



# **Improving Information Access With Collaborative Knowledge Management:**

*Making Semantic Searching a Reality*



## Contents

<b>Synopsis .....</b>	<b>2</b>
<b>Introduction .....</b>	<b>3</b>
<b>Problem Statement .....</b>	<b>4</b>
<b>Previous Options .....</b>	<b>5</b>
<b>The Ardenno Solution .....</b>	<b>7</b>
<b>Supporting Multiple Repositories .....</b>	<b>7</b>
<b>Identifying Data Destinations .....</b>	<b>8</b>
<b>Automated Metadata Assignment.....</b>	<b>8</b>
<b>Ardenno Collaboration .....</b>	<b>11</b>
<b>Ardenno Town Hall® .....</b>	<b>11</b>
<b>Implementation .....</b>	<b>13</b>
<b>The Ardenno ROI Story .....</b>	<b>13</b>

## Synopsis

Knowledge is power. Though said in the 16<sup>th</sup> century by author and philosopher, Sir Francis Bacon, the meaning of the phrase is even more applicable today as companies and organizations around the world are trying to deal with overwhelming amounts of data. The information, documents, spreadsheets, presentations, graphs, image files, photographs and other content trapped in information silos across the enterprise, contains a significant amount of knowledge that can benefit the organization. Unfortunately, much of this information remains untapped, unsearchable and trapped in disjointed systems that often hide data relevancy from an accessible common view.

Today, there are ever increasing legal and commercial pressures to ensure that electronic data is effectively managed and accessible. Companies have struggled to deploy and integrate multiple applications in an attempt to add context to their data repositories. Much hope has been placed on the potential of semantic search tools and collaboration solutions.

Ardenno is one of the first commercially available tools to turn this hope into a reality.

*“Search in the future will look nothing like today’s simple search engine interfaces. If in 10 years we are still using a rectangular box and a list of results, I should be fired”.*

—Susan Dumais, Microsoft Research,  
The New York Times, March 11, 2007

**Microsoft®**

## Introduction

Information overload is confronting each and every one of us. The cheap cost of storage and the development of multiple infrastructure solutions has made it easy for everyone to store everything. Unfortunately, as the volume of data explodes, our ability to find information from it continues to drop dramatically.

Thomas A. Stewart identified “Intellectual Capital” as the company's greatest underutilized competitive weapon. Intellectual Capital, he defined as the untapped, unmapped knowledge of organizations. It is the collective knowledge embodied in a company. Until now, this capital has not been effectively managed. Resolving this issue, is key to the success of an organization.

Organizations are also subject to increasing commercial and legal pressures to ensure that their electronic data and records are effectively managed. For many organizations, the underlying problem is that there are so many types of data, much of it unstructured, existing outside of database systems or stored across the network in disparate repositories, all without contextual relevancy.

The resolution of these business pains also appears to be compounded by the apparent contradictory needs of management, who want control, security repeatability and predictability and those of the employee, who want speed, flexibility and accessibility.

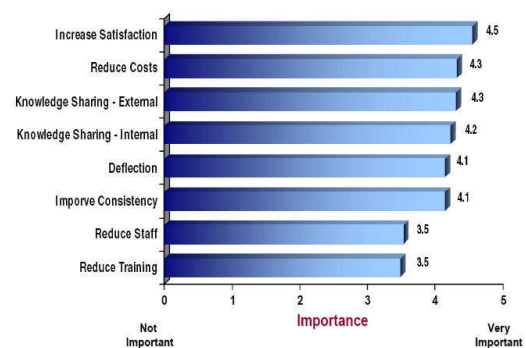
There are many examples of companies that have developed bespoke knowledge management and

collaboration tools in an attempt to capture what has been described as "the water fountain or coffee machine **know-how** exchange".

Most of today's employees have high expectations of search, wanting instant, accurate results. Much of the information and knowledge of the enterprise is stored and trapped in silos and is not truly searchable. Properly defining and categorizing a documents metadata is extremely important to the success of a unified strategy. It helps reduce redundant activities and enhances the opportunity to reuse existing data.

International Data Corp. (IDC) estimated that poorly managed knowledge, costs the Fortune 500 companies alone about \$12 billion a year. Substandard performance, intellectual rework, and the lack of available knowledge management resources were identified by IDC as the key causes of this cost.

Important Benefits of Knowledge Management



Source: Service Excellence Research Group

Figure 1

Additionally, there has been research by independent consulting organizations (Figure1) that point to the business benefits of improving Knowledge Management capabilities.

## Problem Statement

It is important to define, what is meant by Information, Knowledge, Knowledge Management and Collaboration.

A.C. Foskett defined knowledge as, “**What I Know**” and information he classified as “**What We Know**”. Companies must ensure that the improving internalization of knowledge by individuals (tacit knowledge) is enabled through the effective sharing of explicit (recorded or formal) knowledge. Companies must also protect themselves against important intellectual property and tacit knowledge being lost when an individual leaves their organization.





For an organization to be truly a learning organization, not only should its employees acquire knowledge and wisdom, but the organization as a whole should acquire knowledge and develop its intellectual capital. For this to happen, companies must ensure effective collaboration at all levels of the organization. The challenge is that this is simply not just one problem. It is a collection of business pain issues that must be addressed throughout the company.

Companies must ensure that the data produced from both internal and external sources is stored effectively. Often, there is too much and too many types of data. Many of the internal systems deployed to manage this data, take no account of how a worker really wants to work. Organizations continue to build complex “systems” with the hope of a productivity return. Many of these systems however, require excessive administration and support. There

are many examples, where even with the full resources of their IT departments, companies have been unable to fully disclose electronic records, and as a result many legal battles have been lost and million dollar fines incurred.

It is often said that the major concern of a CIO when evaluating a new software solution is, “Will people actually use it.” The answer to this is simple, yes, but only if the system actually makes the users life easier and the solution is easy to use.

Staff members must be able to collaborate at all necessary levels;

-  They will need to be able to access all necessary information, as and when it is needed.
-  Employees must be able to semantically search the entire company infrastructure from a single, easy-to-use interface.
-  Employees should be able to find all data, information and knowledge that they have permission to see, even when they don't really know what it is they are looking for, or where it is stored on the company's network or in what database.
-  Users must also be able to locate information, not just by file names or keywords, but by metadata such as project, contract, author, etc.

The use of metadata is essential if a company wants to manage binary or proprietary file formats including images and pictures.

What is required is a single system, that can backup or archive data as it is produced and automatically attach

metadata tags, as part of an automated process. It must enable effective record and content management of both structured data (stored in a database) and unstructured data (not stored in a database). It must do this while complying with the company's security and access protocols. Finally, the system must work with the company's existing infrastructure such that it can be easily and rapidly deployed and must acquire rapid user acceptance while providing the organization with a proven return on their investment

### Previous Options

Solutions to these issues are as numerous and as varied as there are companies, IT groups and departments. Multiple applications have been deployed to deal with specific business pains, but until now, there has not been a holistic approach to the overall problem. IT departments have tried to resolve the situation by using database applications, data warehouses or enterprise content management systems.

These approaches have resulted in overly complicated solutions with low user acceptance. Companies generally want to ensure that primarily key documents are retained in their document management system. Users have also found data can only be stored in a database as a "blob". Consequently, they prefer to store their data in folders on the company's network, where at least they can have easy access to it.

To add context to this approach, companies have looked to develop naming conventions for folder structures. They have looked to develop taxonomies

and ontologies to add unified context to the files stored within these structure. This approach, however, only works if everyone within the company always adheres to the defined procedures, and everyone follows the agreed procedures as to where to locate the data and new files produced.

Document Management and Enterprise Content Management Systems also tend to rely on this folder metaphor. As a consequence, when files are added to such systems or "e-Rooms for the first few months, they are constantly relocated to various folders, as users try to reconcile their final location. Often, this creates a maze of files and iterations leading to the business pain of version control management.

A heterogeneous approach exploiting various aspects of these solutions is very common, but many companies still incur unnecessary costs. Not because they have not retained their electronic data, but because they continue to struggle to extract information from it. The heterogeneous approach of multiple different repositories and data systems forces users to practically know where data is located before they look for it and to know how to navigate those storage systems.

To address these challenges, companies have developed intranets and looked to enable the use of web search tools as a means of extracting information and knowledge from their electronic records. This approach can entail a high maintenance overhead and may require the assignment of full time administrators to ensure that these intranets are kept up to date. In recent years, these web search tools have been enhanced so that they now can be used against the data folder structure of a company's network.

Current search engines however, have been found to be limited in terms of their practicality for business users. Many of these search engines are optimized for browsing websites, and not unstructured documents and folder structures of the network. Many of these tools do not support the searching of metadata. As a result, users struggle to find relevant information.

Often, when a user begins a search, they do not know what they're looking for. They will know it, when they find it. To conduct such a search, users must be able to search semantically, which requires metadata tag support. Current tools are limited by their poor security, inability to offer a centralized index and keeping that index up to date with the latest changes on the network. The inadequacies of current search tools have led many to begin working on the next generation of search engines. The press often refers to this as Web 2.0.

The needs of companies constantly challenge the limits of information technology, but even after the deployment of multimillion dollar systems and infrastructures, developing and sustaining that success continues to demand new enhancements.

Companies must enable their employees and teams to work much more effectively together. They must reduce the cost of poor quality and reworks and they must ensure that the know-how and tacit knowledge of their key employees is shared amongst others. Companies and organizations are now actively developing knowledge management projects that they often describe as their internal university.

At the core of these projects, companies are looking to develop tools to enable their employees to collaborate much more effectively. The lack of commercially available and affordable collaboration tools has been a major challenge. Companies have had to develop their own collaboration solutions, often based on some commercial technology.

The e-mail system of many companies has been "press ganged" into the role of the collaboration environment. Users and project teams use e-mail and attachments as their primary method of collaboration. As a result, the number of file versions on a company network and the amount of e-mail traffic within an organization is spiraling out of control. This approach demands that someone, usually the originator of the e-mail, has to consolidate all of the edits and comments.

Each of the approaches described have their pro's and cons. The challenge is how to build upon the pro's and leverage the existing IT infrastructure and applications.

## The Ardenno Solution

Ardenno looked at the fundamental business pains that hinder effective knowledge management and team collaboration. Our approach was to address the root causes of the problems, unlike previous offerings that focused superficially on treating specific “pains” with complex and expensive software systems.

Market research quickly found that although these issues and problems were greater in certain departments, such as research and development, they were common across disciplines and industries. This meant, that the solution had to be departmental and industry independent.

The ideal solution had to work with a company's existing infrastructure, data repositories, and file formats. It was essential that the system could be installed and deployed within days. The goal was to develop a system that would exploit the best features of the company's existing infrastructure and technology investments. It had to enable users to be able to find information and increase the knowledge and intellectual capital of the organization. Ardenno also had to be able to provide easy to use collaboration to everyone within the company.

With the cost of storage continuing to fall and data volumes doubling every 6 to 18 months, the challenge confronting businesses is how to extract relevant information and content from these data repositories. It is essential that the system support semantic searching and metadata tagging. For metadata tagging to work, Ardenno had to enable the automatic association of metadata and files.

Research has found that of the data produced within a company, less than 20% is structured data, i.e. stored in a database system or enterprise content management solution. This 20% however, was identified as being some of the company's most important information. So Ardenno would have to support indexing, searching and collaboration of database records, in addition to the unstructured data found on the company's network and intranets.

Ardenno delivers on three main criteria:

- 1: Ardenno delivers a proven ROI
- 2: Ardenno enjoys a high level of user acceptance
- 3: Ardenno reduces user workload and enables them to collaborate smarter and to be much more productive and effective

## Supporting Multiple Repositories

Ardenno's immediate strength begins with the ability to work with ANY datasource and ANY type of data. Datasources is the term used by Ardenno to describe individual data repositories. Normally these are unique folders on the company's network, as well as local folders including the numerous “MyData” repositories. A datasource can also be a database or ECM entity (SharePoint, Livelink, etc.).

Ardenno automates the polling of these datasources for any new or changed data and files. When new files are located, they are indexed and automatically copied or moved to designated destination(s). Alternatively, they can be left in their present location. A company's Standard Operating Procedure (SOP) may determine that when a file is produced, it must be relocated to a backup folder and/or to a file storage repository. It may

even require that a file, when moved, is zipped in order to maximize the file storage space. Ardenno provides all these capabilities as standard. The extensive Ardenno security layer enables a company to configure exactly what data any given user has access to. Ardenno will also integrate with a company's existing LDAP or Active Directory security profiles.

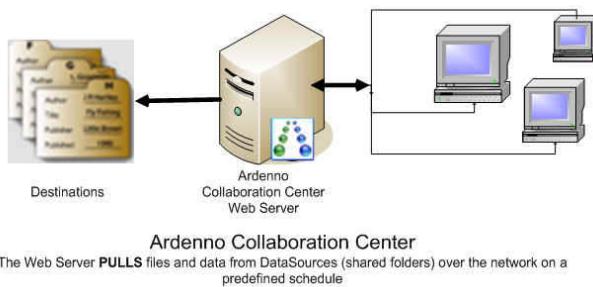


Figure 2

Ardenno will automate the "Pulling" of data from the various datasources

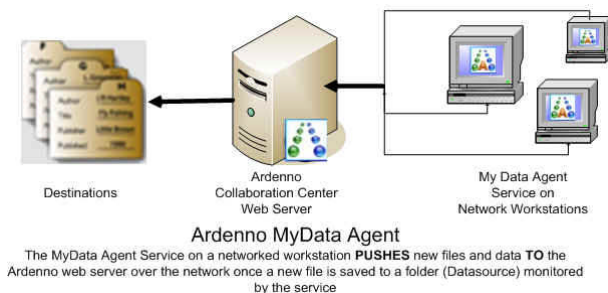


Figure 3

Alternatively, the "Ardenno MyData Agent" can be installed on local workstations and laptops and be configured to automate the "Pushing" of data from the various datasources to the Ardenno application. This module has proven very popular with users who work remotely and intermittently connect to the company's main network server.

## Identifying Data Destinations

Data Destinations in Ardenno describe information about where data and files are physically stored, once processed by the Ardenno automation engine. Data Destinations are usually folders and subdirectories on the company's secure files storage system. Unlike previous solutions, Ardenno does not attempt to become the key file storage system. Ardenno fully leverages a company's existing storage infrastructure. Ardenno uniquely stores pointers and semantic metadata tags for these processed files. Ardenno then monitors the infrastructure to ensure that these pointers and the index is current and valid and will advise the administrator as appropriate. This approach enables Ardenno to be implemented and deployed within days.

The Ardenno data destination approach also enables the automatic creation of additional native file copies in different file storage locations and systems. This seamless process provides a method of creating original and collaboration copies of a file, while ensuring secure file processing. Users no longer need to know where data and files are actually stored.

## Automated Metadata Assignment

Metadata is information about information: more precisely, it's structured information or knowledge about a data file or record. This metadata can be based on a controlled vocabulary, an ontology or taxonomy. Metadata provides enhanced context for records and files. Users have more scope to locate information with the best possible recall. The promise of metadata is to enable semantic searching.

Ontologies classify information into logical multiple tiers and categories. Ontologies supercharge semantic

search and retrieval capabilities. Unlike search technologies alone, ontologies reveal the overall structure of a knowledgebase in a hierarchy that adds tremendous relevancy and visibility to the user community. The user navigates through sub-categories to narrow the search, a process that helps avoid false hits that are outside the area of interest. When used with Ardenno's semantic search, ontologies increase efficiency by limiting the volume of data returned in a search to the most relevant results.



**Edit Tag Group**

Allows to maintain a Group of Tags in order to facilitate consistency for tagging DataSources

Name:

Description:

Tags:  (Click here to add a new tag)

!	Tag	Type Name	Default Value	Description
<input type="checkbox"/>	Contributor	<input type="text" value="X"/>	<input type="text"/>	An entity responsible for making
<input type="checkbox"/>	Coverage	<input type="text" value="X"/>	<input type="text"/>	The extent or scope of the cont
<input type="checkbox"/>	Creator	<input type="text" value="X"/>	<input type="text"/>	An entity primarily responsible f
<input type="checkbox"/>	Date	<input type="text" value="X"/>	<input type="text"/>	A date associated with an event
<input type="checkbox"/>	Description	<input type="text" value="X"/>	<input type="text"/>	An account of the content of the
<input type="checkbox"/>	Format	<input type="text" value="X"/>	<input type="text"/>	The physical or digital manifest
<input type="checkbox"/>	Identifier	<input type="text" value="X"/>	<input type="text"/>	An unambiguous reference to th
<input type="checkbox"/>	Language	<input type="text" value="X"/>	<input type="text"/>	A language of the intellectual cc
<input type="checkbox"/>	Publisher	<input type="text" value="X"/>	<input type="text"/>	An entity responsible for making
<input type="checkbox"/>	Relation	<input type="text" value="X"/>	<input type="text"/>	A reference to a related resourc
<input type="checkbox"/>	Rights	<input type="text" value="X"/>	<input type="text"/>	Information about rights held in
<input type="checkbox"/>	Source	<input type="text" value="X"/>	<input type="text"/>	A reference to a resource from v
<input type="checkbox"/>	Subject	<input type="text" value="X"/>	<input type="text"/>	The topic of the content of the r
<input type="checkbox"/>	Title	<input type="text" value="X"/>	<input type="text"/>	A name given to the resource.

Figure 4

Until Ardenno, there has not been an easy way to automatically add the necessary metadata tags to data, records and files. Previous approaches have required the "end user" to apply tags manually. This has been the main reason why simple tagging systems have failed to take root within the corporate

environment. In addition to automated tagging, Ardenno allows users to add their own unique metadata tags and notes "on the fly" to increase the potential of semantic searching. Ardenno lets employees find information the way they want it, rather than the way structured applications manage it. Ardenno supersedes the folder metaphor. It is the key to empowering your semantic searching and knowledge growth.

Organizations that have captured and stored information in enterprise content management systems now want more advanced semantic searching and relevancy identification techniques to leverage their existing investments. To achieve this, companies with ECM systems such as Microsoft SharePoint, OpenText Livelink ECM and databases such as Microsoft SQL Server and Oracle, have deployed the Ardenno Collaboration Center.

## Searching

Ardenno is not simply a search engine, although we provide seamless integration to the world's leading open source enterprise search engine Lucene. Lucene has excellent tools and configuration flexibility to meet the needs of the largest enterprises. Ardenno took the Lucene enterprise engine and enhanced the relevancy algorithms to exploit the Ardenno metadata and indexing capabilities. This enabled advanced semantic searching, which was needed to support the specific requirements of the business community.

Users can perform the most sophisticated searches in the context of their own work and projects. You and your users can also use the metadata tags to enable more creative semantic searching. When performing searches with Ardenno, a user can simply begin, by typing into a

tag field and Ardenno will display the possible matching tag labels and values using its auto complete feature. It's important to note that no one needs to know or learn a query language such as SQL!

Helping users to find information through the semantic searching capabilities, far from being the endpoint was a regarded by Ardenno is only the beginning. The Ardenno solution enables users to extract information from their search results and to develop knowledge. The results of a search in Ardenno are sorted by content relevancy or can be viewed directly in the Ardenno town hall. Data can also be viewed in its native application. Ardenno does not change the native file, nor does Ardenno attempt to convert any files into a proprietary format. This approach enables Ardenno to search for any file type, without requiring end user customization.

### Enhanced User and Team Collaboration

Ardenno enables users and project teams to create Knowledge Bags. These are collections of important files that they and the project team want to collaborate on. Entries in a knowledge bag are not copies of the original files but rather pointers to where those files are located on the company's infrastructure. Users can now have the comfort of a "MyData" type entity, but without having to constantly store the files locally on their workstation.

Ardenno takes "MyData" to the next level. Knowledge Bags are much more than "MyData" replacements. They are project and team data sets, which everyone in the team can access for rapid collaboration. The user can select any file in a

knowledge bag and review it and all related relevant data relationships visually in Ardenno's Town Hall.

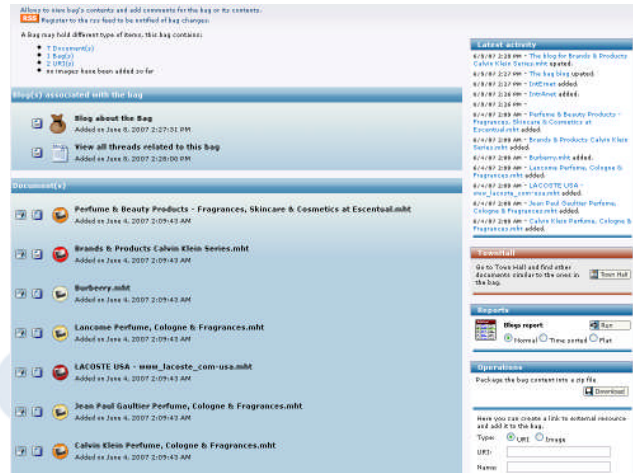


Figure 5

Ardenno recognized that companies would want to control who had access to their valuable information. So, in addition to supporting LDAP or Active Directory, Ardenno also allows customers to configure what user groups, departments or specific individuals will have access to the information contained within any given knowledge bag.



Figure 6

## Ardenno Collaboration

Many people today “collaborate” via email. To get colleague’s opinions and feedback, files are sent as e-mail attachments and updated as feedback is received. Across the company, inboxes are inundated with everyone’s comments. Ardenno provides a much more efficient alternative, a “Collaboration Blog” that stops email from being your collaboration platform.

Figure 7 shows an example of a multi-threaded contextual conversation that occurs between a team of collaborators. Access to a specific blog can be limited to members of a bag’s security profile or by individual user access settings. In addition, using RSS feeds (figure 5) users of a knowledge bag are automatically alerted whenever one of their knowledge bags or blogs is updated by another member of their collaboration group.



Figure 7

## Ardenno Town Hall®

One of the keys to the success of the Ardenno solution has been Town Hall. Town Hall provides the ability to display users, files and database records with content relationships in a multi-dimensional clustered display. Its social networking (community framework) offers knowledge workers a way to provide “social structure” between information, individuals, knowledge and their organization. Town Hall social networking provides available information and greater logical relevancy through advanced display technologies.

Town Hall displays a selected file at the center of the graphic (the center of town...). Around it are

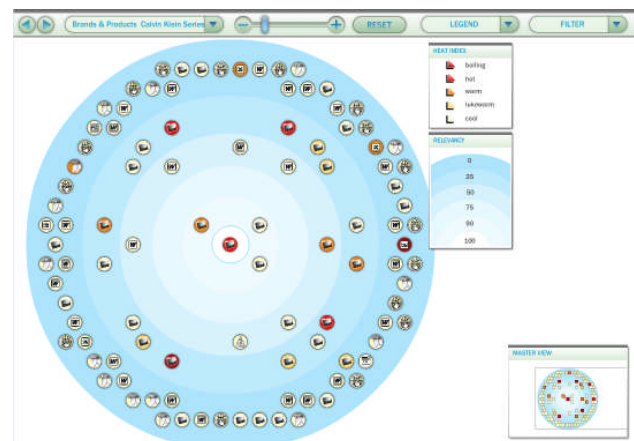


Figure 8

displayed relevant files, records and knowledge bags, related to the selected Town Hall file that this user has permission to see. The closer files, records and knowledge bags are displayed on the graphic to the central file, the more relevant their content.

Town Hall provides a clustered display of multiple relevant files in a single view and lets users pan or zoom the display for enhanced readability and fully leverages your mouse track ball pointing and zooming.

The icons of the various files are displayed in different colors based on their “*Knowledge Heat Index*”.

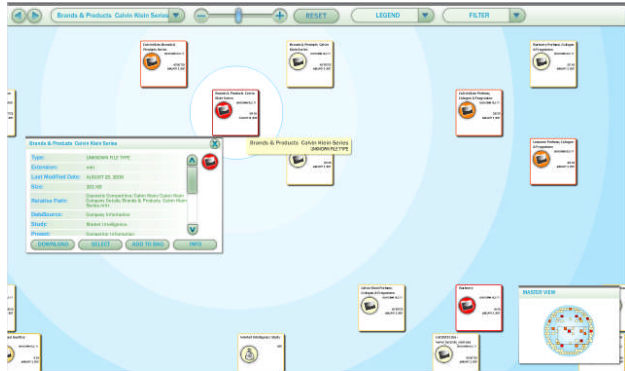


Figure 9

Figure 9 shows details of a selected file, including the associated metadata. The Knowledge Heat Index is constantly changing based on the number of times that other users utilize this file, and the actions that they perform on it. The Heat Index rapidly becomes a very useful tool; as it enables users to quickly gain a visual indication of a file’s “usefulness” by others within the organization.

## Ardenno and Enterprise Content Management (ECM)

Ardenno also leverages and enhances your existing enterprise content management systems such as Microsoft SharePoint, Open Text Livelink ECM, Oracle based document management systems and data systems that support JDBC.

Ardenno provides a single searchable portal to ALL your enterprise data files and records, whether they are structured or unstructured and whether the content is stored and saved in the document management system or not. Ardenno applies to all

parts of the organization; it provides a singular view of your entire data. Figure 10 is an example of how Ardenno displays files contained in Microsoft SharePoint. In the background, one can see files newly added to the SharePoint repository. The foreground screen shows how Ardenno is able to “point” to the SharePoint database of documents.

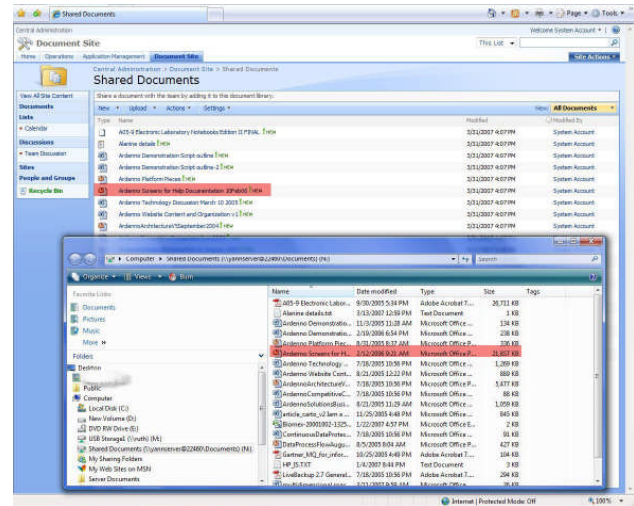


Figure 10

This enables Ardenno to process these data files / records like any other datasource. Ardenno can also query and produced Townhall displays of content from this and other data storage systems such as Microsoft SharePoint, Open Text Livelink ECM, Oracle or any other storage system that has a JDBC driver. All this is done, while complying with the company security requirements.

## Implementation

The success of an implementation is measured by how quickly new applications are adopted throughout the organization. Ardenno's product training is oriented towards passing the knowledge you need to be self-sufficient rapidly. The Get Ardenno Started (G.A.S.) configuration and training session takes you from zero to 100 in a few days. You learn how to configure and support Ardenno administratively. We'll teach you how to share the power of Ardenno and how to maximize the value of Ardenno semantic metadata tagging with your existing infrastructure. Ardenno guarantee that by the end of the first day you will be able to use the collaboration center on your existing infrastructure.

## The Ardenno ROI Story

The Ardenno ROI story is not complicated and focuses on quantifying the productivity value to organizations. A detailed Excel based ROI calculator is available for you to determine your own return on investment from an Ardenno deployment. Simply contact your account manager to get the ROI calculator. It will highlight to you the various areas where Ardenno will positively affect your organization and bottom line.

To help you monitor this return over time, Ardenno provides sophisticated metrics (Figure 10) with respect to the data and files you produce.

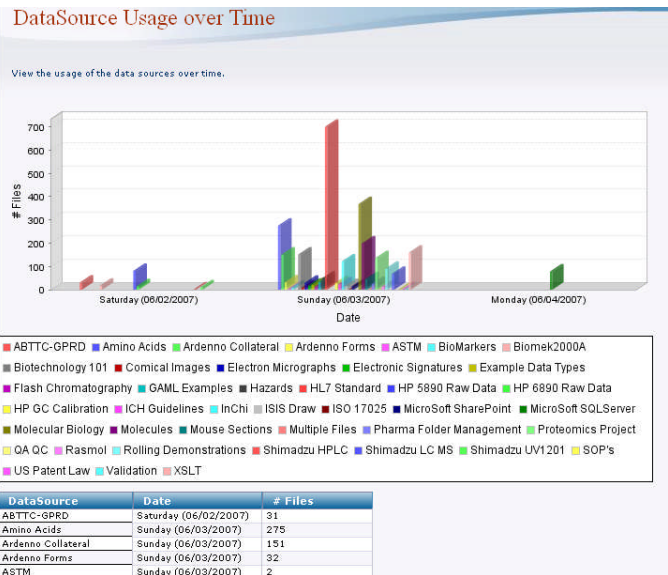


Figure 10

## Ardenno has a positive ROI in less than 90 days!

The ROI chart below shows an example of quantitative proforma measurements and the categories tracked using the Ardenno ROI calculator.

### Ardenno ROI Business Benefits

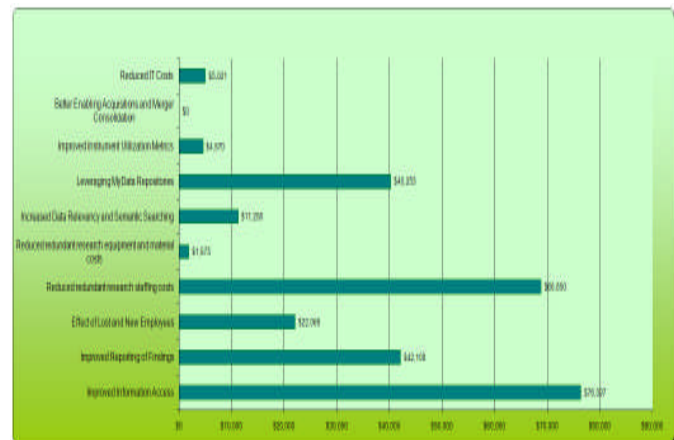


Figure 11

## Summary

Ardenno helps organizations develop their intellectual capital by delivering the right information at the right time with context and relevancy. We help companies reap dividends from increasing staff efficiency and knowledge.

Ardenno delivers a simple approach, to make your data more accessible. It enables you to extract information from it when you find it. Thus, you are able to better leverage the information you've invested so much to create.

The Ardenno approach is the only known off-the-shelf solution that is uniquely qualified to help you better manage your structured and unstructured data.

The Ardenno Town Hall provides a multi-dimensional visual display of files and records with a focus on file relevancy and a file's / record's Heat Index. Town Hall offers unparalleled business intelligence to aid users in their efforts to better locate and utilize the data that exists within your organization.

Get started now with your own Ardenno implementation. See for yourself how Ardenno can enhance your existing technology environment.

## Improving Information Access with Collaborative Knowledge Management: *Making Semantic Searching a Reality*

June 2007

© 2007 by Ardenno Solutions, Inc.  
All rights reserved.  
Published 2007.

### Restricted Rights

Printed in the United States of America.

The information contained in this document is proprietary and confidential to Ardenno Solutions, Inc.

No part of this document may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording, for any purpose without the express written permission of Ardenno Solutions, Inc.

This document is subject to change without notice, and Ardenno does not warrant that the material contained in this document is error-free. If you find any problems with this document, please report them to Ardenno in writing. Ardenno, Ardenno Solutions, the Ardenno logo, Ardenno Collaboration Center, Ardenno Town Hall are trademarks of Ardenno Solutions, Inc. All other company and product names may be trademarks of their respective owners. The information contained herein is subject to change without notice.

Ardenno Solutions, Inc. All rights reserved.

This document contains or may contain statements of future direction concerning possible functionality for Ardenno's software products and technology. All functionality and software products will be available for license and shipment from Ardenno only if and when generally commercially available. Ardenno disclaims any express or implied commitment to deliver functionality or software unless or until actual shipment of the functionality or software occurs. The statements of possible future direction are for information purposes only and Ardenno makes no express or implied commitments or representations concerning the timing and content of any future functionality or releases.