



Self-Knowledge Through Metadata with Deepfile

Analyst: Anne MacFarland

Management Summary

Information is power. Information is dangerous. But neither of these aphorisms is pertinent if you do not know the extent of the information. Additionally, electronic information can be easy to lose and hard to find, as it has no physical manifestation to track. As the data stored by enterprises of all sizes grows, the problem of bulk grows as well, particularly for unstructured files outside of databases and structured business process flows that more obviously characterize the nature of the information.

Today, there is little commonality in how people regard the information they touch in the course of the business day. Storage managers see it as occupied media, to be measured and managed by capacity and made available by authorization schemes and segregated by zoning. Business workers see their own organization of files, often ignoring the bulk of other casual files they just happen to have saved. Even content management applications are more concerned with the push and pull than with the nature of the content. The effectiveness of searching, indexing, and sharing schemes - for both fixed and evolving data - depends on a proper definition of the domain on which these processes must run. **You want, after all, the information, all of the information, but only the relevant information. Unfortunately, as long as knowledge of business information is limited to separate strands, developing a comprehensive sense of the knowledge locked in an enterprise - which is necessary to achieve this goal - is difficult.**

A company named Deepfile Corporation, based in Austin, Texas, takes the approach of characterizing the data in place as it is, before trying to do anything with it. It has architected a solution to be politically - as well as functionally - attractive, avoiding the use of agents, and recognizing that the draconian nature of many schemes for file management (often based on fixed capacity limitation rather than need) has hampered their implementation. **Deepfile proposes, not to import data somewhere or process it somehow, but, instead, to analyze it in place, pulling metadata from data's application context, and from the templates used by most enterprises to regularize the organization of often-used forms (like contracts). This metadata is stored in a separate repository.**

This is a top-down approach, which allows more detailed characterization of information with wider relevance, and summary analysis of files not subject to government regulations, with only transitory, local importance. It does not replace storage, content, or compliance management but gives them a more rational environment to work in.

As you think about getting a better handle on the information, that is both the strength of an enterprise and a source of liability, Deepfile's approach is one you should look at. For more details, read on.

IN THIS ISSUE

➤ The Politics of Business File Management	2
➤ Deepfile Modules	2
➤ Using Deepfile	2
➤ Conclusion	2

The Politics of Managing Business Files

There is an inevitable tension of responsibility between the people who produce or gather information and those charged with saving and securing it. Production is inevitably a *more* proposition, while IT system management focuses on limits and quotas. For example, when an enterprise's file obesity is due to a glut of unauthorized MP3 files, most end users see their part in the problem as so small as to be irrelevant. Content management and records management will see the problem as beyond their purview. IT is left with solving the problem, but without knowledge of the character of the files (some MP3s may be legitimate assets), they can only lecture, threaten, and blindly remove. They cannot build a reasonable solution. The problem becomes politicized.

Deepfile's metadata repository exposes the character of retained files in a way that allows business users and IT folks a common view with which to discuss what has been retained and why. It is then possible to set policies – for business governance, regulatory compliance, and IT efficiency – based on the character of the information assets involved.

Deepfile gathers the metadata, sorts files into usage and role-based classes, and provides a dashboard to users to show their file profile characteristics and appetites. With this dashboard, compliance elements can be added to let users monitor their own compliance with policies. The file management overview can also be used as a way to monitor regulatory compliance initiatives (which are not end-user controlled). At the department level, Deepfile lets managers get a better handle on productivity. Moreover, it gives IT a rational way to deal with the proliferation of uncharacterized files.

This is not automatic indexing. Instead, Deepfile focuses on planning what metadata is needed to support use of files as information. This leads to the ability to classify. Let's take a closer look.

Deepfile Products

Deepfile's original two products, *Auditor* and *Enforcer*, are now complemented by a third, called *Sentinel*. *Auditor*, the original product can be bought separately, or with either of the other two products, or with both of them.

Deepfile Auditor

The *Auditor* creates the metadata repository and reports the characteristics of all the information in enterprise systems, organizing it and characterizing the files from the top down by file type, organizational use, and other relevant aspects. It works out of band and uses no agents on servers. An ad-hoc query

interface allows the administrator to drill down to learn more about the nature of a situation.

Deepfile Enforcer

Enforcer is an out-of-band rules engine that leverages the metadata repository to enforce policies on files in groups. It provides the centralized policy enforcement.

Deepfile Sentinel

Sentinel takes the characterization derived from *Auditor*, and adds the rules developed with *Enforcer*. At first release, it reports on file usage. At second release, later this summer, *Sentinel* will link these reports to the policies that are violated. With the third release (by this year's end), *Sentinel* will be a full sense-and-respond system able to generate the action plans to assure compliance with policy.

Use of Deepfile

Deepfile allows IT storage management to be business-process focused. It allows content management to extend its purview to include ad hoc files created as part of doing business. It gives a rational basis for file retention policies, and by year-end, a rational way to comply both with government regulations and with internal needs for tighter, more effective self-management.

CynergisTek, also based in Austin TX, has partnered with Deepfile to offer HIPPA compliance solutions. Sarbanes-Oxley (SOX) solutions will be available soon. Other distribution partners, like SANZ and SiliconIron, see Deepfile's software as a way to manage data storage better.

Deepfile's approach is good for the long run. It sets up a managed environment for all information that lets the management be extended and modified as mandates change. It is not bound by an application. It does not require retained information to be processed. It includes the end user as part of the process, instead of isolating him or her as nothing more than a source of trouble and error. It brings together the relevant parties and makes them all part of the solution.

Conclusion

If your enterprise has a problem of too many files about which you know too little, Deepfile's strategy of in-place-analysis of your mass (or mess) of files gives you the way to characterize the problem, and to determine what is needed in which information domains. This lets people develop a solution that is right for their enterprise. This may be the right starting point for all to move forward.



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 Enterprise Architectures and Infrastructure 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. 28. (Please dial “1-28” when you hear the automated attendant.)***

Regarding Trademarks and Service Marks

The Clipper Group Navigator, The Clipper Group Explorer, The Clipper Group Observer, The Clipper Group Captain's Log, and “*clipper.com*” are trademarks of The Clipper Group, Inc., and the clipper ship drawings, “*Navigating Information Technology Horizons*”, and “*teraproductivity*” are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

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.