DP library in OpenDP Commons
- Form Ad Hoc Security Review Committee
- Find DP Applications Leader(s), COVID-19 use case
- Establish partnership model, more collaborations
- Fundraising

**Fall 2020:**
- Launch the OpenDP Commons with working library
- Establish Editorial Board & Committers to review contributions
- Release MVP of 1st OpenDP System, with Dataverse integration
- Second OpenDP Community Meeting

**Beyond:**
- Expand functionality and deployments
- Form Steering Committee
- Sustainability through community commitment
What can you do?

Follow our plans
- Many more details in the whitepapers at http://opendp.io/
- Watch for emails and posts from us

Contribute
- Participate in breakout discussions
- Send feedback & suggestions to info@opendp.io anytime
- Stay tuned for more opportunities

Collaborate
- See Collaborations plenary & breakout