OpenDP Advisory Board Report

Thursday, June 25, 2020

Gerome Miklau and Adam Smith, chairs. John Abowd, Gilles Barthe, Barbara Bierer, Sarah Bird, danah boyd, Cynthia Dwork, Ulfar Erlingsson, John Friedman, Jeff Gill, Daniel Goroff, Frauke Kreuter, Orran Krieger, David Lazer, Margaret Levenstein, Katrina Ligett, Carlos Maltzahn, Kenneth Mandl, Ilya Mironov, Helen Nissenbaum, Kobbi Nissim, Dina N. Paltoo, Jules Polonetsky, Aaron Roth, Aleksandra Slavkovic, Dawn Song, Latanya Sweeney, Omer Tene, Stefaan G Verhulst and Andrew Vyrros, members.

The OpenDP advisory board met on Friday, May 15, 2020 at the conclusion of the first OpenDP Community Meeting.

The meeting covered a number of topics. The board was generally positive about the current direction of OpenDP and its mission. The discussion generated a number of recommendations for the executive committee (EC), as well as a number of important points for which the board provided no recommendation, but that the board encourages the EC to look into further.

We've grouped the main points under four headings, though this doesn't exactly mirror the stream of the discussion.

1 Governance

The board was impressed with the overall proposed structure for OpenDP. A large part of the discussion centered on the kind of institution that OpenDP should be, and how it relates to other institutions and projects.

- How should OpenDP be chartered as an organization? Several board members recommended that the EC look at establishing OpenDP as a nonprofit separate from

Harvard while, if possible, keeping an affiliation with Harvard. The existence of a separate nonprofit provides flexibility and can facilitate public engagement. On the other hand, university affiliation makes certain aspects much easier, notably accepting and managing grants and donations.

- How should OpenDP relate to other open-source projects and foundations? A related point was about how OpenDP could take advantage of the ecosystem of open-source projects. One particular suggestion was to take further advantage of programs and support structures offered by NumFOCUS. Another suggestion was to affiliate with an open-source foundation (e.g. the Apache or Linux foundations) for community-building and licensing arrangements.
- How should and does OpenDP relate to other projects that use or touch on differential privacy? There was some confusion in the larger community and among board members about how the "OpenDP systems" will relate to OpenDP. Will the systems themselves be open-source? Under specific licenses? How would projects like Social Science One be integrated?

  A related point is that there are several existing and developing open-source libraries or projects. For example, Google and IBM have open-source libraries. The OpenMined project is developing PySyft, which includes code (or wrappers for other libraries) for differentially private analysis as well as secure computation. OpenDP should ideally have a clear position on how it relates to these other efforts.

Another theme of the governance discussion was on resource estimation. In addition to keeping a careful eye on adequate staffing for the project's engineering and administrative needs, the board encourages the EC to look into the time needs and incentive structure for the review process (editorial board, committers). If the project is successful, the demand on the time and attention of the editorial board and committers could be substantial. Can the review process scale up appropriately with volunteers whose core obligations lie elsewhere? Lastly, cultivating an engaged community around OpenDP will require substantial resources—see "Community Role", below.

Although there was some discussion about licensing and contributor agreements, the board was not able to engage in sufficient depth to make a particular recommendation. We hope to discuss these issues with the executive committee further going forward, as part of the existing process for making licensing decisions.

There was discussion of the composition of the advisory board. On one hand, some members felt that "data users" were not adequately represented, with room for more members from the social and bio-medical sciences, for example. (That said, there were concerns that for now it may be hard to find social scientists willing to engage at appropriate depth.) On the other hand, it was pointed out that there are almost no representatives of "data subjects" on the board—for example, civil rights organizations or others who could speak for members of the public affected by technology like OpenDP but not well represented among the technologists building it. We did not have time to discuss term lengths or processes for evolving the board's composition.

More generally, there was a strong recommendation that OpenDP diversify all the teams within OpenDP. Integrating diversity, equity and inclusion (DEI) at all levels of OpenDP will make the project more likely to succeed, both because of the breadth of perspectives it brings, and the specific potential impact of DP on use cases that affect historically marginalized communities. Commendably, the project already has a commitment to DEI in the whitepaper. Further best practices for DEI involve defining specific goals and mechanisms for ongoing evaluation of their accomplishment.

Finally, there was extensive discussion of whether and how OpenDP should take an active role in outreach to the public and advocacy. We report on this discussion below (see "Mission and Role in the Community"), but note here that this question relates to OpenDP's institutional status (as a part of Harvard or a separate nonprofit, for example) and the reception of its initiatives.

2 Use Cases

The board agreed that use-cases are important, noting that the right use-cases early in the life of OpenDP would offer opportunities for publicity and legitimacy of the project.

It was observed that, in the early stages of OpenDP, the best use cases will be those that make available data that has not been shared before, rather than cases where DP is used to provide alternative privacy-protected data that may be criticized as lacking utility. A use case engaging social or biological scientists, and exercising the statistical functionality planned for OpenDP would be welcome. In addition, results from the use-case breakout were raised, showing a surprisingly large interest in "small" data (hundreds of thousands of records or less) over internet-scale data.

In terms of specific application domains, there was general enthusiasm about near-term use cases focused on COVID19 data, as well as interest in deploying DP to safely release system traces from data centers and inside companies. There was also a suggestion that a basic tool supporting the DP release of SQL-like counts would make a good MVP.

In addition to providing algorithms for computing DP outputs, board members wondered how OpenDP would engage with the myriad of issues surrounding a concrete deployment of DP including: whether users are protected or merely individual actions by users are protected, the setting of the privacy loss budget, and the circumstances under which the privacy loss budget may be "refreshed" or "reset".

There is a potential tension between the near-term goal of deploying a minimum-viable-product (MVP) and remaining open to outside contributions, which may not be aligned with the MVP but which nevertheless deserve support from OpenDP. Members of the board diverged in how best to resolve this tension. Regardless, it is critical to be intentional about navigating this tradeoff early.

Finally, the board suggested that OpenDP consider whether the adoption of DP in the use-case may ultimately lead to more invasive collection of personal information, because it can be unduly justified by the use of strong privacy technology.

3 Community Role / Mission

The advisory board discussed a broad spectrum of community engagement and advocacy efforts, described in the paragraphs below.

There seemed to be broad consensus that each of the following are important for differential privacy as an emerging technology. Thus, whether and how OpenDP addresses these efforts is an important question. In addition, the board recognized that each of the outreach efforts below would require significant resources to carry out successfully.

User communities Those adopting OpenDP's tools and/or using data products produced by OpenDP tools should be directly engaged. This includes both technically sophisticated users (academics, statisticians, data scientists) who are unfamiliar with DP, as well as those further from data analytics and statistics who, while unlikely to understand the fundamentals of DP, nevertheless need an "on ramp" to understand the functional role DP can play in data governance.

Educational efforts Differential privacy lacks a coherent set of educational materials; support and publicity for such materials, including accessible papers, books, online courses, and bootcamps, would greatly support the mission of OpenDP.

Data subjects The tools and applications built as part of OpenDP will have significant impacts on data subjects, i.e. the individuals whose data is collected and analyzed. The board discussed how and whether OpenDP could engage directly with data subjects, including community and civil rights groups representing those who have been harmed in the past by government and industry data collection and are worried that technological solutions will harm them further. One possible structure would be a dedicated working-group within OpenDP to engage data subjects and consider their needs, perspectives, and concerns. More broadly, which part of the organizational structure will consider larger ethical questions?

Advocacy There are now, and will continue to be, political battles underway about the use of differential privacy, its impacts and its legality. The board discussed whether OpenDP would defend DP against misrepresentations of the technology, whether OpenDP would contribute to amicus briefs in legal disputes concerning DP, and how OpenDP, as an institution, would respond to inquiries from the press. There should ideally be a clear point of contact designated to speak on behalf of OpenDP.

While advisory board members recognized that some may consider these outreach efforts beyond the scope of OpenDP, they also wondered whether OpenDP would be able to avoid some role in each of the above domains, particularly under the pressure of stakeholders. Established practices for responding to these issues will be important for the ongoing success of OpenDP.

4 Finances

The board also discussed OpenDP's financial sustainability. There were several suggestions for specific sources of support (notably programs at the NSF and NIH). At the organizational level, the board stressed the importance of developing guidelines for how support will be acknowledged and policies on how financial support can affect the trajectory that OpenDP follows.

The board generally encouraged OpenDP to seek industry support in addition to collaboration. Two pitfalls of this approach were discussed: on one hand, OpenDP should be sure to maintain its independence in structuring the Commons and choosing which partnerships to pursue; second, OpenDP should be careful of the potential perception that it is closely tied to specific industry interests. Clear public guidelines could help mitigate those dangers. OpenDP could also benefit from the experience of other open software projects that have received industry support while maintaining their independence. Examples include Jupyter, STAN, ML Commons, and the Partnership on AI.

Conclusion

The advisory board was enthusiastic about OpenDP's mission and the progress the team has made so far. Among the issues raised above, some require attention in the short term—diversity and organization structure, especially—while others call for consideration over the next few months.

Appendix - OpenDP Advisory Board Members

Adam Smith, Professor, Computer Science, Boston University
Gerome Miklau, Professor, College of Computer and Information Sciences, University of Massachusetts, Amherst
John Abowd, Chief Scientist and Associate Director for Research and Methodology, US Census Bureau, Edmund Ezra Day Professor of Statistics and Information Science, Cornell University
Gilles Barthe, Director at Max Planck Institute for Security and Privacy, IMDEA Software Institute, Part-time Research Professor at IMDEA Software Institute
Barbara Bierer, Faculty Director, MRCT Center of Brigham and Women's Hospital, Harvard Professor of Medicine, Harvard Medical School
Sarah Bird, Principal Program Manager, Microsoft
danah boyd, Partner Researcher, Microsoft Research
John Friedman, Professor of Economics and International and Political Affairs, Brown University
Cynthia Dwork, Gordon McKay Professor of Computer Science, Harvard University, Radcliffe Alumnae Professor, Radcliffe Institute for Advanced Study, Distinguished Scientist, Microsoft Research
Ulfar Erlingsson, Manager, Differential Privacy, Apple
Jeff Gill, Distinguished Professor, Departments of Government, and Mathematics & Statistic, Founding Director, Center for Data Science
Daniel Goroff, Vice President and Program Director, Alfred P. Sloan Foundation (ex officio)
Orran Krieger, PI Mass Open Cloud, Co-Director, Red Hat Collaboratory, Professor Electrical and Computer Engineering, Boston University
Frauke Kreuter, Director and Professor, Joint Program on Survey Methodology, University of Maryland, Professor of Statistics and Methodology, University of Mannheim, Head of Statistical Methods Unit, Institute for Employment Research, Germany
David Lazer, University Distinguished Professor of Political Science and Computer and Information Science, Northeastern University
Margaret Levenstein, Research Professor, Survey Research Center, Institute for Social Research, Adjunct Professor of Business Economics, Ross School of Business, University of Michigan
Katrina Ligett, Associate Professor of Computer Science and Head of the Program on Internet & Society, Hebrew University
Kenneth Mandl, Donald A.B. Lindberg Professor of Biomedical Informatics and Pediatrics, Harvard Medical School, Director, Computational Health Informatics Program, Boston Children's Hospital
Carlos Maltzahn, Adjunct Professor of Computer Science & Engineering, UC Santa Cruz, Founder & Director of the Center for Research in Open Source Software (CROSS)
Ilya Mironov, Research Scientist, Facebook AI

Helen Nissenbaum, Professor, Cornell Tech Information Science, Director, Digital Life Initiative, Cornell University
Kobbi Nissim, Professor, Department of Computer Science, Georgetown University
Dina N. Paltoo, Assistant Director, Scientific Strategy and Innovation, Immediate Office of the Director, National Heart, Lung, and Blood Institute
Jules Polonetsky, CEO, Future of Privacy Forum
Aaron Roth, Professor of Computer and Information Science, University of Pennsylvania
Aleksandra Slavkovic, Professor, Associate Dean for Graduate Education, Eberly College of Science, Pennsylvania State University
Dawn Song, Professor, Computer Science Division, University of California Berkeley
Latanya Sweeney, Professor of the Practice of Government and Technology and Director of the Data Privacy Lab, Harvard University
Omer Tene, Vice President and Chief Knowledge Officer, International Association of Privacy Professionals
Stefaan G Verhulst, Co-Founder and Chief Research and Development Officer, GovLab, NYU
Andrew Vyrros, Technology Consultant and Principal, Wafflehead Labs