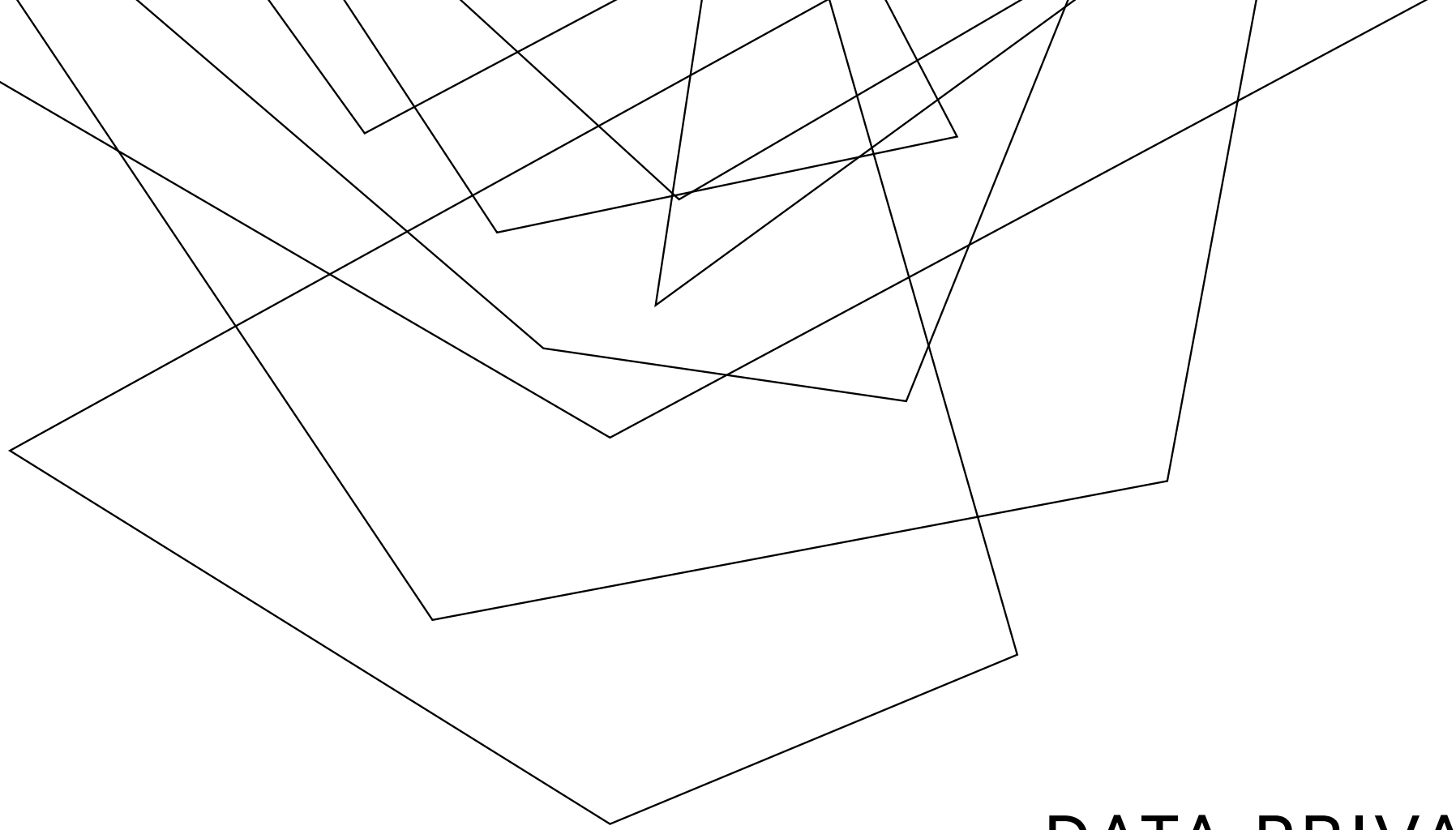


DATA PRIVACY

YOAN BOLDUC

My experience

- 16 years of experience with SAS Products as Administrator, Architect and Manager.
- Strong interest in Cybersecurity
- Different areas of activity: logistics, education, healthcare and finance.

HOW TO KEEP SENSITIVE DATA PRIVATE

HANDLING PRIVATE INFORMATION WHEN DEALING WITH LARGE DATASETS

WHAT IS SENSITIVE DATA?

It depends on the context and who you talk to:

- Health record
- Customer list
- Recent transactions
- A customer classification
- Generally speaking, sensitive data is information that, if exposed, could cause significant harm, discrimination, or financial loss to an individual or organization

LEGAL CONTEXT

Canada:

- PIPEDA (Personal Information Protection and Electronic Documents Act)

Europe:

- GDPR (General Data Protection Regulation)

United States:

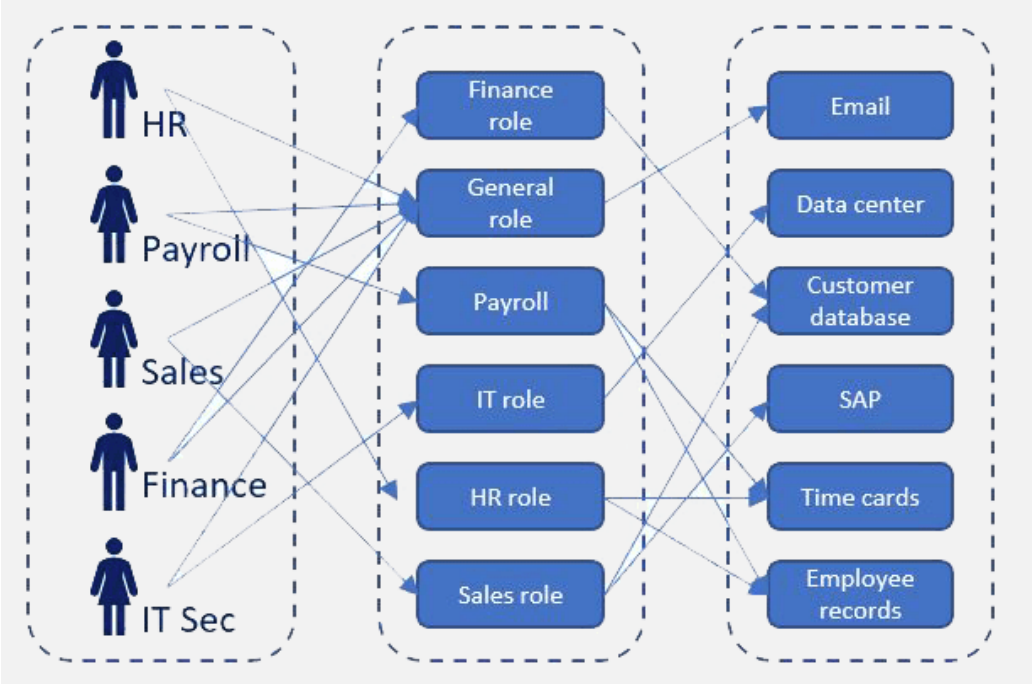
- HIPAA (Health Insurance Portability and Accountability Act)
- FERPA (Family Educational Rights and Privacy Act)
- GLBA (Gramm-Leach-Bliley Act)
- CCPA (California Consumer Privacy Act)
- CPRA (California Privacy Rights Act)

THE BASICS: ACCESS CONTROL

Techniques for connecting

- Users are added to security groups.
- Permissions are granted to security groups on resources (databases, folders, systems)
- Some approval process is implemented for new requests.
- Single level of governance.

THE BASICS: RBAC



- Multiple levels: role and technical groups
- Role group: represents a business function
- Technical group: represents a capability on a set of resources (servers, folders, databases, etc...)
- Users are members of role groups
- Roles groups are members of technical groups.
- Permissions on objects are assigned to the technical groups.
- Multiple level of governance with implied trust.
- Possibility of punctual group assignments

GOOD IT PRACTICES

General good practices in IT:

- Keep your software up to date.
- Use modern firewalls to segregate networks and monitor traffic.
- Implement Endpoint Detection and Response tools.
- Use a SIEM (Security Information and Event Management tool) to corelate logs and have a big picture view.
- Use vulnerability scanners (Nessus, OpenVAS, Wiz, Aqua Security, etc...)

FOOD FOR THOUGHTS

Techniques for connecting

- Make eye contact with your audience to create a sense of intimacy and involvement
- Weave relatable stories into your presentation using narratives that make your message memorable and impactful
- Encourage questions and provide thoughtful responses to enhance audience participation
- Use live polls or surveys to gather audience opinions, promoting engagement and making sure the audience feel involved

FOOD FOR THOUGHTS

How many data elements do you need to identify someone?

- One: Social Insurance Number, Health Card Number, Driver License Number.
- Two: First name and last name? First name and home address?
- Three: ZIP Code, birth date and gender (Latanya Sweeney: works for 87% of americans)

FOOD FOR THOUGHTS

What amount cases that are on the extremes?

- Like age? ~366 000 people are aged over 90... Only 11k aged over 100... Add a street name, a village name or maybe just their last name...
- According to Forebears: Canada has 858,996 unique surnames that is an average frequency: 43 people per surname.
- What about ethnicity?
- Clearly, context matters.

HOW TO KEEP DATA PRIVATE BUT ENABLE ANALYTICS

Different techniques

- Restricting the available data (smaller data mart)
- Derived values:
  - Replace date of birth with age, age range, etc...
  - Replace postal code with FSA or some geographic division.
  - Replace specific sales figures by range
- Value substitution: change name and last name
- Key hashing

IT MAY STILL BE INSUFFICIENT

In some case, a large number of abstracted facts can lead to reidentification:

- Multiple timestamps events (list of visits to a healthcare specialist).
- Relationships to other items (cars owned by a person)
- Grades in multiple courses
- Recent transactions on checking account
- List of mortgages

IT MAY STILL BE INSUFFICIENT

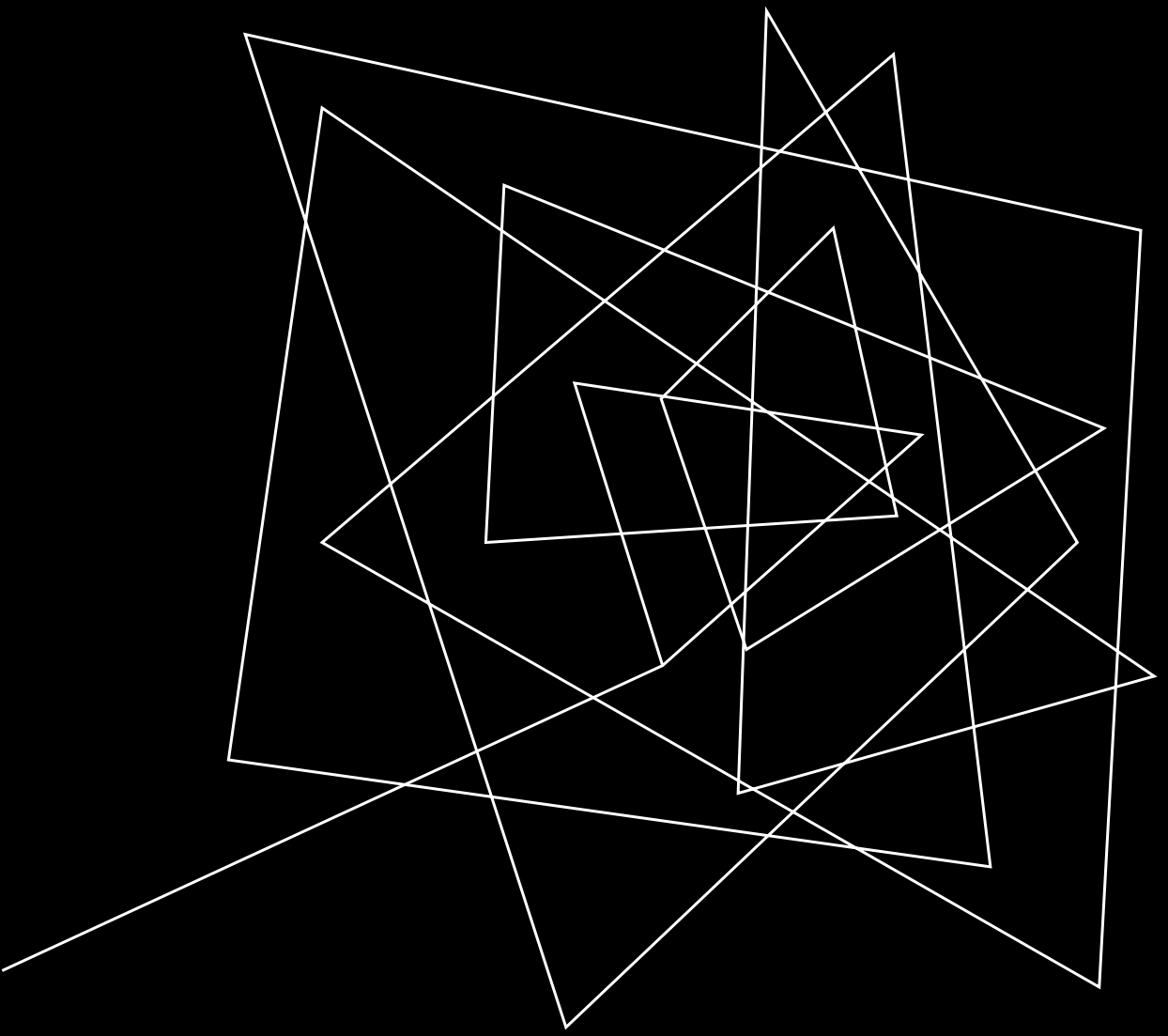
Linking multiple distinct datasets

- Marketing
- Finance/Accounting
- HR
- Sales

THE SOLUTION

On-demand purpose-built environments and datasets

- Analyst identify data requirements and submits a request.
- Request is analyzed and assigned a risk rating.
- Risk is evaluated and approved or rejected.
- An on-demand datamart is produced:
  - Hashed keys, abstracted values, minimal attributes, etc...
- The datamart is made available in a single-use restricted environment.
- The result is reviewed and vetted before being extracted from the environment.



QUESTIONS???