Ethical Concerns, Conduct and Public Policy for Re-Identification and De-identification Practice: Part 3 (Re-Identification Symposium)

October 2, 2013 (https://blog.petrieflom.law.harvard.edu/2013/10/02/ethical-concerns-conduct-and-public-policy-for-re-identification-and-de-identification-practice-part-3-re-identification-symposium/) mmeyer (https://blog.petrieflom.law.harvard.edu/author/mmeyer/) Empirical (https://blog.petrieflom.law.harvard.edu/category/empirical/), Genetics (https://blog.petrieflom.law.harvard.edu/category/genetics/), Health Information Technology (https://blog.petrieflom.law.harvard.edu/category/health-information-technology/), Human Subjects Research (https://blog.petrieflom.law.harvard.edu/category/human-subjects-research/), Medical Privacy (https://blog.petrieflom.law.harvard.edu/category/medical-privacy/), Re-Identification Symposium (https://blog.petrieflom.law.harvard.edu/category/blog-symposia/re-identification-symposium/)

This post is part of Bill of Health's *symposium on the Law, Ethics, and Science of Re-Identification Demonstrations. Background on the symposium is here (https://blogs.law.harvard.edu/billofhealth/2013/05/13/online-symposium-on-the-law-ethics-science-of-re-identification-demonstrations/). You can call up all of the symposium contributions by clicking here (https://blogs.law.harvard.edu/billofhealth/category/re-identification-symposium/). —MM*

By Daniel C. Barth-Jones (https://www.mailman.columbia.edu/our-faculty/profile?uni=db2431)

In Part 1 (https://blogs.law.harvard.edu/billofhealth/2013/05/29/public-policy-considerations-for-recent-re-identification-demonstration-attacks-on-genomic-data-sets-part-1-re-identification-symposium) and Part 2 (https://blogs.law.harvard.edu/billofhealth/2013/10/01/press-and-reporting-considerations-for-recent-re-identification-demonstration-attacks-part-2-re-identification-symposium/) of this symposium contribution I wrote about a number of re-identification demonstrations and their reporting, both by the popular press and in scientific communications. However, even beyond the ethical considerations that I've raised about the accuracy of some of these communications, there are additional ethical, "scientific ethos", and pragmatic public policy considerations involved in the conduct of re-identification research and de-identification practice that warrant some more thorough discussion and debate.

First Do No Harm

Unless we believe that the ends always justify the means, even obtaining useful results for guiding public policy (as was the case with the PGP demonstration attack's validation of "perfect population register" issues) doesn't necessarily mean that the conduct of re-identification research is on solid ethical footing. Yaniv Erlich's admonition in his *"A Short Ethical Manifesto for the Privacy Researcher*" (https://blogs.law.harvard.edu/billofhealth/2013/05/23/breaking-good-a-short-ethical-manifesto-for-the-privacy-researcher/) blog post contributed as part of this symposium provides this wise advice: "*Do no harm to the individuals in your study. If you can prove your point by a simulation on artificial data – do it.*" This is very sound ethical advice in my opinion. I would argue that the re-identification risks for those individuals in the PGP study who had supplied 5-digit Zip Code and full date of birth were already understood to be unacceptably high (if these persons were concerned about being identified) and that no additional research whatsoever was needed to demonstrate this point. However, if additional arguments needed to be made about the precise levels of the risks, this could have been adequately addressed through the use of probability models. I'd also argue that "data intrusion scenario" uncertainty analyses which I discussed in Part 1 of this symposium contribution (https://blogs.law.harvard.edu/billofhealth/2013/05/29/public-policy-considerations-for-recent-re-identification-

demonstration-attacks-on-genomic-data-sets-part-1-re-identification-symposium) already accurately predicted the very small re-identification risks found for the sort of journalist and "nosy neighbor" attacks directed at the Washington hospital data. When strong probabilistic arguments can be made regarding potential re-identification risks, there is little possible purpose for undertaking actual re-identifications that can impact specific persons.

Looking more broadly, it seems more reasonably debatable whether the earlier January re-identification attacks by the Erlich lab on the CEPH – Utah Residents with Northern and Western European Ancestry (CEU) participants (https://snp.cshl.org/citinghapmap.html.en)could have been warranted by virtue of the attack having exposed a previously underappreciated risk. However, I think an argument could likely be made that, given the prior work by Gitschier which had already revealed the re-identification vulnerabilities of CEU participants (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2668019/pdf/main.pdf), the CEU portion of the Science paper also might not have served any additional purpose in directly advancing the science needed for development of good public policy. Without the CEU re-identifications though, it is unclear whether the surname inference paper would have been published (at least by a prominent journal like *Science*) and it also seems quite unlikely that it would have sustained nearly the level of media attention.

Ethical Dilemmas

In trying to sort out these admittedly complex questions about ethical conduct of re-identification research, I find it interesting to juxtapose a section from the 2009 Gitschier paper "*Inferential Genotyping of Y Chromosomes in Latter-Day Saints Founders and Comparison to Utah Samples in the HapMap Project* (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2668019/pdf/main.pdf)" (which laid out the foundational issues supporting the Erlich Lab's surname inference attack method) with some comments reported in the news about the surname inference attacks:

[From page 256 in the Gitschier paper]

> *Each of the matching SMGF haplotypes is associated with a surname, and the obvious question emerges as to whether these surnames indeed correspond to the surnames of the CEU contributors themselves. This direct question is unanswerable. In consultation with investigators at the University of Utah, where the samples were collected, we* ***jointly concluded that confirmation of the predicted surnames would violate the ethical constraints of informed consent obtained during the collection of these samples*** *because the names of the subjects would be used in the analysis. Moreover,* ***in deference to the privacy of those who contributed the CEU samples I have not included the predicted surnames*** *in Table 2. Instead, I have attempted to assess the accuracy of the predictions with several simulations, as follows: To challenge the power of using a collection of only 17 STR markers to accurately screen the SMGF database, I submitted the 17-marker subset of alleles, shown in Table 2, corresponding to the consensus haplotypes for both Joseph Smith and Brigham Young. In response to each of these queries, conducted without regard to surname, I retrieved largely the same set of individuals that I had uncovered with the larger set of ~40 markers. This very limited test suggests that a reasonable guess as to male ancestors may be made by a relatively small set of highly informative markers, at least within this targeted database.* [Emphasis Added]

I found the section above to be a helpful example illustrating judicious care not to advance the same potential harms that one is seeking to expose and prevent.

Providing some additional interesting interplay on these same issues is this section from the New York Times article on the surname inference attack (https://www.nytimes.com/2013/01/18/health/search-of-dna-sequences-reveals-full-identities.html):

> *...Dr. Jeffrey R. Botkin, associate vice president for research integrity at the University of Utah, which collected the genetic information of some research participants whose identities were breached, cautioned about overreacting. Genetic data from hundreds of thousands of people have been freely available online, he said, yet there has not been a single report of someone being illicitly identified. He added that "**it is hard to imagine what would motivate anyone to undertake this sort of privacy attack in the real world**." But he said he had serious concerns about publishing a formula to breach subjects' privacy. **By publishing,** he said, **the investigators "exacerbate the very risks they are concerned about."***
>
> *Yaniv Erlich, a human genetics researcher at the Whitehead Institute, which is affiliated with M.I.T. He stresses that he is a strong advocate of data sharing and that he would hate to see genomic data locked up. But when his lab developed a new technique, he realized he had the tools to probe a DNA database. And he could not resist trying.* [Emphasis Added]

Arguably, a point could also be made, for example, that, given the existent knowledge we had from the 2009 Gitschier paper, the subsequent Y-STR surname inference attack (https://www.ncbi.nlm.nih.gov/pubmed/23329047) was more akin to a "weaponization" of this attack rather than a new revelation of its possibility.

Misaligned Incentives

In Part 1 of this symposium contribution (https://blogs.law.harvard.edu/billofhealth/2013/05/29/public-policy-considerations-for-recent-re-identification-demonstration-attacks-on-genomic-data-sets-part-1-re-identification-symposium), I wrote about my concerns that there might be incentives for re-identification scientists and the media to join forces (unintentionally or not) in overstating real-life re-identification risks. Editors and writers might be tempted to drum up a media audience and researchers may have motivations to promote their research. Yet, as much as re-identification researchers can provide uniquely insightful contributions regarding the re-identification issues that policy-makers should be anticipating and addressing, as academics, they will be primarily rewarded only for producing research publications. Low profile privacy policy advocacy efforts will do little to further privacy researcher's careers, particularly if re-identification risk research is left (at least temporarily) unpublished so that researchers may first work with data holders and policy-makers in order to reduce re-identification risks prior to the public exposure of such vulnerabilities. Unfortunately, the fact that academics presently receive very little career recognition for unpublished public policy service activities adds to the potential for misaligned incentives and associated ethical concerns.

Re-identification researchers, academic leaders, publishers and popular press journalists all need to engage in continued open discussion/debate together regarding ethical issues for health data de-identification (https://www.tandfonline.com/toc/uajb20/10/9) and whether the current incentives underlying the conduct and reporting of re-identification research are actually aligned with their ethical obligations to the individuals being re-identified and to society as a whole. To advance our discussion of these ethical issues, I offer the following thoughts about the complex role of the concepts of beneficence and justice in re-identification research.

Of Beneficence and Justice

While Dr. Erlich was off to a great start in his earlier *"A Short Ethical Manifesto for the Privacy Researcher"* (https://blogs.law.harvard.edu/billofhealth/2013/05/23/breaking-good-a-short-ethical-manifesto-for-the-privacy-researcher/) symposium post with his guidance to "do no harm", his vision was ultimately incomplete when he only advised privacy researchers to "*Do no harm to the individuals in your study*". The Belmont Report's ethical principle (https://www.hhs.gov/ohrp/policy/belmont.html) of *Beneficence* (which routinely helps to guide Institutional Review Boards (IRB) though their human subjects research reviews under the Common Rule) actually involves two complementary ethical rules. Because the simplistic "do not harm" adage fails to adequately address the complex ethical quandary as to when it is justifiable to seek certain benefits despite the risks involved, and when such benefits should be foregone because of the risks, the second axiom underlying the principle of Beneficence is the necessity to "*maximize possible benefits and minimize possible harms*".

As the Belmont report wisely recognizes:

> *The **obligations of beneficence** affect both individual investigators and society at large, because they **extend both to particular research projects and to the entire enterprise of research**. In the case of particular projects, investigators and members of their institutions are obliged to give forethought to the maximization of benefits and the reduction of risk that might occur from the research investigation. In the case of scientific research in general, members of the larger society are obliged to recognize the longer term benefits and risks that may result from the improvement of knowledge and from the development of novel medical, psychotherapeutic, and social procedures.* [Emphasis Added]

Because re-identifications can occur when it is possible to link persons who are unique on their combined characteristics within a dataset to these same characteristics for unique persons listed within a comprehensive population register for the larger population (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2076397), it should be intuitively clear that re-identification risks aren't typically equally distributed among all of the members of a population. Individuals with rarer characteristics will often be at higher risks of re-identification if these characteristics have not been appropriately protected by effective statistical disclosure control methods. For example, for re-identifications based on combined sets of demographic characteristics, the elderly (who, of course, become increasing rarer in frequency as they grow older) and racial/ethnic minority groups can be disproportionally at risk of re-identification.

It is for this reason that scientific reports of re-identification demonstrations need to be extremely clear as to when they've targeted only particularly vulnerable sub-populations in order to provide a "proof-of-concept" for a re-identification attack approach. If re-identification researchers aren't explicit about having selected especially re-identifiable sub-populations, the general public (and even policy evaluators) are likely to miss that the risks facing the general population haven't been well represented. Proper public policy assessment requires that, whenever it is possible, researchers should seek to produce reliable estimates for the average re-identification risks for the entire population. This is a much preferable stance to that of only selecting a particularly susceptible sub-population because the demonstration will show greater risks. And because the media will inevitably be drawn to reporting on re-identification attacks demonstrating greater risks, researchers need to carefully convey their results to the media in a fashion that accurately reflects these population-wide risks and actively work to correct the record if media attention does not reflect this in order to avoid distorting public perceptions. It is still important to convey influential factors

that heighten re-identification risks for particular persons or sub-populations to help make potentially impacted persons aware of these risks, but purposely selecting only an especially susceptible sub-population in order to obtain more impressive results cannot be justified.

The often inequitable distribution of re-identification risks also leads us directly to consideration of another of the Belmont Report's basic ethical principles: namely, *justice*. Injustice occurs when benefits of research for individuals are unequally denied and/or when risks are inequitably distributed (without good reason). Such concerns about justice within re-identification research, of course, interact in a complex fashion when consideration is given to the "*maximize possible benefits and minimize possible harms*" dictum in our ethical obligations to provide beneficence. Hence, my open question from Part 1 of this symposium contribution (https://blogs.law.harvard.edu/billofhealth/2013/05/29/public-policy-considerations-for-recent-re-identification-demonstration-attacks-on-genomic-data-sets-part-1-re-identification-symposium), as to whether it is an ethically compromised position, particularly in the coming age of personalized medicine, if we end up purposefully masking the racial, ethnic or other group membership status information (e.g. American Indians or LDS Church members, etc.) for certain individuals, or for those with certain rare genetic diseases/disorders, in order to protect them against supposed re-identifications. In making this ethical determination, we must, of course, recognize that by doing so, we would also deny them the benefits of research conducted with de-identified data that could help address their health disparities, find cures for their rare diseases, or facilitate "orphan drug" research that would otherwise not be economically viable.

The ethical calculus of these questions is especially challenging if these hypothetical re-identification threats would not actually be implemented in the real world, but we have been falsely alarmed by demonstration attacks by re-identification scientists who operate free of the usual economic constraints that govern motivations to re-identify.

Complex Ethical Calculus and Ethical Equipoise

With all of the complexities surrounding the risks, benefits, costs and motivations that come into play in re-identification research, it is not surprising that this relatively nascent field invokes a host of complicated questions regarding the ethics of re-identification demonstrations (https://dataprivacylab.org/events/2013a/), including:

- What responsibilities do re-identification researchers have to individuals in de-identified data?
- What responsibilities do re-identification researchers have to communities or to individuals who were not in the study but who may be affected by the re-identification research?
- When, and in what circumstances, should re-identification methods/results be made public?

We face admittedly challenging ethical questions and calculus regarding not only potentially causing harms to persons re-identified through conduct of re-identification research, but also possibly reducing important societal benefits if we do not accurately estimate and represent the true risks. Clearly, there is an important motivation for mitigating possible privacy risks and harms by researching and exposing re-identification vulnerabilities. However, we must also struggle with even broader longer-range questions as to how we can best balance re-identification risks with the considerable benefits that come from research analyses conducted with de-identified data.

I've argued elsewhere (https://healthaffairs.org/blog/2012/08/10/the-debate-over-re-identification-of-health-information-what-do-we-risk/) that we must achieve an ethical equipoise between potential privacy harms and the very real benefits that result from the advancement of science and healthcare improvements which are accomplished with de-identified data. For those who are seeking a yet undiscovered cure for a disease that afflicts them, our societal failure to move medical science forward isn't just some vague "unrealized benefit" but rather a quite real and palpable harm. When re-identification risks are exaggerated, we need to recognize that the resulting fears cause needless harms. Such fears can push us toward diminishing our use of properly de-identified data, or distorting the accuracy of our statistical methods because we've engaged in ill-motivated de-identification and have altered data even in cases where there was not anything more than *de minimis* re-identification risks. Disturbingly, the resulting inaccurate

conclusions we will reach during the conduct of medical science and making of healthcare decisions due to using overly de-identified data are not easily detectable. Unless we use de-identification methods that double check against the original data and minimize such statistical distortions, we risk becoming systematically blinded to reality by overuse of de-identification in cases when it isn't needed to produce privacy protections. If we've unnecessarily distorted the data that we use in our attempts to conduct science and better understand the world because of exaggerated re-identification reporting, society as a whole suffers. Even more damaging in this complicated harm/benefit-balancing calculus would be slowing, or abandoning, scientific progress unnecessarily because unfounded fears have led us to the oft-tweeted, but ultimately errant, conclusion that "de-identification is a myth" (https://www.ipc.on.ca/images/Resources/anonymization.pdf).

Re-identification researchers need to come to some meaningful consensus regarding this formidable thicket of ethical questions which can help guide their research conduct and reporting; and, as they do so, it is important to recognize some historical legacies that stem from the two quite different research communities which have been drawn together in addressing these questions.

Culture Clash

I've spent a lot of time over the past several years reflecting on the cultural norms and ethos of the twin disciplines "*Statistical Disclosure Control*" and "*Privacy-Preserving Computer Science*" which have primarily contributed to the body of scientific work we have regarding data de-identification and re-identification. The statistical disclosure literature began somewhat earlier and evolved primarily within the cultural norms and ethos of the medical, statistical and research communities. While the focus of this work was still protecting the privacy of sensitive (often medical) information, there was, quite admittedly, nowhere near the sense that constant vigilance and "provable privacy guarantees" were imperatives in order to protect against all-pervasive re-identification threats as is now more commonly held by their computer security/privacy colleagues. Much of this cultural gulf arises from the social norms and training of the two groups. To wit, if you're a cryptographer, it is not surprising that you might be more inclined to suspect that everyone's a spy – it's just part of your training to do so. But some of this culture clash results from the later entry of computer scientists into the field during a far more interconnected, data rich and vulnerable time in history. So, it should not be too surprising then that white hat hackers conducting "penetration testing" likely think that other researchers are just fooling themselves if they rely on social and cultural norms, data use contracts and other legal protections, and "security by obscurity" as part of the total package which prevents the occurrence of re-identification attempts.

Yet at the same time, medical researchers may well be equally as skeptical about computer scientists' claims that virtually "Everything is PII" (https://dl.acm.org/citation.cfm?id=1743558) (Personally Identifiable Information). While Arvind Narayanan (https://blogs.law.harvard.edu/billofhealth/2013/05/26/reidentification-as-basic-science/) is indeed correct that "*any information that distinguishes one person from another can be used for re-identifying anonymous data*", there is in reality a great difference between "can" and "will". There is typically a wide chasm between being "distinguishable" and "identifiable". The extent of this "information gulf" is highly dependent on the information correctly known (or knowable) to the data intruder and the effort that a data intruder would exert in order to accurately obtain this information. As explicated by my earlier "perfect population register" points, even for information that is highly distinguishable, achieving correct re-identifications is usually non-trivial due to issues of data availability, completeness and divergence. Given the statistical data utility damage caused by overuse of de-identification, whenever the real re-identification risks are already very small, not differentiating between theoretical "can" and the practical "will" ultimately becomes a very harmful and costly set of starting assumptions.

Having been exposed to the ethos of both of the statistical disclosure and computer security milieus, I personally believe that we'll all be much better off in finding acceptable answers to these complex re-identification research ethical questions if these two communities would better meld their baseline assumptions and ethos to more closely fit the reality of the particular situations and contexts (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=534622) in which de-identified data is utilized.

Privacy "Guarantees" and Differential Privacy

To the statistical and medical research communities, computer scientists may seem to be obsessed with privacy "guarantees" and associated mathematical proofs. This "suspect everyone" conventional wisdom from the computer security community can be appropriate when addressing security contexts where background assumptions of all pervasive threats are realistic. But, as we've seen from the body of evidence that has accumulated in the decade since the HIPAA privacy rule has gone into effect, one can point to very few, if any, cases of persons who have been harmed by verified re-identification attacks (https://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0028071), particularly if we exclude re-identification demonstrations conducted by academics from this tabulation.

The suppositions of Differential Privacy (https://www.scientificamerican.com/article.cfm?id=privacy-by-the-numbers-a-new-approach-to-safeguarding-data), for example, essentially assume that all researchers accessing any differentially private dataset are omniscient, omnipotent and constantly co-conspiring data snoops. In order to provide privacy guarantees that aren't dependent on the background knowledge of a data intruder, it's presumed that all data intruders know everything — except whether a target individual is in the data set (near omniscience); or would have the time, money, computational resources and motivation to acquire such near all-pervasive knowledge (omnipotence). Finally, it's implicitly assumed that everyone who accesses the data is not only a data snoop, but also working in "cahoots" with everyone else accessing the data to attack the privacy of every person represented by the data, thus, warrantying a shared fixed "privacy budget" for all queries against the data base. Once this privacy budget has been exhausted, the dataset must be closed to any additional study — regardless of its expense or future research value.

Now, medical researchers are an admittedly bright and capable bunch, but differential privacy's presumed omniscience and omnipotence also comes with an assumption of perpetual malevolence, so it's not surprising that differential privacy's ridiculously over-the-top assumptions simply are not a good fit when we're considering the context of data access by medical and public health researchers. Researchers have spent many years of their lives getting M.D. and Ph.D. degrees in order to help people and not to be able to snoop on them. Data de-identification researchers Fida Dankar and Khaled El Emam have provided a useful overview of the important limitations of differential privacy for healthcare data (https://www.tdp.cat/issues11/tdp.a129a13.pdf). For this health research context (and many others), the baseline assumptions of differential privacy are just as absurd as our blindly accepting some dictum that all emergency room doors should now be replaced with bank vault doors.

Admittedly, Differential Privacy may be a helpful solution in situations where the necessary noise being added to the data can be appropriately tolerated and the risk of re-identification attempts could be high (e.g., when data is being publically released without data use agreements and access controls on the internet). Privacy guarantees and associated mathematical proofs are of value if they are able to lead to workable solutions for our immediate needs, but if that is not the case, we may need to rely on empirically workable heuristic approaches.

To any readers who are skeptical about my point here, my rejoinder is that it is likely that the door to your house is locked even though it won't "guarantee" security. Fortunately, we have some fairly effective heuristic approaches for separating out when re-identification risks have been importantly reduced by de-identification methods so it is not all that difficult to find solutions that are "good enough" and provide both dramatic privacy protections and useful analytic data. The bottom line is that both the computer science and statistical privacy communities should hopefully recognize that actually achieving effective protection of privacy should always be more important than whether we've been able to devise and utilize some elegant mathematical proof in our process of getting there.

With that in mind, I turn to some governance and regulatory related suggestions that can add importantly to the technical solutions provided by statistical disclosure control methods and privacy-preserving computer science solutions.

Some Sensible Steps Forward

I have previously recommended several best practices for the use of de-identified data (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2076397) that should be considered by legislators and regulators as possible required de-identified data use conditions. First and foremost, these recommendations include strong prohibitions on the re-identification, or attempted re-identification, of individuals and their relatives, family or household members. Congress should consider establishing civil and criminal penalties for unauthorized re-identification of de-identified data (and for HIPAA Limited Data Sets). A carefully designed prohibition on re-identification attempts could still allow research involving re-identification of specific individuals to be conducted under the approval of Institutional Review Boards (IRBs), but would ban such re-identification attempts conducted without essential human subjects research protections.

Robert Gellman has proposed a well-conceived voluntary legislative-based contractual solution (https://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?article=1277&context=iplj) that, with some appropriate modifications and enhancements, could serve as a suitable foundation for such legislative efforts. Coalitions from the information industries, which depend on de-identified data for their operations, and privacy advocates alike, should press forward toward a consensus in support of legislation prohibiting data re-identification in order to protect the substantive privacy protections afforded by properly implemented de-identification practices.

Probabilistic re-identification risk assessments that do not attempt to match direct or overt identifier data (e.g., HIPAA Limited Data set identifiers (https://privacyruleandresearch.nih.gov/pr_08.asp)) to actual individuals should fall outside any legislative definition for re-identification attempts and, thus, would escape this legislative prohibition. While these probabilistic risk assessments will most likely be conservative because they will not have accounted for false positive re-identifications with full validation studies, they would still allow a great deal of useful re-identification research to move forward without a need for IRB oversight.

Data protection requirements of the sort suggested would impose some additional modest impositions on the use of HIPAA de-identified data, but would help to provide recourse for actions against data intruders. Because properly conducted de-identification should consistently assure that no more than a very small proportion of the de-identified data would be re-identifiable, economic evaluations using cost-benefit analyses of re-identification attempts based on linkage to external data sources show that they are typically not economically viable as small-scale efforts targeted at a specific person, or a small number of individuals. To achieve any reasonable economic payoffs sufficient to counter the expense (in terms of time, effort, requisite computer and mathematical skills and data costs) involved in constructing a near perfect population register, data-based re-identification attempts need to be large-scale efforts. Such large-scale re-identification efforts would be susceptible to detection and prosecution using a combination of statistical analyses similar to those used in investigating employment discrimination allegations and computer forensics.

Final Comments

This symposium has provided a much needed start to a critical debate that should be faced head on and openly engaged by re-identification scientists and the academic/popular press who report on this important and rapidly accelerating issue. Unfortunately, as has been displayed by some recent re-identification demonstrations, attacks focused on vulnerable, but rare, targets are sometimes used in order to demonstrate that the attack is possible. And, as I've discussed elsewhere (https://www.concurringopinions.com/archives/2012/09/re-identification-risks-and-myths-superusers-and-super-stories-part-ii-superusers-and-super-stories.html), such improbably realized demonstrations may serve to usher the general public and policy-makers into improbable and unfounded fears of the inevitability of ultimately low probability events. Unless re-identification researchers carefully attend to the ethical implications of the

conduct and reporting of re-identification demonstrations, we risk furthering the very privacy harms that we seek ultimately to prevent and additionally creating far broader harms to the scientific and healthcare improvement progress that society routinely derives from the "data commons" built with de-identified data.

Re-identification research can and does provide a valuable service to society when it helps the public and policy-makers to accurately perceive and respond to disclosure risk scenarios where there are realistic motivations for re-identification attacks with credible risks and harms. Hopefully an ongoing debate on these issues will result in a meaningful consensus regarding ethical conduct of re-identification research that will better balance the full range of societal interests, both in importantly protecting individual privacy and in preserving the considerable public good that results from research conducted with properly de-identified data.

LEAVE A REPLY

You must be logged in (https://blog.petrieflom.law.harvard.edu/wp-login.php?redirect_to=https%3A%2F%2Fblog.petrieflom.law.harvard.edu%2F2013%2F10%2F02%2Fethical-concerns-conduct-and-public-policy-for-re-identification-and-de-identification-practice-part-3-re-identification-symposium%2F) to post a comment.

This site uses Akismet to reduce spam. Learn how your comment data is processed (https://akismet.com/privacy/).

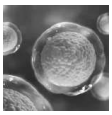
‹ Nanocourse in Public Health at HSPH: FIRST SESSION TOMORROW (https://blog.petrieflom.law.harvard.edu/2013/10/02/nanocourse-in-public-health-at-hsph-first-session-tomorrow/)

Worth Reading This Week › (https://blog.petrieflom.law.harvard.edu/2013/10/02/worth-reading-this-week-31/)

Search...

MOST POPULAR

(https://blog.petrieflom.law.harvard.edu/2024/04/04/eu-and-us-regulatory-challenges-facing-ai-health-care-innovator-firms/) EU and US Regulatory Challenges Facing AI Health Care Innovator Firms (https://blog.petrieflom.law.harvard.edu/2024/04/04/eu-and-us-regulatory-challenges-facing-ai-health-care-innovator-firms/)
by The Petrie-Flom Center Staff (https://blog.petrieflom.law.harvard.edu/author/petrieflom/)



(https://blog.petrieflom.law.harvard.edu/2024/04/07/cell-therapies-and-their-legal-discontents/) Cell Therapies and their Legal Discontents (https://blog.petrieflom.law.harvard.edu/2024/04/07/cell-therapies-and-their-legal-discontents/)
by Adithi Iyer (https://blog.petrieflom.law.harvard.edu/author/aiyer/)

(https://blog.petrieflom.law.harvard.edu/2024/04/08/salus-populi-training-the-judiciary-in-the-social-drivers-of-health/) Salus Populi: Training the Judiciary in the Social Drivers of Health (https://blog.petrieflom.law.harvard.edu/2024/04/08/salus-populi-training-the-judiciary-in-the-social-drivers-of-health/)
by The Petrie-Flom Center Staff (https://blog.petrieflom.law.harvard.edu/author/petrieflom/)

(https://blog.petrieflom.law.harvard.edu/2024/04/18/two-years-on-from-a-landmark-abortion-decision-in-kenya/) Two Years On From A "Landmark" Abortion Decision in Kenya (https://blog.petrieflom.law.harvard.edu/2024/04/18/two-years-on-from-a-landmark-abortion-decision-in-kenya/)

by Joelle Boxer (https://blog.petrieflom.law.harvard.edu/author/jboxer/)

(https://blog.petrieflom.law.harvard.edu/2024/04/06/protecting-health-privacy-is-a-royal-pain/) Protecting Health Privacy is a Royal Pain (https://blog.petrieflom.law.harvard.edu/2024/04/06/protecting-health-privacy-is-a-royal-pain/)
by Bobby Stroup (https://blog.petrieflom.law.harvard.edu/author/rstroup/)

GET OUR NEWSLETTER

SUBSCRIBE NOW (HTTPS://PETRIEFLOM.WPENGINE.COM/SIGN-UP-FOR-THE-PETRIE-FLOM-CENTER-NEWSLETTER/)

HOT TOPICS

Select Category

ARCHIVES

Select Month

POWERED BY

The Petrie-Flom Center for Health Law Policy, Biotechnology, and Bioethics at Harvard Law School (https://petrieflom.law.harvard.edu/)

PAGES

About Bill of Health (https://blog.petrieflom.law.harvard.edu/about/)

Policies (https://blog.petrieflom.law.harvard.edu/policies/)

Symposia (https://blog.petrieflom.law.harvard.edu/symposia/)

In Focus Series (https://blog.petrieflom.law.harvard.edu/in-focus/)

SIGN UP FOR OUR NEWSLETTER

SUBSCRIBE (HTTPS://PETRIEFLOM.WPENGINE.COM/SIGN-UP-FOR-THE-PETRIE-FLOM-CENTER-NEWSLETTER/)

FOLLOW US

(https://www.linkedin.com/company/the-petrie-eflom/)

(https://www.facebook.com/petrieflom/)

(https://vimeo.com/petrieflom)

(https://twitter.com/PetrieFlom)