**Group name:** Curating Data Breach Notifications Across States

**Group members:** Peter Cellier

**Topic:** Data breaches affecting the personal information of individuals, consumer protection.

**Target audience:** Policymakers interested in regulating data privacy and intervening after major breaches; journalists interested in tracking trends in data privacy and breaches across states; people, users, consumers, etc. who are affected by data breaches and who want to understand their risk by comparing information on data breaches across states.

**Goals for the protocol:** This project will attempt to curate information/data on data breaches reported by multiple states (at least 2) to create a new, curated database of data breach notifications across states. All 50 states plus the District of Columbia (and possibly territories, etc.) require that companies, organizations, and other entities that collect and store personal information notify the state government (usually the attorney general's office) of any "major" data breaches that compromise the personal information of that state's residents (usually above a threshold number, like 500 residents affected). Currently, these data are often siloed in state attorney general websites, sometimes in interactive tables ([see Oregon's page for an example](https://justice.oregon.gov/consumer/databreach/)) but more commonly in static PDF reports on annual data breach statistics and trends ([see California's page for an example](https://oag.ca.gov/privacy/databreach/reporting)). Some states with more progressive open data policies, such as Washington, allow users to access data breach information through the state's open data portal, and might provide different cuts and/or visualizations of the data in different formats (see [here](https://data.wa.gov/Consumer-Protection/Data-Breach-Notifications-Affecting-Washington-Res/sb4j-ca4h/about_data), [here](https://data.wa.gov/Consumer-Protection/Instances-of-PI-Breached-by-Type/cnur-mzu7), and [here](https://data.wa.gov/Consumer-Protection/Data-Breaches-by-Business-Type/bhqz-yv73). for three examples from Washington state's open data portal).

**Some goals I have for the protocol, in addition to serving the users described above, include:**

Gain experience curating, organizing, and combining data from different sources and formats (e.g., CSV files over here, PDFs over there, maybe some web scraped data over there),

Develop a user-friendly protocol that people could potentially use and implement beyond me and this project, including the users listed above, and

Practice different curational techniques, including those covered in this course and also possibly tools like (web) scraping to gather data from PDFs or static web pages.

**Group name:** Curating Data Breach Notifications Across States  
**Group members:** Peter Cellier  
**Topic:** Data breaches affecting the personal information of individuals, consumer protection.  
**Target audience:**Policymakers interested in regulating data privacy and intervening after major breaches; journalists interested in tracking trends in data privacy and breaches across states; people, users, consumers, etc. who are affected by data breaches and who want to understand their risk by comparing information on data breaches across states.  
**Goals for the protocol:**This project will attempt to curate information/data on data breaches reported by multiple states (at least 2) to create a new, curated database of data breach notifications across states. All 50 states plus the District of Columbia (and possibly territories, etc.) require that companies, organizations, and other entities that collect and store personal information notify the state government (usually the attorney general's office) of any "major" data breaches that compromise the personal information of that state's residents (usually above a threshold number, like 500 residents affected). Currently, these data are often siloed in state attorney general websites, sometimes in interactive tables ([see Oregon's page for an exampleLinks to an external site.](https://justice.oregon.gov/consumer/databreach/)) but more commonly in static PDF reports on annual data breach statistics and trends ([see California's page for an exampleLinks to an external site.](https://oag.ca.gov/privacy/databreach/reporting)). Some states with more progressive open data policies, such as Washington, allow users to access data breach information through the state's open data portal, and might provide different cuts and/or visualizations of the data in different formats (see [hereLinks to an external site.](https://data.wa.gov/Consumer-Protection/Data-Breach-Notifications-Affecting-Washington-Res/sb4j-ca4h/about_data" \t "_blank), [hereLinks to an external site.](https://data.wa.gov/Consumer-Protection/Instances-of-PI-Breached-by-Type/cnur-mzu7" \t "_blank), and [hereLinks to an external site.](https://data.wa.gov/Consumer-Protection/Data-Breaches-by-Business-Type/bhqz-yv73" \t "_blank) for three examples from Washington state's open data portal).

Some goals I have for the protocol, in addition to serving the users described above, include:

* Gain experience curating, organizing, and combining data from different sources and formats (e.g., CSV files over here, PDFs over there, maybe some web scraped data over there),
* Develop a user-friendly protocol that people could potentially use and implement beyond me and this project, including the users listed above, and
* Practice different curational techniques, including those covered in this course and also possibly tools like (web) scraping to gather data from PDFs or static web pages.