大成 DENTONS

Privacy regulation of de-identification and anonymization is a game-changer - How to make it work in practice

Did you know Dentons produces podcasts on a variety of topics?

- Arbitration
- Employment and Labour Law
- Intellectual Property
- Tax
- Transformative Technologies and Data
- Entertainment & Media Law
- Life Sciences and Health Care
- Women in Leadership and Entrepreneurship
- Toronto Business Insights
- Smart Cities
- Agribusiness
- Banking and Finance
- Mining

Visit our Podcast page and subscribe https://www.dentons.com/en/insights/podcasts

Privacy regulation of de-identification and anonymization is a game-changer: How to make it work in practice

We also have blogs in various areas

- Insurance
- Regulatory
- Venture Technology
- Drone Regulation
- Employment and Labour
- Privacy and Cybersecurity
- Technology, New Media and IP Litigation
- Canadian Occupational Health & Safety
- Tax Litigation
- Commercial Real Estate
- Commercial Litigation
- Transformative Technologies and Data
- Entertainment & Media
- Mining
- Doing Business in Canada

Visit our Blogs and Resources page at dentons.com/en/insights/blogs-and-resources

大成 DENTONS

Privacy regulation of de-identification and anonymization is a game-changer - How to make it work in practice

Thursday, November 4, 2021

Presenters

Chantal Bernier
National Practice Lead, Privacy and Cybersecurity
+1 613 783 9684
chantal.bernier@dentons.com

Dr. Khaled El Emam
CEO, Replica Analytics
kelemam@replica-analytics.com

Setting the stage

- For the first time in Canada's legal history, "de-identified"and "anonymized" information are defined in and regulated by general privacy legislation: Québec Bill 64 an *Act to modernize legislative provisions as regards the protection of personal information* provides these definitions:

- "*Information is de-identified if it no longer allows the person concerned to be directly identified*."

- "Information anonymized if it irreversibly no longer allows the person to be identified directly or indirectly."

The historical first

- Anonymized information “ is brought within the scope of privacy law - beyond the minimal reach of current provisions of personal health information protection on “de-identify”/ “*anonymiser*”:
  - the information must be anonymized according to “*generally accepted best practices.”*
  - It must be used for “*legitimate and serious purposes*.”
    - Section 23 of the new *Act respecting the protection of personal information in the private sector.*
  - Anyone who identifies or attempts to identify a natural person using de-identified information without the authorization of the person holding the information or using anonymized information, may be subject to a fine.
    - Section 91 of the new Act.

Other legislative proposals: from former Bill C-11

- C-11, only covered “de-identified information” but through a definition that corresponds to Bill 64’s “anonymized information”:
  - “*de-identify means to modify personal information — or create information from personal information — by using technical processes to ensure that the information does not identify an individual or could not be used in reasonably foreseeable circumstances, alone or in combination with other information, to identify an individual*.”
- It would have allowed an organization to use personal information without consent to de-identify it. (proposed clause 20)
- It would have restricted its use to “Research and development” (section 21) and its sharing with specific organizations and for “socially beneficial purposes.” (proposed clause 39)

…From Modernizing Privacy in Ontario White Paper

- "de-identified information" would be:

  " *information about an individual that no longer allows the individual to be directly or indirectly identified without the use of additional information.*" It could be used with appropriate safeguards.

- "anonymized data" would be:

" personal information that has been altered in such a way that it is no longer identifiable in relation to an individual."

- "*to incentivize the use of anonymized data*" it would be removed from privacy rules altogether."

The policy issues behind the legislative proposals:

- “De-identified” and “anonymized data” are necessary for innovation.

BUT

- The risk of re-identification must be addressed.
- Anonymization is no longer possible.

The facts to ground the legislative policy:

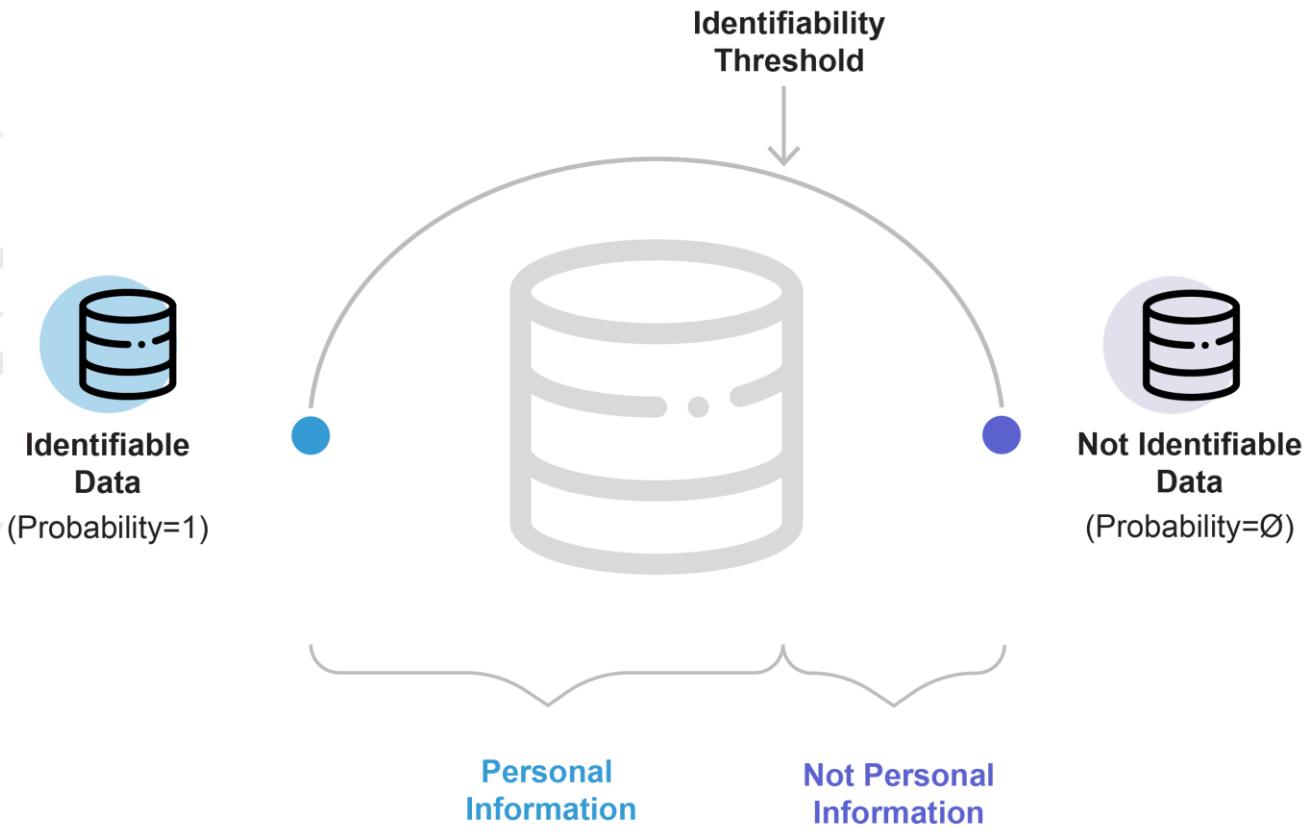
- Is anonymization truly impossible now?
- What is the real risk of re-identification?
- What are the "*generally accepted best practices*" of de-identification and anonymization?

A Brief Introduction to De-identification

Khaled El Emam

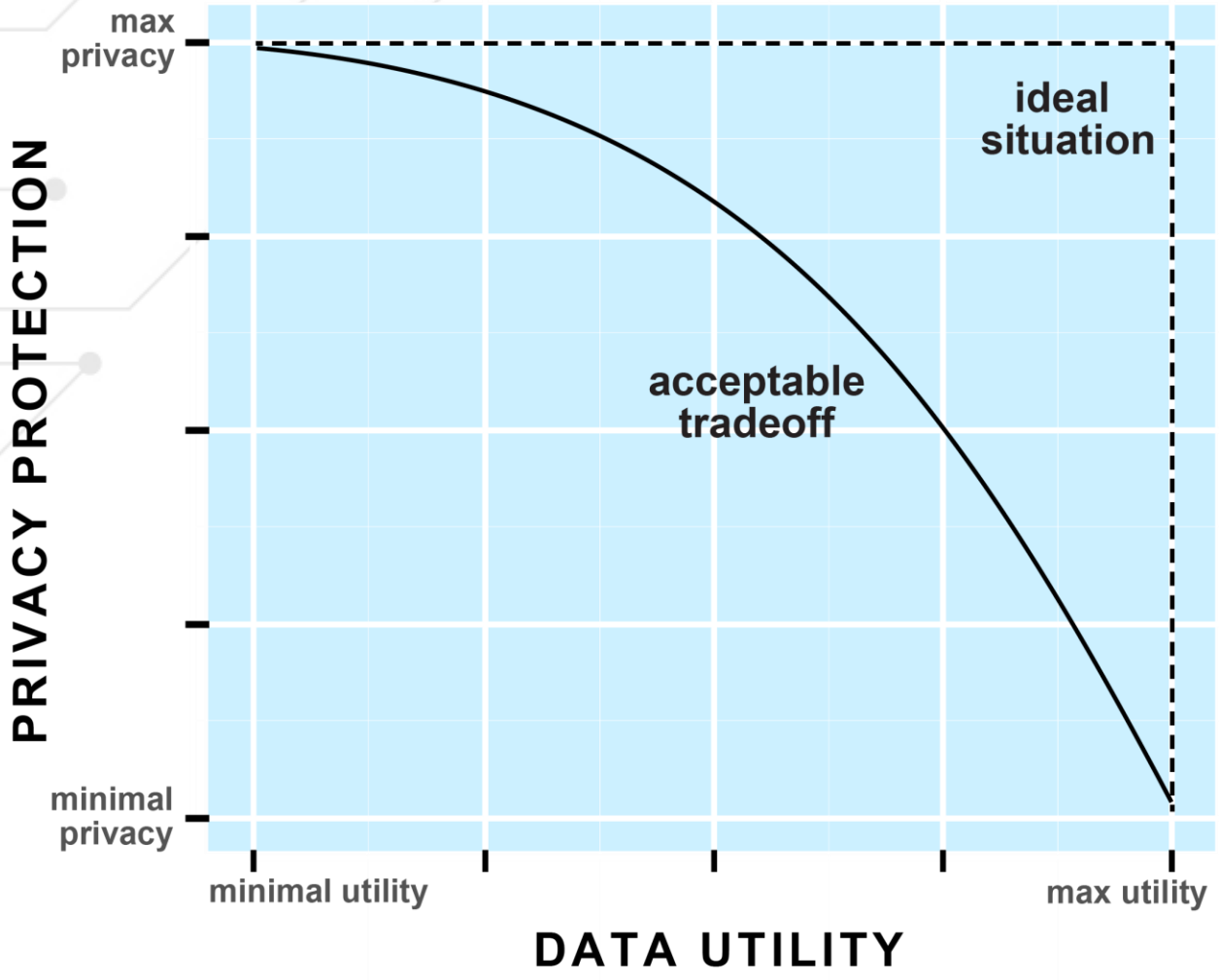
kelemam@replica-analytics.com

Identifiability spectrum and risk thresholds



Replica Analytics

There is a trade-off between the extent of privacy protection and data utility



Replica Analytics

A common approach that has worked well in practice is risk-based anonymization

- Generalization
- Suppression
- Addition of noise
- Microaggregation

- Security controls
- Privacy controls
- Contractual controls

Replica Analytics

Commonly mentioned PETs

01

RISK-BASED DE-IDENTIFICATION

Using methods like k-anonymity to measure re-identification risk, and data transformations are combined with controls to manage overall risk.

02

DATA SYNTHESIS

Models are built from data, and these are used to generate new datasets that retain the statistical patterns.

03

FEDERATED ANALYSIS/SECURE MULTIPARTY COMPUTATION

Computations are distributed among multiple parties, either as data sources or as computing nodes, or both.

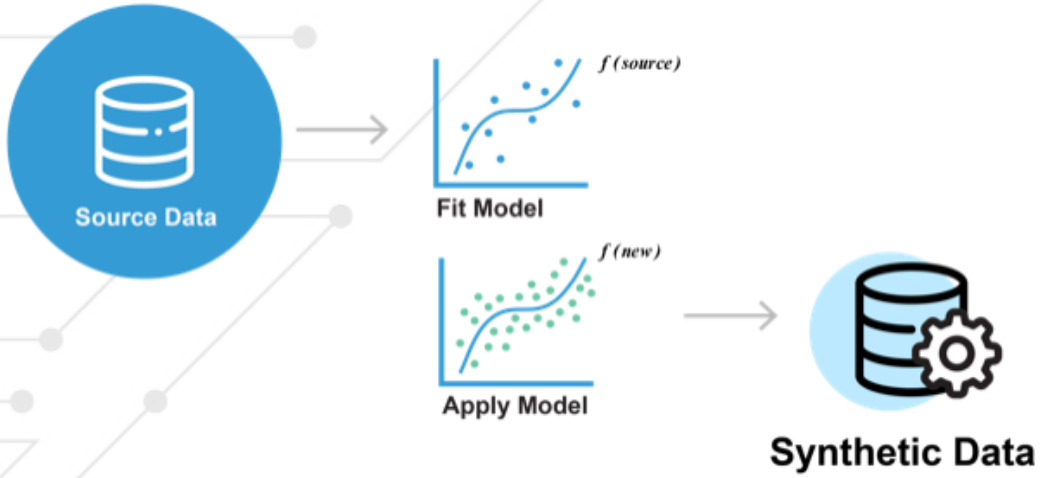
04

DIFFERENTIAL PRIVACY

Interactive system that adds noise to the results of interactive queries to manage re-identification risk.

Replica Analytics

The Synthesis Process



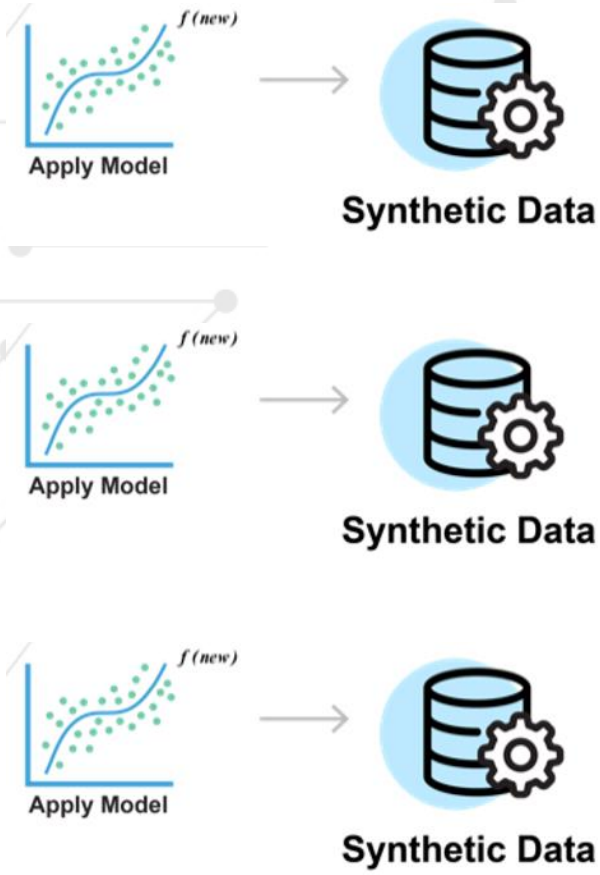
COU1A	AGECAT	AGELE70	WHITE	MALE	BMI
United States	2	1	1	1	33.75155
United States	2	1	1	0	39.24707
United States	1	1	1	0	26.5625
United States	4	1	1	1	40.58273
United States	5	0	0	1	24.42046
United States	5	0	1	0	19.07124
United States	3	1	1	1	26.04938
United States	4	1	1	1	25.46939

Additional Clarifications

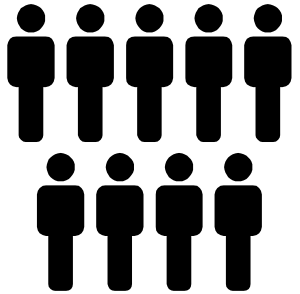
- The source datasets can be as small as 100 or 150 patients. We have developed generative modeling techniques that will work for small datasets.
- The source datasets can be very large – then it becomes a function of compute capacity that is available.
- It is not necessary to know how the synthetic data will be analyzed to build the generative models. The generative models capture many of the patterns in the source data.

Replica Analytics

A simulator exchange allows data to be made available without sharing actual data



Data Consumers



Additional Clarifications

- The simulators would not be given to the data consumers – they would only have access to them through an interface.
- This access would be monitored and throttled to reduce the risk of attacks on the models.
- Data consumers would also need to agree to terms of use around the access to the simulators.

Replica Analytics

Thank you

Chantal Bernier
National Practice Lead, Privacy and Cybersecurity
+1 613 783 9684
chantal.bernier@dentons.com

Dr. Khaled El Emam
CEO, Replica Analytics
kelemam@replica-analytics.com