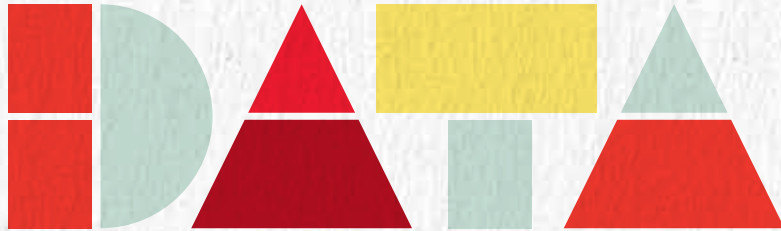


Leveraging Underused



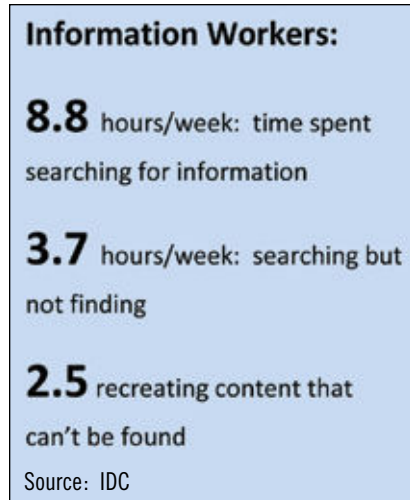
By Kevin J. Harrang

Every one of us faces at least one common problem on a daily basis: finding the documents and information we need to do our jobs. This is important because all of us are increasingly required to do more with less — in less time and with less cost.

The good news is that we're awash in documents and information. The bad news, as we've all noticed, is that the vast increase in available information actually seems to have made it more difficult to find any specific thing we're looking for.



FIGURE 1



Finding something we know exists is hard enough, but it's the tip of the iceberg compared to *not* finding high-value information because we simply don't know it exists. Thus, we're losing a colossal opportunity. Indeed, one of the most underutilized assets in legal organizations is information we already have but lack a practical way to find. And law workers are not alone in this situation (see Figure 1).

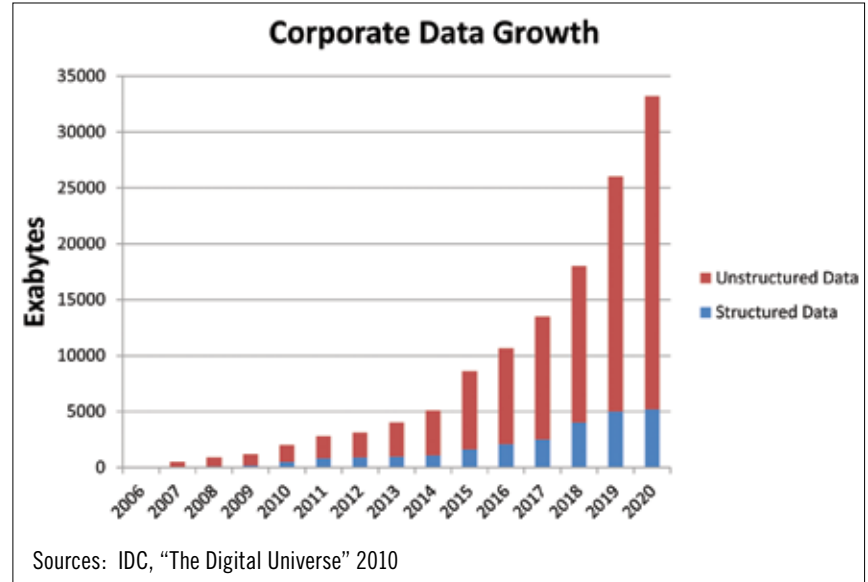
Compare the challenge we face in locating relevant information to what we do when we find things on the web. The internet is the largest collection of information ever assembled, used by millions daily to find everything from facts to products to people, but it has no centralized organization, no digital equivalent of the Dewey Decimal System. And yet it works: Far more often than not, we find what we seek, even when we're not sure exactly what we're seeking.



Kevin J. Harrang is the former deputy general counsel for department operations of Microsoft's legal and corporate affairs group (LCA). He has been a frequent CLE speaker, was the keynote speaker at two LegalTech conferences, lectures at the University of Washington School of Law, and is involved in a number of legal, charitable and civic organizations. He is also founder and board member of MetaJure, Inc., a legal technology provider. kharrang@metajure.com

The author would like to acknowledge the invaluable contributions of the following colleagues in the preparation of this article: Marty Smith, MetaJure, Inc.; Marc Miller, interim dean, James Rogers College of Law, University of Arizona; and Steven Levy, founder, Lexician.com.

FIGURE 2



So why is it so hard to find documents on our own computers, and even harder to find documents and other information within our organizations? Why is it you have to know where something is kept in order to find it in your organization, but not on the internet? What would it mean if you could easily access not just some of your information, but all of your information?

How we got here

If you ask your IT experts or record specialists about this problem, they'll probably point to something called *unstructured data* as one of the prime culprits. Unstructured data means information that has not been placed into well-organized repositories. The vast majority of growth in corporate data has taken the form of unstructured data (see Figure 2). The solution to unstructured data,

we're told, is to structure it — to move your documents and information into costly new systems that can bring organization to this mess.

There are only two problems with this diagnosis: It doesn't work, and it's no longer true (if indeed it ever was).

Legal organizations need to make the same leap with their information and documents that all of us made with everyday information over the past few years with internet searching. What follows is a roadmap for such an approach.

Why finding information is so difficult

When attorneys and other professionals look for information in the form of documents, they are most often attempting to do one of three things:

1. Find a specific document they know exists (e.g., the confidentiality agreement we signed last year with that company);
2. Find examples of a specific type of document (e.g., other examples of confidentiality agreements); or
3. Find everything that exists about a certain topic, company or individual (e.g., everything about confidentiality agreements, or everything we've ever done with that company).

Sometimes, the people looking for information recall exactly what they are looking for, which we can call **searching**, but often they need the assistance of tables of contents and other lists to trigger their memories, called **browsing**. Browsing leverages the familiar, *“I can’t remember it but I’ll recognize it when I see it”* phenomenon.

For centuries, with paper documents, attorneys have relied upon methodical storage that allows later retrieval and use. If you were to toss all your paper documents into a big pile, you’d find it almost impossible to find anything later.

For years, the heart of the law firm or legal department’s information system was its file room, in larger firms led by a trained librarian and staffed by file clerks to do all the sorting, filing and retrieving. File rooms work well for finding specific documents and information that is known to the searcher, such as a specific will or pleading — Item 1 in our list of legal search goals. So, if you know exactly what you want, you can quickly locate that document in well-managed organizations, albeit with a good deal of behind-the-scenes staff assistance (this can work less well after hours).

If you didn’t know exactly what you were looking for (Items 2 or 3), traditional file rooms were not very useful because of their lack of subject and content cross-references. If you were looking for similar types of documents or general information on a topic, the file room was typically not set up for this purpose. Instead, many lawyers would wander down the hall and ask their colleagues if they knew of any similar prior work that would make a good starting point.

As electronic documents and information became more common, it soon became apparent that, unlike file rooms full of paper documents, digital collections could be very useful, not merely for finding something specific, but also for finding examples and doing research more generally.

Unfortunately, the value of these possibilities has largely been thwarted in countless legal organizations by the way they have implemented their systems. Sticking with the familiar way of doing things, attorneys created documents locally on their personal computers, with the expectation that they would file important documents into document management and other centralized systems when they were done.

One result of this approach, intentional or not, was to shift much of the filing duties that had formerly been done by lower-cost file clerks onto the attorneys or other legal professionals. Not surprisingly, compliance with the expectation that everyone would upload and then accurately file their documents into a centralized system has been more miss than hit in most organizations. The first step toward a better digital world, we were told, required lawyers to file more rather than fewer documents, often profiling or tagging documents with additional information.

The problem, of course, is that this took the time of the most valuable and highly compensated individuals in the organization, and saddled them

For years, the heart of the law firm or legal department’s information system was its file room, in larger firms led by a trained librarian and staffed by file clerks to do all the sorting, filing and retrieving.

with duties formerly done by clerical staff, together with additional chores like document profiling. It shouldn’t surprise anyone that this has never worked well.

Even in the rare situation where document management procedures were aggressively enforced to capture work product such as signed contracts, such systems were never designed to house all the different types of documents and information generated by or for the organization.

Like the file rooms of old, many electronic filing systems contain only significant documents such as executed agreements, whereas drafts, contracts that were never signed for some reason, and a lot of other work product never makes it into the centralized system.

Worse yet, word processing documents are only the tip of the data iceberg. A huge amount of work product now resides in email, which is almost never shared, much less archived, in centralized systems.

In addition, most corporate legal departments have more than one, if not many, special purpose systems for patents, trademarks, litigation, corporate filings and myriad other specialties. All these systems contain a wealth of useful data, none of which will ever make it into the document repository.

Another ubiquitous example is electronic billing systems, which contain a wealth of information about prior work product, but are almost never tapped to

GLOSSARY

Structured data: data in the form of documents and other electronic files that have classification and cataloging data in fixed fields such as databases

Unstructured data: data in the form of files such as the free-form text of a word processing document or email

Document management: systems for managing collections of documents, especially word processing documents. One of the principal functions of document management systems is to convert unstructured word processing documents into more structured forms by prompting users to put information about the document into fixed fields of a document profile (e.g., name, type, parties, keywords, etc.).

Metadata: data that describes other data

answer significant questions, such as: *Have we ever paid for this research before?*

Yet as problematic as corporate legal department documents may be, consider the myriad documents produced at great expense by outside counsel, which typically exists somewhere but is seldom archived in a repository easily accessible to the department that paid for the work product.

The tip of the iceberg

The results of this natural evolution of our information systems are twofold:

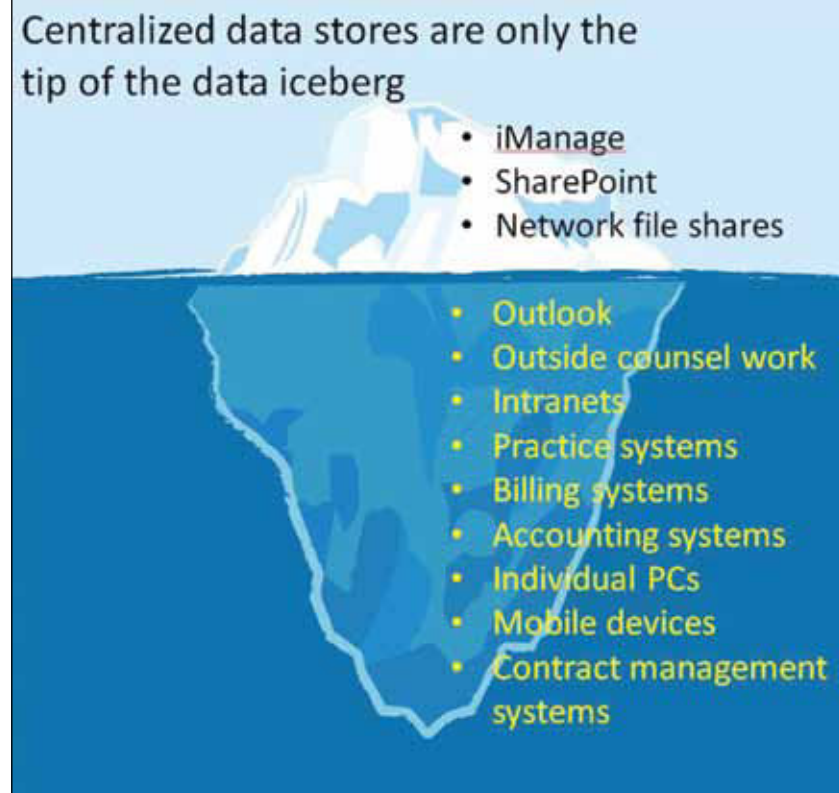
1. There has been an explosion of unstructured data, as described; and
2. The majority of important documents and information within a given legal organization reside outside its centralized repositories — document management, SharePoint and the like.

Where are all of these documents? Legal departments that attempted to find out discovered that a significant percentage of documents uniquely reside on the computers of attorneys and paralegals — despite available centralized storage systems such as SharePoint, as well as internal policies that mandate uploading of important documents into them. Because few, if any, attorneys grant access to their computers to their colleagues, this means that the vast majority of documents are not available to anyone except the authors and recipients.

And recall that these documents are just the tip of the data iceberg. Taken together, this all means that the professionals in corporate legal departments do not have access to most of their documents and information, much less a usable way to search them.

Old thinking

The response of records management and IT teams to this evolution of organizational information has largely been backward-looking. By the old way of thinking, the rise of unstructured data



means that users must not be doing what they're supposed to.

With many organizations having purchased specialized (and expensive) document management systems — or at least repositories such as SharePoint — pointing out that much of the legal organization's data resides outside such systems only elicits the response that this is a policy compliance problem, at least from the persons whose job it is to maintain such systems.

So the offered solution to the veritable tsunami of unstructured data is to demand more time from the legal professionals to convert it into structured information and upload it into centralized systems. Not only is this unwise and inefficient, it's not clear that this approach actually solves the core problem of information retrieval even when compliance is high.

It's no longer even necessary.

Addressing the problem requires violating long-held assumptions.

Taking a different approach will likely make your records and IT experts uncomfortable, at least at first.

New thinking

A new approach can be summarized in five principles.

Principle 1: Embrace messy, unstructured data

The first step is to stop viewing unstructured data as a liability, and instead, embrace it as a key organizational asset. If you do this, there is a surprising and counter-intuitive positive result: The bigger the mess, the better.

If you can access all of your data and easily search it to find what you want (more on that in a moment), then it turns out that more data is better.

It may be hard to remember, but one of the early questions users had of the public internet was: If everyone could publish whatever they wanted on the

web, how on earth would anyone find anything? For the internet, companies such as Google pioneered solutions that enabled users to find the websites they were looking for.

Document search within the organization requires different software, but the result can be the same: It's possible to find documents despite the fact that they have *not* been catalogued and filed, and despite the fact that *you don't know where they are kept*. The keys are to realize that there is great value in all the information that will never make it into your structured repositories, and to recognize that it does not need to be structured to be searched.

Principle 2: Centralize access, not storage

Rather than centralizing the storage of your information, the key lies in centralizing access to your data. Recall that internet search works because you

don't need to know where in the world the website or data is actually stored.

The problem with document search within the typical organization is that you need to know to look in the billing system for billing records, the document system for documents, etc. And as we saw previously, most documents and data are not even accessible because they are stored locally on users' computers.

On the other hand, centralizing access through search can leave all your information exactly where it is, thus obviating the painful conversion exercises to move legacy data. Centralized access also solves the problem experienced by anyone who has attempted centralized storage of information, namely that such repositories quickly become out of date without universal compliance. In order to be up-to-the-minute, information needs to be accessed where it resides, not where it

was supposed to have been moved according to policy (but wasn't).

Principle 3: Search instead of browse

Electronic systems can index every single word in every single document, a task that would have essentially been impossible before, and it turns out that searching the index takes a fraction of a second.¹ (Note that the way many document management systems ask users to profile their documents when saving them to add information is a largely manual and incomplete form of indexing.)

It is now possible for systems to simply index all the contents of your documents and rely on advanced searching techniques to locate what you want, all without additional profiling by the user.

If you centralize the index of your documents, then you can use the magic of the latest search technologies



ACC Webcasts

**ACC Webcasts offer CLE/CPD credits.
They're right here.**

You need to get up to speed on the latest legal developments and fulfill your CLE/CPD requirements. ACC Webcasts are your answer. Convenient and cost-effective, live webcasts are offered weekly and dozens are available on demand. Expand your legal knowledge and earn CLE/CPD credits. Everything you need is right here.

webcasts.acc.com

ACC Association of
Corporate Counsel

not only to locate documents by keywords, but also increasingly by concepts. For example, search systems can now be trained to locate something called *Confidentiality Agreement* even though the search terms you used were *Non-Disclosure Agreement*.

Principle 4: Share by default

As with the internet, the value of networked information systems grows dramatically with the number of contributors. It can be a roadblock, then, if the members of an organization resist information sharing.

While it could be viewed as information hoarding that attorneys keep so many documents locally on their computers, it is more probably a result of the disincentives of centralized repositories. Busy professionals may see little upside to spending time profiling and uploading their documents into centralized repositories.

Where you have behavior that will benefit the group more than the particular individual, the solution is to flip the presumption. Rather than a system in which everything is private unless affirmatively shared with the group, the system should facilitate sharing of everything except that which is affirmatively withheld. This way, the incentives align in a way that advantages the group.

Rather than making a lot of work for your colleagues to opt into your document systems, make opting out of this knowledge base the exception rather than the rule. Note that this makes the most sense among trusted groups of professionals, such as a law firm or legal department, where the advantages of information sharing are most immediate.

Principle 5: Be realistic about security

Another reason often cited for lack of legal work product sharing is the requirement for confidentiality. Certainly, this is a real issue, and probably one of the main reasons that legal

departments often resist participating in enterprise-wide search systems.

Within the relatively small groups of trusted legal professionals that make up many law firms and legal departments, however, there is much less reality to this concern.

Information systems are often improperly judged relative to perfection rather than the real world. Systems are judged insecure if even a single colleague could conceivably access a document they don't need to see, despite the fact that this possibility exists with every unlocked office or file room. Worse, attorneys forget or are unaware of how many relatively low-level IT personnel could gain access to their email and electronic documents if they really wanted to.

The point is not to disregard the real issues surrounding security, but to do a realistic cost-benefit analysis. It is worth remembering not just what harm a document could do in the wrong hands, but also what *good* the right document could do in the hands of the right trusted colleague at the right time.

It's important to stress that these principles do not spell the end of your structured systems such as SharePoint or iManage. Rather, it means that you recognize the value in the majority of your organization's information that resides outside such repositories, and which will never make it into them.

If all you can do with your current systems is to find the signed agreement you're searching for (which in itself would be a triumph for many organizations), imagine what it would be like if the same search could also locate things like a recent unfiled amendment to that agreement done by a colleague? Or correspondence and memoranda written by outside counsel, and the billing records for that work?

None of this is science fiction anymore. But achieving these results requires embracing a different approach to information than we've taken in the past.

ACC EXTRAS ON... Data management

ACC Docket

The Five Pillars of In-house Ediscovery (Dec. 2012). www.acc.com/legalresources/resource.cfm?show=1321063

The SMB's Guide to Developing an Electronic Data Management Program that Can Reduce Legal Risk and Lower Ediscovery Costs (May 2012). www.acc.com/docket/smb-ediscov_may12

Presentation

Outside Counsel Management: Leveraging Data and Technology for Successful Value-Based Relationships (Oct. 2010). www.acc.com/oc-mgt-d&t_oct10

ACC HAS MORE MATERIAL ON THIS SUBJECT ON OUR WEBSITE. VISIT WWW.ACC.COM, WHERE YOU CAN BROWSE OUR RESOURCES BY PRACTICE AREA OR SEARCH BY KEYWORD.

Fortunately, this approach will be familiar now that all of us are regular users of the internet. Think of this discussion the next time you find what you are looking for among the millions of webpages. Or better yet, when you find something surprising that you weren't directly searching for. We take it for granted that this is how the public internet works.

This is the real world.

This could be your organization. **ACC**

NOTES

- 1 If you've ever wondered how searching can be so fast, here's a quick layman's explanation: When you search a large number of documents for keywords, the system is not actually looking at each document while you wait. That would be too slow, even at computer speeds. Rather, the system has first pre-searched all the documents and catalogued all the words into an index. Just like a book, it's easier and faster to look something up using the index than it would be to read the entire book each time you wanted to find something.