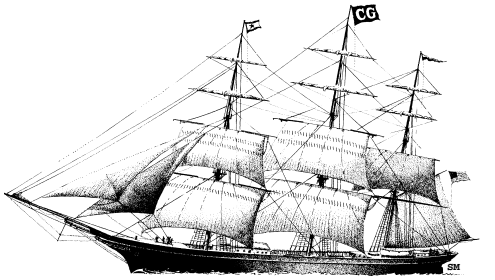
THE CLIPPER GROUP
Explorer™



Navigating Information Technology Horizons℠

Published Since 1996 Report #TCG2005076 October 31, 2005

The Role of Enterprise Data Classification

Analyst: Anne MacFarland

Management Summary

When you are building something complex, like an organization, you focus on hiring the right talents and experience, designing the relationships and governance, and getting everyone to work together. Titles and vocabulary are often an *ad hoc* afterthought. However, the titles and vocabulary that classify *who* does *what* are important to newcomers to the organization, who see only a sea of faces. In an organization of any size, few are the jobs that don't require a good understanding of the larger organizational structure.

Information Technology has been focused for the past few decades on *building* things through a bottom-up process of design, test, and add. One need only look at current infrastructure vocabulary, such as virtualization and Web Services, to realize that the coherence of such names to a larger audience was probably under-researched. Users have been told to learn enough about their IT environment to do their jobs, and not to try to understand the bigger picture. However, **anyone involved in the policymaking of governance, whether in IT or on the business side, needs a better big-picture top-down view of how the technology side fits together. In particular, they need to understand the extents and uses of the data that the enterprise generates, captures, and uses.** Knowing data as a sea of great capacity, even if that capacity is segmented by applications, is not good enough.

It is not good enough because information is not merely the vehicle by which business is done, but an asset with value - and with liabilities. The success of an enterprise depends on its use. Many enterprises have moved from a model where information is owned by departments to a model where information is shared – across the organization, and with partners. Deft use of information enables new products and services to be delivered with speed, care, and precision to customers. Quick access to information has high value in the sales cycle. **As information is shared, its treatment becomes less a matter of local policy and more a matter of corporate strategy.**

There is a need to make explicit an articulation of the roles bodies and flows of information play in the enterprise. The value of information is not absolute. It depends on the breadth of context in which it is relevant, the confidentiality of the information that may limit its use, and the period of time for which the information is valid. These will depend on how, why, and by whom the information was generated or captured. You cannot simply put it out there, index it, and search it - without ongoing management - and expect things to work well.

Enterprise Data Classification clarifies the links between information and the functions of an enterprise. The bulk and growth of data in enterprise systems requires it to make data useful for governance. For more details, read on.

IN THIS ISSUE

The Different Roles and Characteristics of Classification

Classification is - by its nature - comprehensive, structured, and targeted at a particular goal, such as a learnable hierarchy to speed finding or a rational hierarchy to promote understanding of the whole and the relationship of its parts.

Search and its *indexing* have accelerated and enormously enriched the speed of finding. Enterprise Data Classification is focused not on finding, but on understanding the whole of enterprise data as a set of information classes characterized by their use in the enterprise. For it is the business *use* of information, not just its *subject* or its *age*, that determines the information management policies that govern that use and, by extension, where and how the information is stored over its lifetime (ILM).

How an enterprise classifies data will depend on what it does, what kinds of data it captures and produces, and how it uses information. This classification will change, albeit slowly, over time. Classification is not a rigid standard you adhere to, though it is generally helpful if a standard vocabulary is used. **As an enterprise evolves, classification must adapt.** While a semantic pancake flipping of reclassification will confuse everyone, enshrining the classification scheme as a revered object will make it hated and, over time, less useful.

Low Level Classification

Some IT administrators are familiar with the concept of classification as something they do as part of data migration from one array to another. File systems and log files can similarly document how data has been used, but give no intelligence or capability on how information might be used more broadly as a component of value to the business. The intelligence they provide is like a bar code on an item in a grocery store. It gives the stocker and checkout scanner much needed information that makes store operations much more efficient, but doesn't help the person buying the food. This is low-level, not enterprise-level, classification.

Descriptive Classification

XML and other descriptors help describe information in its individuality or small-scale aggregate. In a similar vein, in Web Sites, classification describes the formats of page zones where information must be frequently updated. The scope of understanding given by XML descriptors and Web-site classification facilitate process. Like the bar codes of low-level classification, the benefit is specific rather than comprehensive. You may learn that a tiger is feline and striped, but if you do not also know that that a tiger may also eat you, if he desires, and you are on safari, the information may be insufficient.

Enterprise Data Classification

Enterprise Data Classification gives a comprehensive view of enterprise information and its uses. It is a multi-dimensional take on description, encompassing how data has been used, and also how it has been shared, and how it can and cannot be re-used. All of this is something an enterprise user will know about the information he or she uses, but it needs to be made explicit in a comprehensive fashion to set information policies properly. And the process of classification has been a part of our education since we were young.

As an example, a retailer may have a litany of several advertising campaigns over the course of a year, each of which will have standard forms of collateral, and most of which will involve a predictable need for storage (holiday campaigns will require more). The information may have a broad relevance for a limited amount of time, and probably will not be subject to a legal discovery search. The enterprise might be interested, over time, in trends of what is effective for Web and brick and mortar channels, but they won't care what is sold.

All of this information, undoubtedly from multiple applications, can be assigned to a class with policies about how and by whom it is used, how it is stored, how long it is kept and whether it is to be archived. These policies can be amended, at the classification level, as corporate policies dictate.

This process turns enterprise data from an unruly sea into a few dozen categories. These categories can be subdivided as needed to fit the shape of the particular enterprise and the particular, granular needs of certain kinds of enterprise data.

Reasons for Enterprise Data Classification

This top-down view of enterprise data has measurable value for the data center and for the enterprise. For the data center, classification can enhance the efficiency of operations. The liability of an organization (e.g., for what has been said or promised by its employees) demands that e-mail, instant messaging, and other forms of information be treated, carefully, as any source of liability deserves. However, a vast majority of enterprise information does not fall into that category, and scrupulous audit trails cost a lot in the latency of instrumentation.

Enterprise-level classification will reveal that that many data is not a source of risk, and the value that might be derived from it at the present time does not merit keeping it available at a nanosecond's notice. **Different classes of data have different lengths of useful life, and some are more confidential than are others.** The characterization afforded by classification puts more intelligence into your data center's ILM strategy. The business value of Enterprise Data Classification is more pervasive and may be hard to measure, but the benefits include the following.

Clarifying The Present

Corporate governance is enhanced greatly by an accurate mapping of the information involved in business processes to corporate structure. This mapping has become more important as enterprises have sought to extract greater process efficiencies by breaking down departmental fiefdoms of process. As data centers have to map the dependencies between applications, so enterprises must map and understand enterprise dependencies on information sources. The multidimensional nature of classification's mapping reveals communities of interest around data that map to usage and not to departments. The knowledge of the extent of these communities enables the creation of better enterprise information strategies and policies.

Understanding The Past

As documentation of the past, enterprise information can be a source of enlightenment. Some trends can only be identified over years. Old approaches that might be otherwise forgotten can be reused, for "fresh" is often a reincarnation of "sufficiently aged."

Enterprise information, over time, can also be a source of risk. This risk lies not just in its documentation of malfeasance, but also in documenting best-fix solutions that, in retrospect, may not have been the best solution. Enterprise data can also be just plain obsolete. There are many reasons not to keep it all forever. Enterprise Data Classification can determine classes of data, not subject to governmental regulation, that can be destroyed.

Predicting The Future

Modeling and analysis to anticipate future demand has become part of business as usual for businesses and organizations of all sizes. **Classification might identify more data sources that are relevant. As a result, enterprise modeling becomes richer, and more cognizant of the enterprise as a whole.** Then it can reveal, better, the ways a project can leverage the capabilities and experience of the enterprise as a whole, something that is otherwise very hard to do. If the enterprise is an aggregate of acquisitions, this enterprise level of understanding becomes even more valuable.

Getting Started with Classification

Taxonomies, in the form of a list of terms common to your industry, or your organizational nomenclature are often a good place to start. Both can be adequate – but both can also be insufficient. Add in your enterprise strategy and road map, and you will have much of the vocabulary needed to begin Enterprise Data Classification. Because it is a top-down process, it can start at a comprehensive level and expand over time. Nevertheless, there are some kinds of enterprise data that deserve particular attention.

Consider *customer information* and how it has changed over the last decade. It has become very important to distinguish between the broad characteristics of customer information that can be used to size and analyze markets (demographics, geographical information, affinities, etc.) and the customer information, such as credit card numbers, that has no business being shared or reused at all. There are many equivalents of customer information

in an enterprise, where there are divisions of usage within a category of information.

E-mail is an obvious example of a classification category. Within that category, depending on the nature of your organization, there is some e-mail - particular customer-facing email of an advisory nature - that may be more sensitive than routine correspondence. Your lawyers know all about these differences.

Take a look at your databases. It is likely that you already have been sharing database information with partners and suppliers. It is often useful to classify information by how widely it is shared, because this affects its scope of relevance, time to obsolescence and sensitivity.

Normalizing the Classification

It is important to realize that classification of any kind is - to some extent - a cultural process. Anyone who has been through a merger knows the frustration of inconsistencies of nomenclature. Enterprise Data Classification, like any process, has its own lifecycle. There is a need for a periodic reconciliation of terms, a normalization process. There is a great value in using consistent terms and deriving consistent policies across classification categories. **The goal is to use the characteristics of classification to use information more effectively in the business and to craft a manageable number of data policies that will cover most, if not all, of your information.** Therefore, classification should not be an uncoordinated, farmed-out process. Inconsistencies in classification will defeat its purpose and effectiveness. Yet, over time, inconsistencies will arise. It is good to plan from the start to address them periodically[1].

Conclusion

Enterprise Data Classification can enhance the purview of corporate governance. It can enhance the functionality of information lifecycle management (ILM). The grooming of descriptors and the queries of search can help implement data policies, but they do not help in writing them. For that, you need the broad understanding that classification provides.

If you are befuddled by the vastness and incomprehensibility of your enterprise's data, relief does not lie in the bottom-up of file systems, or attributes, or search, but in the top-down approach of data classification. Look to Enterprise Data Classification to get the best use out of your enterprise information assets.

[1] The field of Library Science is obsessed, of course, with classification and its evolution. Vehement arguments over taxonomy are pursued at conferences, and the rest of the community just has to live with the results until the next arbitration. This cautionary example would indicate that normalization be rare, brief and brutal.

About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

- ***The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.***

About the Author

Anne MacFarland **is Director of Infrastructure Architectures and Solutions for The Clipper Group.** Ms. MacFarland specializes in strategic business solutions offered by enterprise systems, software, and storage vendors, in trends in enterprise systems and networks, and in explaining these trends and the underlying technologies in simple business terms. She joined The Clipper Group after a long career in library systems, business archives, consulting, research, and freelance writing. Ms. MacFarland earned a Bachelor of Arts degree from Cornell University, where she was a College Scholar, and a Masters of Library Science from Southern Connecticut State University.

- ***Reach Anne MacFarland via e-mail at Anne.MacFarland@clipper.com or at 781-235-0085 Ext. 128. (Please dial "128" when you hear the automated attendant.)***

Regarding Trademarks and Service Marks

Disclosure

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.