



An Introduction to Data Classification

Finding the Important Stuff

David Bullas, Senior Technical Specialist

Agenda

- Introduction
- Why classify data?
- Common Compliance Regulations
- How does it work?
- What next?



The background of the slide is a solid dark blue. In the center, there is a large white circle. Surrounding this white circle are several concentric circles in different colors: a thin cyan circle, a purple circle, a thick white circle, and a thin green circle. The word "Introduction" is written in a bold, dark blue font in the center of the white circle.

Introduction

Data classification is the process of sorting and categorizing data into various types, forms or any other distinct class. Data classification enables the separation and classification of data according to data set requirements for various business or personal objectives. It is mainly a data management process.

- Techopedia

What does that mean?

Understand your
Rules

Identify where
sensitive
information is

Put that data in the
right place

Repeat

Common Terms

- Classification
- Sensitive Data
- PII
- Compliance
- Governance



Classification: The process of separating data





Sensitive Data: Information protected against unwarranted exposure

Examples:

- Credit Cards
- Medical Records
- Home Phone Numbers
- Email Addresses
- Minister Memos
- Project Status Updates



PII: Personally Identifiable Information



Transcript Request

Office Use Only

Notes

- Transcripts are normally produced in three to five business days.
- Unofficial Transcripts are available on Bear Tracks at www.beartracks.ualberta.ca for all current and/or former students who were in attendance since Fall Term 2002.
- Please print clearly.

Personal Information

Surname on Record		Other Names in Full	
Address			
City	Province	Country	Postal Code

Do you want the above address updated to our database? ☐ Yes ☐ No

Student ID Number	Date of Birth	Former Name (if applicable)
[Area Code] Home Telephone	[Area Code] Business and/or Cell Phone Number	Email Address

Indicate most recent faculty and year you attended.

Faculty	Last Year Attended
---------	--------------------

Have you attended postsecondary studies in Camrose prior to September 2006? ☐ Yes ☐ No If Yes, fax the completed request form to 780.679.1527.

Student Signature	Date
-------------------	------



Compliance:

?

We could talk about this one all day...

But basically: Making sure the right people are following the right procedures, especially in protecting data



Governance:

**Establishment
of policies to
achieve proper
compliance**



Photo: Sigi Johanne Malheu, Redwan HAE © CSO/CSG 2017

Her Excellency the Right Honourable
Julie Payette
Governor General and
Commander-in-Chief of Canada



Son Excellence la très honorable
Julie Payette
gouverneure générale et
commandante en chef du Canada

Governance is also...

... the piece that nobody wants to look at until something goes wrong, and then it's scrutinized within an inch of its life.

So you need to make sure it's tight in the good times to protect you in the bad.

The background of the slide is a solid dark blue. It features a series of concentric circles centered on the slide. From the center outwards, the circles are: a white circle, a thin cyan circle, a thin purple circle, a thick white circle, and a thin green circle. The text is centered within the white circle.

**Why classify
data?**

There are lots of files out there



Only some of them matter



But they aren't always easy to find



From the time information is created until it is destroyed, it should be labeled with a classification designation to ensure it is protected, stored and managed appropriately. ”

Brian Evans, Tom Walsh Consulting

The background of the slide is dark blue. It features a series of concentric circles: a thin green circle, a thick white circle, a thin purple circle, a thin cyan circle, and a thick white circle. The text is centered within the white space of the innermost thick circle.

Common Regulations that require Data Classification



Compliance Regulations: PIPEDA

Personal Information Protection and Electronic Documents Act (PIPEDA)

Sets the ground rules for how private-sector organizations collect, use, and disclose personal information in the course of for-profit, commercial activities across Canada.

It also applies to the personal information of employees of federally-regulated businesses such as banks, airlines, telecommunications companies



Compliance Regulations: The Privacy Act

Relates to a person's right to access and correct personal information that the Government of Canada holds about them.

Also applies to the Government's collection, use and disclosure of personal information



Compliance Regulations: The Federal Banking Act

Regulates the use and disclosure of personal financial information by federally regulated financial institutions.



Compliance Regulations: PCI DSS



The Payment Card Industry Data Security Standard (**PCI DSS**) is an information security standard for organizations that handle credit cards.



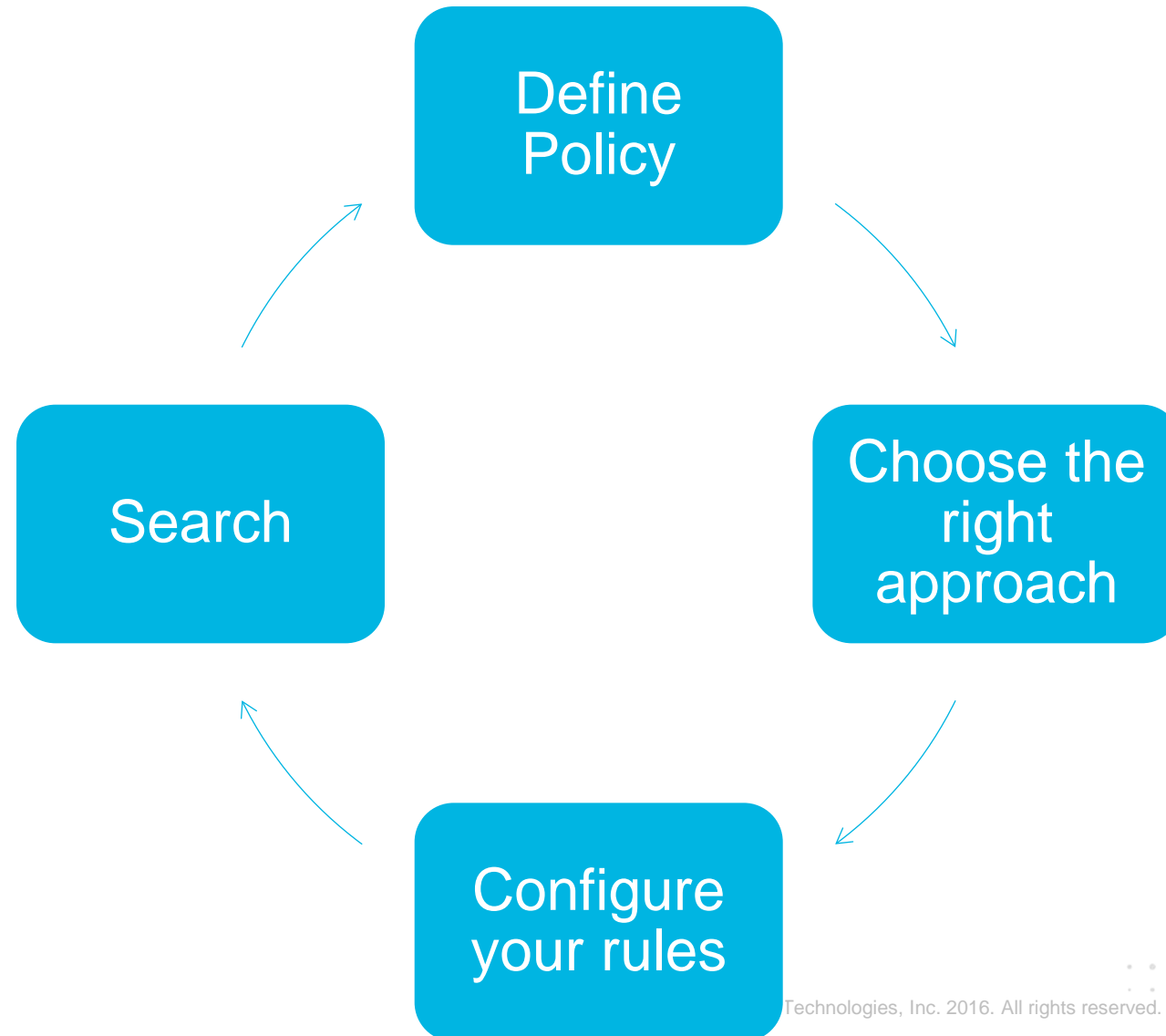
Compliance Regulations: GDPR

The **General Data Protection Regulation (GDPR)** is a legal framework that sets guidelines for the collection and processing of personal information of individuals within the European Union (EU).

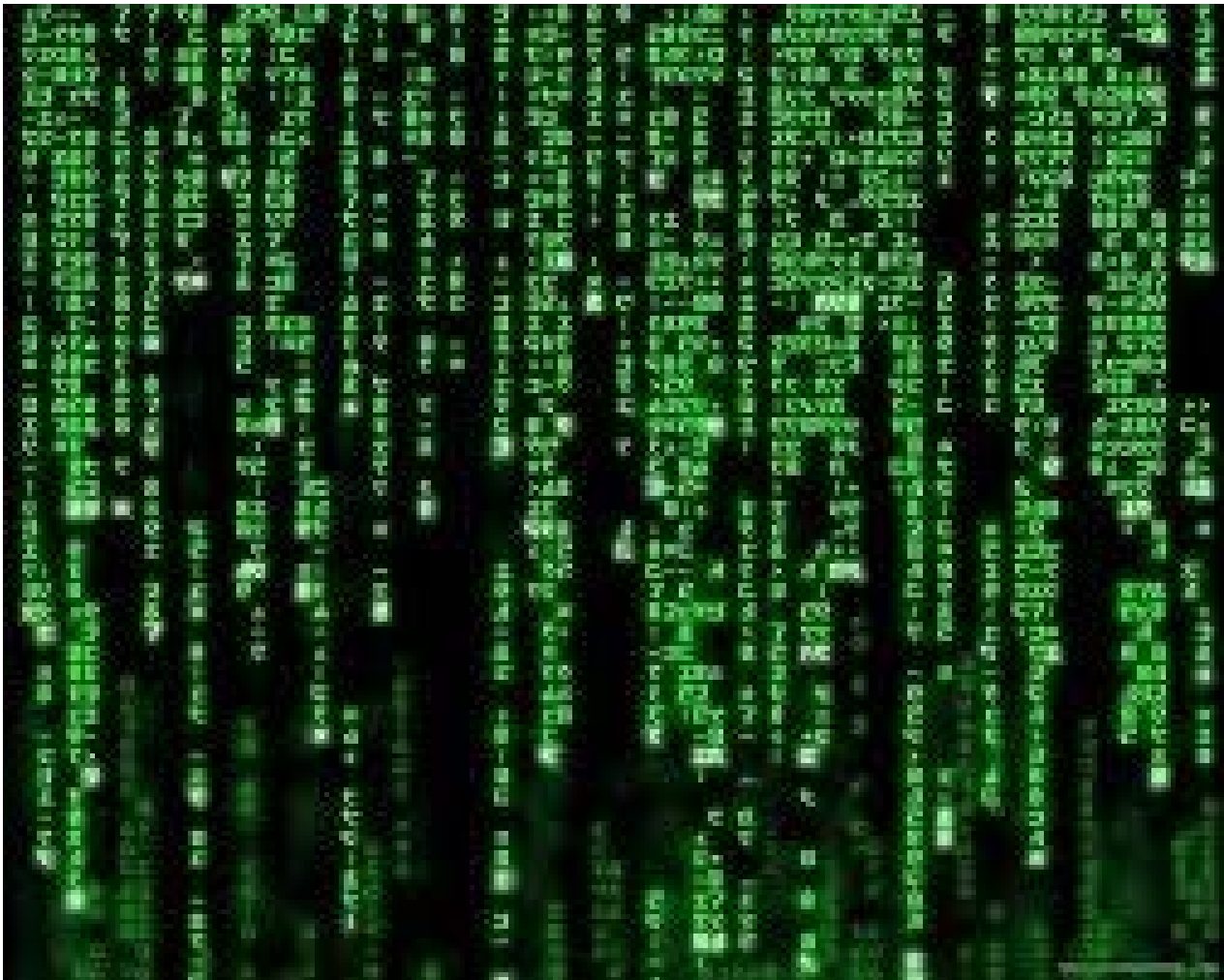
The background of the slide is a solid dark blue. In the center, there is a large white circle. Surrounding this white circle are several concentric circles in different colors: a thin cyan circle, a thicker purple circle, a thin white circle, and a very thin green circle. The text is centered within the white circle.

How Data Classification Works

Data Classification Steps



Defining Policy



Defining Policy

Enumerate the types of data that matter to your business

- Compliance Data
- Project Proposals
- Employee Reviews
- CAD Drawings

You don't have to know it all to start!

Decide on your approach



Manual

Content Analysis

Behaviour



1. Manual

- This is the “do it yourself” option.
- Get the users to classify their own files
- Checks the checkbox for classification policies
- Relies on your knowledge of your data and the rules

Is it or Isn't it?

4111 1111 1111 1111

Is it or Isn't it?

4111 1111 1111 1111

YES!

Is it or Isn't it?

4111 1111 1111 1111

YES!

4111-1111-1111-1111

Is it or Isn't it?

4111 1111 1111 1111

YES!

4111-1111-1111-1111

YES!

Is it or Isn't it?

4111 1111 1111 1111

YES!

4111-1111-1111-1111

YES!

4111.1111.1111.1111

Is it or Isn't it?

4111 1111 1111 1111

YES!

4111-1111-1111-1111

YES!

4111.1111.1111.1111

YES!

Is it or Isn't it?

4111 1111 1111 1111

YES!

4111-1111-1111-1111

YES!

4111.1111.1111.1111

YES!

4111111111111111

Is it or Isn't it?

4111 1111 1111 1111

YES!

4111-1111-1111-1111

YES!

4111.1111.1111.1111

YES!

4111111111111111

MAYBE?

Is it or Isn't it?

4111 1111 1111 1111

YES!

4111-1111-1111-1111

YES!

4111.1111.1111.1111

YES!

4111111111111111

MAYBE?

4111 1111 1111 1112

Is it or Isn't it?

4111 1111 1111 1111

YES!

4111-1111-1111-1111

YES!

4111.1111.1111.1111

YES!

4111111111111111

MAYBE?

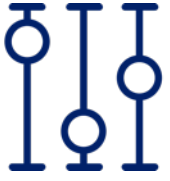
4111 1111 1111 1112

NO!



2. Automated Content Analysis

- Software-driven search for specific content or file attribute
- Relies on rules that look for
 1. Keywords
 2. Patterns
 3. Regular Expressions
 4. File Name, Size, Extension
- Relies on regular review and rule refinement



3. Behaviour

- Search for sensitive content based on who is accessing it:
 - I. If more than 75% of users are members of the Finance department, classify the file as a finance file
 - II. If more than 90% of users are members of the AD group HR, classify the file as an HR file
- Relies on identity information and activity data

Define Rules



Define Rules

- Be specific (we don't want a lot of rework from false positives!)
- Combine terms together
- Use automatic validation (Luhn, Area Codes, Expert Knowledge) where you can
- Refine, Refine, Refine
- This is kind of like the “search” part of “search and replace”

Find the word “he”, change it to “she”

Simple – but be careful that you don't change “the” to “tshe”

Search!



The background of the slide is a solid dark blue. In the center, there is a large white circle. Surrounding this white circle are several concentric circles in different colors: a thin cyan circle, a purple circle, a white circle, and a thin green circle. The text "Now what? Govern!" is centered within the white circle.

**Now what?
Govern!**

Comprehensive Identity Governance

Protect access to **all** applications and data – **on-premises and in the cloud**

- ✓ Access Request
- ✓ Access Certification
- ✓ Provisioning Workflow
- ✓ Access Policies
- ✓ User Risk-based Modeling
- ✓ Password Management
- ✓ Data Classification
- ✓ Activity Monitoring
- ✓ Permission Analysis



Benefits

- Greater visibility into access risks
- Centralize all access to applications and data
- Reduced complexity by providing a consistent set of controls

One Identity, all data

Complete Security Visibility

For each identity, get complete coverage of their structured and unstructured data access

Accounts with access to sensitive data are tagged with a red-letter “S”

The screenshot shows the 'View Identity Catherine Simmons' page in the SailPoint Identity Center. The 'Entitlements' tab is active, displaying a table of permissions. A blue arrow points to the 'X:\CorpData\Finance' attribute, and another blue arrow points to the 'Modify' entitlement, with a text label 'Modify Permission on a Shared Folder'.

Attribute	Entitlement
X:\CorpData\Finance	Modify
RESPONSIBILITIES	Treasury Inquiry
RESPONSIBILITIES	Treasury Settlements Administrator
RESPONSIBILITIES	E-Business Intelligence Alerting
RESPONSIBILITIES	Internet Expenses Reporting and Analysis
memberOf	All_Users
memberOf	FinanceUsers
groups	SatelliteOfficesRDP
groupmbr	ProjectManager



Thank You