

Compliance Guardian Metadata Classification Capabilities

What is metadata?

Metadata is “information about information”. This comes in the form of metatags, which provide information about documents such as title, author, subject, key words, and date created. Think about metadata as the electronic equivalent to cards in card catalogs found at libraries.

Why is metadata valuable?

The beauty of metatags is that they can also provide information that may be useful for searching for that document or appropriately classifying its sensitivity. Consistent, quality metadata tagging by individuals can improve search as well as used organize and structure content so that it can be more widely disseminated. However, that should not stop organizations from augmenting metadata based on machine derivable metadata. Location, context, and the ability to evaluate the data by filtering out “noise words” can aid in adding meaningful information on the fly to either validate or invalidate human-entered data. Validation of the data is as important as the data itself. This allows organizations to improve search and discovery across their data sets, which ultimately empowers users to find the right content quickly in order to improve collaboration throughout the enterprise.

Challenges of Classifying Content with Metadata

Classification with metadata is putting content into classes or categories. The metadata can be embedded into the content, or it can be externalized into a form of managed metadata that is connected, aside, or points to the actual content. Additionally, the content can then be examined to add this metadata, which can be machine generated, human generated, or both. There are pros and cons to every one of these methods that organizations must consider when implementing metadata schemes. First, content-based automatic classification systems are very powerful but they can be fooled into saying the content is something that it is not. User assignment or cataloging could also have the same problem of the user either intentionally tricking the classification process or just setting the value incorrectly.

Many people refer to the practice of classification as Taxonomy. This Semantic, usage of metadata classification approach, lends itself to a highly ordered set of data and useful content management if implemented properly. Additionally, it gives users a superior way to use and interact with data and people. Within the Taxonomy, we can also have controlled vocabularies and/or user-generated vocabularies.

Most organizations possess thousands of pieces of content, including documents in common use as well as sensitive and classified content such as customer information, employee information, intellectual property, and organizational sensitive information. There is a high risk of having hard-to-use content or exposing sensitive content to the wrong sets of eyes. Organizations must properly catalog all of their information to ensure the content is only available and accessible to the people who should have it and protected from the people who should not.

Compliance Guardian Metadata Classification Capabilities

Compliance Guardian Classification System

The best solution for classification with metadata is a system combined with user collaboration to properly catalog content for increasing searchability, usability, and security, while having the ability to dispose of content that breaks organization classification rules. AvePoint [Compliance Guardian](#) helps organizations handle the creation of metadata, allowing users to create metadata automatically or manually, as well as take disposal-related actions in accordance with organization's specific needs.

Through standardization Compliance Guardian allows you to consolidate and improve knowledge management, as well as fully implement capabilities to own and execute governance. Compliance Guardian ensures that "control" doesn't have to mean "limit". Tag content, and evaluate and act on existing metadata and file properties in order to appropriately classify that information before entering it into the data lifecycle.

Trust and Verify

Many enterprise organizations implement a concept of "trust and verify." Trust that your users are the best people to determine what a document is about and empower them to tag and classify the document. On the other hand, verify and validate that they have done so properly.

Compliance Guardian supports the concept of user-assisted metatagging, and also can automatically tag documents as they are entered and/or scanned in SharePoint. As users upload content to SharePoint, Compliance Guardian also can recommend an appropriate tag to them based on autoscanning against pre-determined policies. For organizations utilizing file shares, Compliance Guardian supports the implementation of autotagging for documents residing in those repositories.

Technical Overview of Compliance Guardian Classification System

- Flexibly work with one or more taxonomy and/or controlled vocabularies, content, and structures, as well as enhanced site properties related to the content to be classified
- Tag content with embedded metatags (on file shares or within SharePoint) or system metadata, including full support for SharePoint Managed Metadata, or stand columns via a Metadata Classification Engine
- Support user-driven (user assisted) tagging through native Ribbon-based document panels or automated tagging for SharePoint environments
- Support automated tagging for file share environments
- Identify and resolve inconsistencies between document and SharePoint metadata, and synchronize to ensure metadata consistency across platforms
- Based on content classification, assign permissions, route to the appropriate location, and block or quarantine to comply with information governance policies
- Be able to classify content based on undefined relationships in data but have a rule language capable of customization to do the same
- Facilitate content search, data loss prevention, content management, and risk management
- Move, manage, delete, or quarantine content based on classification

Compliance Guardian Metadata Classification Capabilities

Author and Automatic Classification Interactions

Function	Description
Supports Embedded Metadata	Allow the embedding of data into documents so that the metadata travels with the data.
Supports Custom Taxonomies	Allow users of content management systems to create their own taxonomy and implement controlled vocabularies to create a more refined approach of data cataloging.
Auto Classification	Allow for content to be automatically classified based on complex rule types, taking into account advanced vocabulary support.
Allow User to Enter Metadata	Users can add metadata to the document.
Ensure Quality of User-selected Metadata	Since users are allowed to choose metadata, a classification system should be able to identify errors and correct if necessary.
Ensure Quality of User-entered Metadata	In freeform Text Fields, users can enter metadata and the systems need to validate that the text does not violate policy and matches the document.
Transfer Security-enabled Metadata	Encryption level must be determined based on sensitivity of content to provide content with site-specific classification.

Classification Rule Types

Rule Type	Description
Does Text Exist	Location of Text in a document or in document metadata as a way that can be used to assign a metadata element (Tag) value.
Conditional Text	Like the above rule, Location of Text in a document or in document metadata as a way that can be used to assign a metadata element (Tag) value. It provides a more complex way of looking at text relationships to determine classification.
Dictionaries	Another text-based rule to determine if the system can assign a controlled vocabulary term based on the existence of one or many words.
Element Validation	Whether user-entered metadata or structural elements, one can find meaning from element-based data.
Enhanced Elements	This rule type looks deeper into the structure of content to find information about elements that exist in elements of content to find meaning.
External Content	Searching for content within content that is actually only a reference to content existing on external sites.
RegEx	Location of regular expression match in a document or in document metadata as a way that can be used to assign a metadata element (Tag) value. It provides a powerful method to classify content based on pattern matching.
Conditional RegEx	Location of regular expression match in a document or in document metadata can trigger another regular expression pattern search as a way we can use to assign a metadata element (Tag) value.
Transport	Transport tests the usage and level of usage of the HTTPS protocol that is being used to serve content.
Cookie	Aligned or Connected Data attached to content can be tested to determine classification.
Custom	A custom check type allows the performance of external functions if a condition is found.
Batching/Super Rules	To combine one or more rule types in sequential logic to produce controlled vocabulary or simple classification outcomes.