

# Netflix Needs to Put "Privacy Risks" in Their Queue

September 30, 2009

Netflix recently announced winners of the one-million-dollar "Netflix Prize" and its plans for a new competition, creatively dubbed "Netflix Prize 2." Although details of this second contest still haven't been made public, the [New York Times has reported](#) [1] that competitors will be challenged to "model individuals' 'taste profiles'," based on a dataset that will hold "demographic and behavioral data," including information about members' ages, gender, ZIP codes, genre preferences, as well as their rental histories and movie ratings.

The announcement of Netflix Prize 2 illustrates the continued blinders that companies have about the ease with which individuals listed in supposedly anonymized datasets can be identified. The ostensibly anonymized dataset released to the public for Netflix Prize was limited to the video rental histories of 480,000 Netflix subscribers, the ratings (1-5 stars or "no rating") that subscribers gave each movie, and the date that subscribers rated each movie.

As law professor and [CDT Academic Fellow Paul Ohm has pointed out](#) [2], between 2006 and 2008, researchers [Arvind Narayanan and Vitaly Shmatikov showed that](#) [3] if you know just a little information about a friend (or enemy's) movie-viewing habits, using the Netflix Prize database you can likely uncover every movie your friend ordered through Netflix and how she rated it. Add demographic and behavioral data into the mix and how do you even take seriously claims that the data has been and will remain de-identified?

Netflix is far from the only company releasing easily de-anonymized data under pretenses that the sets' contents are unidentifiable. From the user identifications that came out of AOL's [infamous data release](#) [4] to Latanya Sweeney's use of Massachusetts residents' ZIP code, birth date, and gender - all found in public voter rolls - to identify individuals whose "anonymized" hospital records had been [publicly released](#) [5], the pretense of anonymization is time and again revealed as false. Companies need to own up to the privacy risks inherent in releasing such data and find more comprehensive and robust methods for truly de-identifying their data and protecting their customers.

- [tbp](#)
- [privacy](#)
- [netflix](#)
- [CDT](#)
- [anonymous data](#)

Copyright © 2009 by Center for Democracy & Technology. CDT can be freely copied and used as long as you make no substantive changes and clearly give us credit. [Details](#).

**Source URL:** <https://cdt.org/blogs/erica-newland/netflix-needs-put-privacy-risks-their-queue>

## Links:

- [1] <http://bits.blogs.nytimes.com/2009/09/21/netflix-awards-1-million-prize-and-starts-a-new-contest/>  
[2] <http://www.freedom-to-tinker.com/blog/paul/netflixs-impending-still-avoidable-multi-million-dollar-privacy-blunder>  
[3] [http://www.cs.utexas.edu/~shmat/shmat\\_oak08netflix.pdf](http://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf)  
[4] [http://www.nytimes.com/2006/08/09/technology/09aol.html?\\_r=1&scp=2&sq=aol%20data&st=Search](http://www.nytimes.com/2006/08/09/technology/09aol.html?_r=1&scp=2&sq=aol%20data&st=Search)  
[5] [http://arstechnica.com/tech-policy/news/2009/09/your-secrets-live-online-in-databases-of-ruin.ars?utm\\_source=microblogging&utm\\_medium=arstch&utm\\_term=Law%20&%20Disorde](http://arstechnica.com/tech-policy/news/2009/09/your-secrets-live-online-in-databases-of-ruin.ars?utm_source=microblogging&utm_medium=arstch&utm_term=Law%20&%20Disorde)



## Netflix Needs to Put "Privacy Risks" in Their Queue

Published on Center for Democracy & Technology (<https://cdt.org>)

---

r&utm\_campaign=microblogging