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?
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**

<table>
<thead>
<tr>
<th>transcription_request</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Transcript Request</strong></td>
</tr>
</tbody>
</table>

**Notes**
- Transcripts are normally produced in three to five business days.
- The local Transcript Forms are available on the Office of the Registrar's website at [www.registrar.ualberta.ca](http://www.registrar.ualberta.ca) for all current and former students.
- Please print clearly.

**Personal Information**
- **Surname on Record**
- **Other Names as 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)**

**Home Telephone**
- **Business and/or Cell Telephone**
- **Email Address**

**Indicate most recent faculty and year you attended.**
- **Faculty**
- **Year**

**Have you attended postsecondary studies in Canada prior to September 2002?**
- [ ] Yes
- [ ] No
  - If Yes, fax the completed request form to 780.492.5257.

**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
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.
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
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.
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).
How Data Classification Works
Data Classification Steps

1. Define Policy
2. Choose the right approach
3. Configure your rules
4. Search

The process is cyclical, allowing for repeated cycles of policy definition, approach selection, rule configuration, and search.
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

4111-1111-1111-1111

YES!
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   YES!
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?
<table>
<thead>
<tr>
<th>Format</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>4111 1111 1111 1111</td>
<td>YES!</td>
</tr>
<tr>
<td>4111-1111-1111-1111</td>
<td>YES!</td>
</tr>
<tr>
<td>4111.1111.1111.1111</td>
<td>YES!</td>
</tr>
<tr>
<td>4111111111111111</td>
<td>MAYBE?</td>
</tr>
<tr>
<td>4111 1111 1111 1112</td>
<td></td>
</tr>
</tbody>
</table>
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

Image: https://xkcd.com/208/
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!
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”
Thank You