Interview

Key trends and developments of rights information management systems – An interview with Jim Cuff of Iron Mountain Digital

Jim Cuff

is vice president of Strategy for Iron Mountain Digital. Jim's role is to ensure that an appropriate business growth strategy is in place, and that the company has a compelling technology and solution vision to deliver on this strategy. Jim joined Iron Mountain in 1996. Before joining Iron Mountain, Jim has worked in IT in the financial services industry with responsibility for networks, open systems and desktop computing. Jim is a graduate of Boston College with a degree in Finance and Computer Science.

ABSTRACT How dose an enterprise strengthen their compliance posture? Michael Moon interviews Jim Cuff of Iron Mountain, on trends and developments of rights information management systems.

Journal of Digital Asset Management (2009) 5, 98–110. doi:10.1057/dam.2009.1

Keywords: XML; regulatory compliance; XBRL; digital storage; information management; digital preservation

MM: We're here with Jim Cuff of Iron Mountain Digital. Jim, if you would, give us a little bit of a background in terms of your career.

JC: I am currently the vice president of Strategy for Iron Mountain's digital division. I've been with Iron Mountain for 11 years, working in various roles in information technology, and growing our core physical business.

I've been focusing on our digital division since its emergence around 2001. Before that, I worked on the technology side for the financial services industry.

Throughout my technology career, I focused on the deployment of technologies – whether that meant running and managing teams to deploy worldwide networks, or building data centers and ensuring that the infrastructure was there.

I've moved over to the solution-development area at Iron Mountain Digital, to develop digital versions of our products and solution sets.

MM: Great. How about giving us a quick overview of Iron Mountain?

JC: Iron Mountain provides secure, offsite storage and information management for our customers. We help folks with storage

management problems related to cost, scale and capability. We also help customers manage risk and compliance related to physical and electronic information.

We've been doing this since 1951 when the company was started. In the mid- to late-1990s, we went public. We're currently a worldwide organization with somewhere around 20 000 employees that provide services around the globe.

In 2001, in response to what our customers needed, we added digital services – essentially towards solving the same issues that we do for our customers on the physical side. Whereas in the early 1990s we would only pick up customers' tapes and bring them to vaults, now – instead of physically picking them up in a truck – we can transport those same backup capabilities, when it makes sense, over a network to a secure offsite facility. We can do the same with electronic records that need to be archived.

MM: Could you bring us up to speed in terms of some of your core business issues – regulatory and/or operational?

JC: They fall into several different categories. We help our customers on a daily basis to efficiently handle their storage management issues, whether

Correspondence: Jim Cuff
Iron Mountain Digital,
120 Turnpike Road,
Southborough, MA 01772,
USA
E-mail: Jim.cuff@
ironmountain.com

they are physical or digital – in today's world this is more important than ever. We also ensure that organizations have a great option for operational recovery through our data protection offering. In terms of digital preservation, there's a lot of action on the corporate side around the Federal Rules of Civil Procedure – specifically Rule 26. We help there with both the information flow related to Rule 26, and the work-ahead time to figure what you need to keep and what it may be appropriate to destroy in the normal course of business.

MM: Does that correlate with what people now are calling 'electronic discovery (e-discovery)'?

JC: Oh, absolutely.

On the e-discovery side, and from an Iron Mountain perspective, we have always helped our customers with discovery. Some of it may be physical records. Some of it is e-discovery. The physical discovery piece has existed for a long time, so it may be better understood.

electronic-discovery is all about the ability to provide responsive information that's stored in electronic format. A perfect example would be e-mails that need to be retrieved and produced. The problem set is just much larger than it may have been historically on the physical side, where you may have years and years of backup tapes, or an extremely large archive that you need to sift through to find the information you need.

MM: And that is part of what Iron Mountain Digital provides as e-discovery services?

JC: Yes. We provide discovery services that run the gamut from helping companies come up with a strategy around being prepared for a litigation event to do discovery to services for the actual *process* of discovery – whether it be collecting that information from disparate sources, or putting it together so that you can look at it and figure out what you need to have. We've also got high-performance solutions that allow you to look through that information quickly, so that you can save a lot of time and money in actually *going through* those discoverable documents.

The latest advancement in that area has been with our acquisition of *Stratify* in the fall of last year. That gives us a very robust capability that leverages some very innovative technology, which dramatically improves the efficiency of the discovery review process. Every second literally counts here – as a very expensive legal

staff drives this phase. Increasing their efficiency is a huge value to our customers.

MM: So back to the basic domain of digital preservation – you mentioned Rule 26 and other sorts of requirements to have documentation ready for discovery procedures. Those are generally legal claims to inspect records – usually in pursuit of a lawsuit. Is that correct?

JC: Yes. The most common need for preservation is ensuring that documents are kept, or preserved, to meet a regulatory or legal obligation. This needs to be done in a consistent, auditable way where the authenticity of the information cannot be called into question, within a framework of consistent policies and procedures.

MM: What else is going on in the digital preservation space?

JC: It's a very big space. When you extend the time horizon to what is called long-term preservation, there are lots of interesting issues. If you look at archiving as a relatively new phenomenon in terms of electronic archiving, we've hit the first wave of conversion requirements.

I'd say the first version of most archives follows that same technology trend of 'IT Plan A.' That involves bringing in a system/application stack. Most of the systems were brought in on similar platforms to the ones that were used to run operational systems. The problem is, archiving is fundamentally different from your typical systems – and following standard IT practices for system deployment can be problematic. Assume you deploy your standard systems, then you get to that three- to five-year time window after which IT systems would typically be up for 'refresh.' You realize that you're sitting on an extremely large information store – in classic technology measures. Converting that stuff to a new platform with standard methods simply doesn't work on that scale.

When I say convert there are really a few things to start thinking about. One is the physical conversion of the 'bits' from an aging storage infrastructure. This is the easy one to think about, but doing it at an extremely large scale is not something to figure out later. At Iron Mountain we think about bit preservation on the Exabyte scale – and the short answer is

that bit preservation is a never-ending task, not one that is run as a project every few years.

Once you've got a handle on the bit fidelity side of things, the next challenge is in the area of format. A simple example is that Lotus 123 spreadsheet you might have from the early 1990s. Even if you've preserved the bits, can you actually do anything with them. There are several options in this space but it's far from solved. You can also break this problem down to two areas. The first is reading or displaying the information, and the second is being able to actually process the information. So if we stay on the example of the Lotus 123 spreadsheet – do you want to be able to view or print the spreadsheet, or would you like to actually type in some new values and have the formulas work. This is a very hard problem today.

Another interesting slant on this is in the area of data custodianship. Imagine you have left the firm, and the 123 document is still there. We all know that the company is the 'owner' if it is a corporate document, but who is allowed to access it?

In the digital preservation space, you need to think of it as a continuum where some things are still not ready for primetime, and if you have an urgent need, plan on spending a lot of time and brainpower. There are areas where Iron Mountain works to solve these problems knowing that they are better solved once at scale, as this is really a core competency of very few organizations.

MM: That kind of suggests – 'IT Plan A.' What is 'IT Plan B,' now that we have mountains of 'dark,' or unknown content.

JC: I think the next wave on that is well understood on the media side of things. In terms of the company who would, for example have half a petabyte worth of information sitting around, it's a non-trivial problem with 'IT Plan A' to physically move that half-petabyte to a new set of disks. Now there are better options for the physical migration of information.

The next wave beyond that gets into the actual content of what's on the disk, not just preserving it so that you can read the bits, but also making sure that you are able to understand the metadata or encoding. Generally speaking, this is having knowledge about the context of

the information, not just its content. A good example is, knowing which specific application (including version) created the information.

You want to make sure at the point when you are ready to perform format conversions, or take some other action to ensure that the information is useful in the long term, that the metadata can tell you what the source format is. You need to be able to read it or that you will know what type of conversion process you need to go through to read it.

There's a capability in this next wave around indexing and retrieval that's very different from what you would've expected 10 years ago. You could call that almost a Google phenomenon. The post-Google expectation is to have the ability to perform a search for everything in a way that's *meaningful for the user* – as opposed to living within a structured search environment.

MM: Could you elaborate on what we might refer to as, 'Content profiling?'

JC: This is a little bit of art and a little bit of science, I think, in terms of how folks deal with things today.

The classic approach of a very structured metadata model has a tremendous amount of appeal in terms of being able to search for a piece of data and retrieve it based upon a predefined taxonomy or known retrieval method. That's one way of looking at things. The other is this notion of being able to flexibly search for and find something based on an unstructured content view.

I think that the winning combination is to be able to do both. The key is to be able to use that structure or semi-structure as a way to get into the right ballpark, so to speak, and it may even let you drill down to specifics.

But there are other situations where you may be looking for something that doesn't fit that model. Having the flexibility to handle that type of situation is, I think, extremely important.

A good example would be that you may have a taxonomy that drills down in one search method. But if you're looking for something that's not in that taxonomy ... say, for example, you're looking for a purple tree. No one may have conceived a color scheme for trees when they built the taxonomy. You may need to resort to alternate methods.

At the end of the day, the reason people archive data is so that they can find it when they need it, and fast.

MM: When you're using unstructured data storage where you don't have a predefined way of searching the contents – I imagine you'd have a Google appliance that crawls through that.

What other sorts of search technologies does Iron Mountain have in place?

JC: There are different solutions for different problems. Something that works on a strictly text-based search capability works well only in certain situations.

There is also not a single-size-fits-all solution. For example, a Google appliance or something similar is a great fit for getting search results with the goal of retrieving a few documents, but wouldn't necessarily be a good fit if you are looking for high volume retrieval. Iron Mountain has solutions that are optimized not only for finding things, but also high volume data manipulation and movement. A good example of that is being able to deliver responsive information sets for a discovery event. It's also important to be able to figure out how long information should be kept – ensuring that retention and destruction are done in a consistent credible manner are other areas where Iron Mountain moves way beyond the search appliance space.

There are also other specialty needs. A great example is with our *e-discovery tool*. We have some specialized technology that can actually use advanced algorithms to look at documents and put them into like groupings – and which enables you to even make suggestions as to what those groupings may be about.

MM: That would be akin to the technology often referred to as 'text mining'?

JC: Yes. Very similar to that. There's a bunch of technologies in there. A great example is an advanced document clustering technique in our Stratify Legal Discovery System where you end up with something that is what we call a 'Concept folder.' There, you can say, 'Here's a bunch of things on a similar topic. Here's a good place to look.'

You may also want to drill down into that topic. So not only do you have a top-level concept, but you then may also have a nested set of concepts underneath that, to help guide you

to something you're looking for. In the area of e-discovery, that's incredibly powerful.

There are other situations or capabilities where using that structured metadata in new ways makes a lot of sense. For example, if you're in a medical imaging application, you may be looking for all studies of a certain kind that came from a particular modality over a particular data range. That may fit very well into a structured metadata approach, where the concept folder idea wouldn't work quite as well.

MM: In the second wave of digital preservation, you touched upon content profiling. And that there are a number of ways of approaching it – both structured and unstructured. You make the case that a combination of both is probably the optimum way – depending on the point of view or the heuristic outcome that you're seeking to achieve.

Anything else, in terms of some of the things that your clients or users in this second wave grapple with?

JC: I think the biggest thing that I'm still surprised about, is the fact that the scaling issues sneak up on so many folks. It's one thing to start a digital archive, or start down the path of long-term preservation, but the problems related to scale and cost-effectiveness just grow larger and larger every day that you're in a preservation mode.

A lot of times, those solutions – some of which go back to 'IT Plan A' – just don't work in a cost-effective manner when you get to an extremely large data set. Doing some simple capacity planning can help here. If you do the math on a major data manipulation effort, you'll find out very quickly that it may take way longer than you think to complete the effort. You may need to have more than one approach in the long term to make sure you achieve your goals and don't overspend.

MM: Right.

JC: Having those things work so that you're actually, again, able to find what you need in a timely fashion and get it back are extremely important.

MM: Would this suggest then that the third wave in this overall digital preservation mode would then start to deal with the notion of federated storage?

JC: Yes. There are a few different points in there, Michael, which can become big problems in and of themselves.

One is the notion of, 'Is it required that everything be collected and put into one spot?'

In terms of it being in one logical spot, the answer to that is a big yes. In terms of it being collected and put into one physical spot – when you get to actual scale, that becomes an impossibility.

The ability to manage items in place – to migrate those items to new locations for cost or performance reasons – becomes incredibly important. But at no point do you want to be in an all-or-nothing environment.

For example, Iron Mountain runs an extremely large archiving business. We think we do a great job at it. But the operating notion is not, 'Unless it's at Iron Mountain, it can't be managed.' The notion is, 'We can participate in a management ecosystem where the customers can find everything they need, whether it's with us or whether it's in a more active store in their own environment.'

If you look at some of the industry trends around cloud computing – whether it's cloud computing or Iron Mountain Digital's storage-as-a-service model, or some of the new capabilities for technologies that run on a customer premise – there are lots of great options for folks. It's a matter of finding the right mix that works for them.

One thing that needs to be a part of that mix is the notion that whatever you've selected and wherever your content may be, you're able to find what you need when you need it, and you're able to keep it under appropriate control. **MM:** I remember several years ago, we were working with a client that had a DAM system, and unlike a lot of the traditional DAM systems, this particular company put the asset into what they called an object-oriented database.

The real point of enlightenment for me was that each object had a unique IT address. What if every file ever created had a unique IT address? That would fundamentally change DAM from this kind of little private walled garden into – essentially – what we now call Google.

JC: There are some interesting concepts there.

One is from just a basic search-and-retrieval standpoint. That notion there, Michael, gets right to the heart of the issue that some of the classic metaphors for finding and retrieving information just don't work on a large scale. Imagine searching through a giant file system that looks

like a shared drive across a billion files. That's not an optimal use pattern for anyone.

If you knew the document ID for the piece of data you wanted, using your unique IP address example, that's a matter of, 'I know what I want. Here's my ticket. Come and give it to me.' That's a great model. A lot of systems – including some of those that Iron Mountain runs – have something in place that works very similarly.

If you went to one of Iron Mountain's hardcopy facilities, you'll see that every single one of those boxes has a unique barcode on it. You can equate that to what we were doing on the digital side. Every single piece of content that you give Iron Mountain to manage for you will have a unique content ID. Whether that gets exposed to the end application or not depends on whether it's meaningful in that context.

Another interesting aspect to that is that you may have a piece of content that you collect from five or six, or, who knows – 10 or 20 different sources. Each one of those sources believes that they have a unique piece of content, when, in fact, all 20 copies are the same. That's when you can get into a very interesting situation. Especially, when it comes to ownership and rights management and who's allowed to do what.

When you've got a single piece of content that 20 systems think they manage, there are some very interesting things you have to do there, and that you need to consider. Especially, as you look at managing that from a preservation standpoint.

Some interesting things come into place there like *deletion-or-destruction*. Imagine a piece of content that 20 systems are using. One of them says, 'I'm done. Please delete.' Nineteen are still using it as active content. You have to make sure that you're able to accommodate those issues.

This is especially important when you put into play preservation orders or records management considerations. Even when an application may think it's done with a piece of content, an organization may not be done with it for policy reasons. One example is if you actually have an item that is declared an official record, and it reaches the end of its life cycle. When an approved destruction occurs, you want to make sure all copies of that record are

destroyed. This is nearly impossible if you don't know where they actually are.

MM: At Iron Mountain Digital, you traditionally have been in the business of managing business records. While still managing objects, your principal added value entails managing the metadata and the profiles and the rights information about the objects under management. Is that correct?

JC: The way we say it at Iron Mountain Digital is that job number one is managing information consistently. The form with which it's presented is more of the 'how's,' in terms of that, as opposed to the 'what's.'

For example, if you have a memo, you may have typed that memo up on your computer. You may have printed it out. You may have e-mailed it out. It may have then been backed up by a server.

Regardless of all the places that that memo has been put, that memo needs to be managed in a consistent way. It may be casual correspondence. It can be removed from a system after its business use is over. It might be a memo to terminate service for something. In that case, it may be similar to a contract and have a different storage time requirement.

So for Iron Mountain, the notion is, 'Let's talk about managing your information appropriately first.' We'll have whatever capabilities are needed behind that – whether that capability is a 1.2 cubic foot box, or whether that capability is a 100-year electronic archive.

MM: Right. As we've been talking about digital preservation, you had also mentioned two other large concepts that I'd like to now explore. The first one entails compliance. Can you give us an update in terms of compliance-related trends and development as it relates to archiving and information management?

JC: I'd say that the biggest change there is the level of expectation whether it be an industry, corporate governance, legal issues ... I'd say that the bar has definitely been raised as far as what's required to be done. Then, for example, Rule 26 or the Federal Rules for Civil procedure – are more about codifying issues that had already been best practices in a lot of organizations.

The difference now is that it's an expectation that raises the bar above aspiring to use these best practices to being absolutely required to do so.

That's changed behavior in a lot of different organizations. At the end of the day, each organization must ensure that it is meeting its obligations relative to information management.

MM: What other sorts of regulatory framework issues are coming to the fore, specifically as they might relate to Sarbanes-Oxley (SOx)?

JC: Some of it may be new, in general, but is old hat for us. I'd say when it comes to record-keeping and what you're doing with your corporate information, there's a key message that we give our clients for advice and that we hear echoed back very strongly. That is, that *consistency is key*.

An example is that you can design the best problem for SOx compliance on paper, but if you don't have business practices that are followed – and can then be audited to back it up, you're in really bad shape. You would then be in non-compliance with your own policies and practices.

What we've seen is that it has been critically important is to come up with a set of policies and programs that help a company achieve that compliance, in a consistent manner. So it's really important to make sure you know what you're capable of doing, and then make sure you can inspect it to make sure you're following your practices.

I'd say that previously it may have been enough to have a good set of policies. Now it's really something that you have to be able to *inspect and stand behind*. There is almost nothing worse in this area than having documented policies that you don't follow.

MM: This reminds me. Occasionally, we do projects with really large Fortune 10 and Fortune 50 companies. Inevitably, conducting these projects entails that we have to go through the whole procurement and vendor-registration process. In the course of that, we run into what I like to call a 'revenue-suppression system.'

Heretofore known as Enterprise resource planning.

The unbelievably rigorous requirements!

Could you get a little more into how specifically you substantiate policies in your day-to-day operations?

JC: The method there is to be able to ensure consistency. Another point there is that when you look at the way that most companies work, a lot of business gets done not within the four

walls of a company in terms of badged employees.

The notion of outsourcing work to partners makes that even more complex and difficult. The key is to be able to automate. If you have issues that require lots of manual intervention or require people to change the way they work based on a situation, it's very difficult to get consistency.

It's important to make sure you've got pragmatic procedures to outsource, and that you can manage them on an *exception* basis rather than having everything you do be an exception. It's easier said than done.

You also need to make sure that you've got a pragmatic rollout plan, and to make sure that when you look at what you're working on, you're able to also assess your risk in certain areas, and pick a program that's right-sized for what you're doing.

In your example – if a firm's very concerned about specific regulations relative to labor, they may have a screening program in place. The key is to make sure it's the right size for what they're trying to do. And that where appropriate, they invest in the right program.

In that example, investing in a screening program to make things easier – if you want to work with small vendors – may make sense. In other situations, it doesn't make sense at all. We have a consulting group that spends a lot of time making sure that you assess your risks in terms of the number of employees, the places that you do business, and even whether you're in a highly litigated or highly regulated industry or not.

You want to make sure that you right-size those investments to what's required. Some may be a very high investment. The question is which is worse – to spend more money than you need to, or to spend less.

MM: Sure.

One of the things that we track here at the Journal has been the emergence of *policy servers*. We did an extensive interview with Adobe on the Adobe life cycle and other kinds of rights information management systems.

From a digital preservation, compliance-management, risk-management perspective – could you give us your take on key trends and developments of rights information management systems, and perhaps policy servers – which are

more about the day-to-day operational enforcement of rights information?

JC: Yes. I think there are different ways to look at things, but when you look at it from a *preservation* standpoint, the first challenge is to make sure that you're actually capturing all the information in a manner so it can be preserved properly. In some cases, that's done with human intervention. In most cases, if you're able to do it in an automated fashion, you're going to have a much higher degree of success in terms of capturing the right information and ensuring that it can be managed appropriately.

For preservation, you start with capture. Once you've gone through that, then you get into what we'll call the *store* or the *storage architecture* piece. There, it's a matter of, 'Okay. Now that I have this, from a policy standpoint, what do I want to do with it?' In terms of, 'Is this something that needs to be ready for high-speed usage? Is this something that needs to be in multiple locations? Is this something that needs any special encryption?'

Then the last piece we look at is from a *usage-model standpoint*, in terms of, 'Does this need to be made available for particular applications or for follow-on uses? Do I have the right access models in place, in terms of being able to retrieve this information?'

One of the biggest unsolved problems from a long-term standpoint is that – in many cases, the systems that you collect information from are very much from a *user-rights centric model*. For example, my name is Jim Cuff. I have a mailbox. Everything in my mailbox is technically owned by the corporation. But from a user-access standpoint, it's assumed that *I'm the one who's accessing it*.

Now if that information is needed five years from now and for some reason I'm not with the company, the very primary model of user access is, 'I'm Jim. I'm within a certain division. I'm within a certain function of the company.' In many organizations, the directory structure to keep track of those things is far from clean, and certainly not well maintained over time.

You can then move to 10 years from now and ask who should be able to look at the information that was put into an archive by the vice president of strategy for the digital division.

I think on the rights side of things, there's the user-centric view of the world. That really

doesn't hold up in the long term – then you get back into situations where you may need to derive rights from metadata and other contexts around the information. That is a pretty tricky thing to do.

MM: That's great.

Let me shift again to another aspect that we kind of touched on. That was document formats, with an eye towards long-term preservation.

We've read recently about how one of the large standards committees has now come down on a document-storage format. I believe they call it, 'Open Document.' It's based on the Microsoft technology.

JC: Yes.

MM: Are you familiar with that development?

JC: Yes. I think it's very much a great proof point for where things are today, where there are competing views around the right approaches for things.

Imagine there was a time when I was using a word processor called 'WordStar.' There were special commands and keys you could use to make it do certain things.

If in 2008 I want to open up a WordStar document, I'm probably going to have a problem.

MM: I imagine there are little boutique shops that can handle those types of issues?

JC: There absolutely are. It's a matter of, 'What's the value of what's in that WordStar document, relative to the cost of accessing that information?'

MM: Would you have any kind of rule of thumb for the fully burdened cost to take a WordStar document and bring it forward into a readable format?

JC: Michael, that's where it gets to the different approaches. There are certain approaches around, 'Do we all convert to a standard format?'

In the example of a WordStar document, converting to a standard format, you've got to answer some specific questions in doing that.

One is, 'When I convert, do I need to keep the original?' In many situations, the answer is, 'Yes.' If you need to keep the original, you don't need to just look at the conversion cost, but you need to make sure to also look at the storage cost.

Assuming that it's even the same footprint as the old version – which sometimes may not be

the case – you need to consider whether you want to convert every document you have and keep both the original binary version and a 'standard view' version.

There are costs related to that. And that's really an opportunity to choose whether to handle it all at the outset or convert when needed.

The other piece to look at is to understand that *even standards change over time*.

MM: I understand there are lots of different ways of using Iron Mountain services, would you say there's a good/better/best?

JC: Michael, there is. I think the last thing that anybody's going to do to run their business is to always go for the platinum solution, if what you need is the gold or the bronze. There are certain documents that may be of high risk or of high importance, where you definitely go for the best.

MM: Such as invoices, purchase orders, letter of termination, files concerning regulatory agencies?

JC: Exactly. There are other times where the company should make an investment decision and say, 'I'll either take a lower-fidelity version, or I'll do an on-the-fly conversion because I've got a high degree of certainty that the likelihood I'll need that is low.' So it'll cost me less money in the long run to do a conversion on the fly than to do it upfront.

You have to be very careful with those kinds of situations. Again, *scale* comes into play. Just like there's a scale issue if you're keeping two versions of everything, there's a scale issue if you're converting a subset of your archive on the fly. Then all of a sudden, 'on the fly' is not so efficient.

It's really a matter of getting back to, again, what we were talking of on the policy management side of things. The better idea you can have at the outset of what you're storing, the easier it is to make those decisions on a policy basis, rather than on an *ad hoc* basis.

JC: The one other strategy that's out there is the notion of some of the more advanced source systems that are on the create side. Some today already create, for example, high-fidelity and low-fidelity versions of something, especially on the rich media side. In those cases, there may be a preservation strategy that's run or managed even outside of the core preservation system.

They have a better idea around when certain formats are going to be obsolete. The key there is to make sure that the math is done in terms of, 'How long will it take to do the conversion of these documents'?

In those examples, there may be formats you want to convert to that don't exist yet.

MM: Many of the Journal readers have very large data stores.

At what point should a DAMster start thinking about extending the envelope of their preservation system to include an Iron Mountain Digital capability?

JC: There are a couple of terms. Michael, I love the term, 'DAMster.' You introduced me to a new term today, which I appreciate.

The right times to do that is – and again, this gets back to if you move information or data because you have to or because you want to – when you're looking at making some type of change, and you've got capability issues – whether that be on the processing side because you need more capacity – or because you need to do things from a preservation standpoint that aren't in your current capabilities.

Generally, Iron Mountain helps organizations when they need to extend their capabilities in some way. It may be as simple as needing more storage capability and not having the appetite for the cost and capital outlay of doing it themselves. Often it's when a company needs to invest in a compliance program, or when they are getting tired of the grind of e-discovery on their own.

We always recommend to folks that if you have stuff that's running, the last thing you want to do is to interrupt a well-running system. There are times it makes sense in terms of activity level of documents and things like that at the right times. There are also ways to do this in a creative manner.

If you look at stuff when it gets to be 5 or so years old, or when retrieval patterns get extremely low – that stuff is still, in many cases, sitting on the same infrastructure that you need for high-performance use. In those cases, we recommend to folks that rather than buying more high-performance disks because your infrastructure is clogged up with older information – to siphon that stuff off, so that you can continue to leverage the investment you've already made for your active information.

MM: A lot of these larger DAM systems incorporate a hierarchical storage management system.

JC: Yes, and those are great systems, because they have built in the notion of the high-performance area versus the low-performance of the archive area. In those cases then, you want to look at when it's time for media refreshing. What's the right answer to that?

MM: Right. What is the right answer to that?

JC: I'll tell you that the right answer to that is *when you look at media refresh, the biggest mistake that's hard to recover from is starting too late*. That is, if you've got a specifically sized infrastructure.

MM: So I've got a bunch of CD ROMs from the mid-1990s.

At what point do CD ROMs as an optical media begin to run into trouble? As well as laser disks?

JC: The answer is this. When they run into trouble is different for each media type. That's well known.

But imagine, for example, if you have a bunch of CD ROMs and you assume that a year from now, you're getting into the danger zone for media life. If you've got 365 CD ROMs and you can do one a day, if you want to be done in a year, that's great. If you start, obviously, a month before you need to be finished, you're going to need to have someone help you.

Doing a lot in a short period of time is very difficult to do yourself. It's just hard to have the infrastructure to complete it in a timely fashion.

If you look at a service provider – whether it be Iron Mountain or some of the specialty houses that do these kinds of things – they'll invest in that infrastructure, so that you're not paying a huge premium to do things in a shorter period of time.

Of course there are also the laws of physics. If you need to do something in a really short period of time, you may need to look at other options.

For example, we have special rooms we'll use to store media that's aging – for example, in 0° with 0 per cent humidity. Then we can buy some time while we do the conversion processes. There's always an answer that says there's an IT and a capital component of it. But there also may be some creative solutions to buy yourself some more time while you work those things out.

MM: Shifting the subject a little bit, I recall working with some companies in the mid-1990s that were doing electronic journaling of transaction data for large IBM mainframes. The idea was that if you were a bank, you'd have a big IBM mainframe running MVS. In the course of running that day's checks, if there were a catastrophic failure of the system, all of the data that was in the memory of that mainframe would disappear.

The idea is that – in the actual cache memory for these mainframes – they were taking snapshots of the data as it flowed through the mainframe memory, and then were journaling it electronically off to another tape that was offsite. The idea of electronic journaling has been around now for 15 or 20 years.

Can you bring us up to date in terms of what the Web 2.0 version is of this kind of journaling – more for digital content, media assets and that sort of thing?

JC: Yes. I think one of the biggest stories there is the blurring of the backup recovery world and the archive world – relative to content. Especially when you get to fixed content ...

MM: It seems to me, Jim, if you do it right, you get two for the price of one.

JC: You do. The trick is to make sure that you get two for the price of one as opposed to two for the price of two.

MM: How do we do that?

JC: The key is to look at the solutions that are in place. For example, if you look at an archive system in the e-mail management space.

If you're using an e-mail management system that does shortcutting, stubbing, or actually removing content from your mail server and putting it into an archive, the actual mail service store becomes smaller.

In that case, you end up baking up less on your e-mail server, because you've already put that content in a safe place in your electronic archive.

MM: Would you characterize that as a best practice?

JC: It's absolutely a best practice, I'd say. In that case, it's not necessarily the case of two for the price of one. It may be two for the price of one point five. But you're gaining efficiency, and you're starting to manage content in one place. That's probably a good example also for the corporate or IT world – where putting data in

an archive allows your operational systems to run better.

In the rich media world, a lot of the applications already work with a content-centric storage system. In those cases, having a content-centric storage system that puts data into geographically disparate places allows you to have access to the content for operational use. It also can solve – at least for the content side of it – the data protection side.

The key there is to make sure that if you've got issues with versioning or other types of things, you actually do have the recovery points you need. The other key piece there is that a lot of terms – at least the current state of the art is – you can get those solutions for the content. But most systems also have a database or other type of controlling information that you need to make sure is also appropriately protected.

Another issue to be sensitive to is that if you ever decide to not back up data, you must be absolutely sure that the storage system can meet your recovery needs. The one that gets missed most often is recovery from corruption or inadvertent deletion. If your system simply replicates your corrupt data or deletion command to its remote partner, you better make sure you are still running backups. Basically, if you aren't sure that you can recover from these types of scenarios, you better be writing backup tapes and sending them off site.

MM: I recall having the conversation with Jeff van Dyck from Boeing. He has the overall charge or accountability of managing all of the marketing communications for Boeing. He deals with thousands of PowerPoint decks, and tens of thousands of images and documents and PDFs and so on.

He drew a pyramid, and said that the top of the pyramid was his little walled garden of DAM. It's all brightly lit, fully tagged, catalogued and indexed.

As soon as we take it out of our walled garden and put it into one of 7000 servers, spread through Boeing's IT space – at that point, that piece of content that was brightly lit and tagged goes dark. All we have now is just a file name.

He said he began employing XMP – the XML metadata platform standard from Adobe – and started tagging these files that lived off in his server. Then he developed a spider to crawl

through those items – basically, to create an active inventory of all of the files that were out there in server land, and connect that or link that back to the master in the walled garden.

But he said as soon as you took it off the server and pulled it onto a laptop, it was basically dark. There was very little technology to maintain a connection. He said that's the frontier and we'll get to it some day.

Can you bring us up to speed in terms of how Iron Mountain Digital thinks about that pyramid of the walled garden, server land and then the C-drive jungle?

JC: Yes. Let's start with the jungle. It's an interesting place out there. I've done a lot of work in the IT space, and I've talked to a lot of folks that are on the management side of that frontier jungle area.

These folks have worked so hard for so many years. We've finally got management practices and controls for viruses and spyware and things like that.

MM: As well as applications – that's the whole premise, right?

JC: Yes. First you do the virus, and then you do the inventory management. Then you do the software distribution. Then you're feeling pretty good about all the hard work and investment you put in.

There are two pieces that you're missing. One is the notion of the information on those machines that may live only on those machines. You've got to be able to protect that from a backup and recovery standpoint. The other point is that there's information on those machines – some of it may have been in that walled garden, or some of it may be confidential information.

We've all seen in the paper the lost laptop stories. Pick any one you like. From an Iron Mountain perspective, after you've done all that hard work of putting your spyware and your antivirus and doing your inventory management and your software distribution, if you don't add the 1–2 punch of data recovery and protection from loss or theft, you're really falling short. Because it *is* a frontier. These things do go missing.

The last thing any organization needs is to be putting customer or personal information or even corporate secrets at risk. The other thing they don't need to be doing is losing valuable information that exists on those machines.

We think there's a mission for these removable devices – whether they're laptops or other things – the mission is being able to protect that from both kinds of loss. When it's in someone else's hands, or when you don't have it yourself, it's incredibly important.

Although you may not be able to draw that direct line from the walled garden all the way to the end – especially on the laptop side – you've got to have a safety net that protects you.

I was going to say – the other piece is when you look at the *sprawl of information*. Once something is in electronic form, stopping it from moving around is really impossible to do. Again, you may have a small subset of your information where you're able to do that successfully. But the notion of having multiple copies of things in different places – needing to figure out how you can stop them from moving around, or at least knowing when they do – is part of where we're all going.

Keeping one master copy, or being able to archive one and create shortcuts or links to the other is another management technique for at least trying to gain some control.

MM: So two related questions.

One – have you seen any technical developments for conducting a scan of all of the media objects such as Office Documents and rich media objects on laptops or desktops as they connect back into the corporate network?

Two – the Macintosh. A great portion of the high added value media first comes into being on a Mac, and sits on Mac servers. How does Iron Mountain Digital deal with what I'll call the 'Mac User Space'?

JC: There's the ability to get those solutions – some in a direct way, and some as a side effect. For example, if you're doing data protection of your laptops, you've actually got an inventory of everything that's on those remote machines in a centralized location. That's the first question for the scout. Does the scout actually have to go in the field? If there are enough agents out in the field reporting back what's there, you can look somewhere that's actually under system control in a more accessible location. One of the big benefits of Iron Mountain's PC and Mac protection solution is enterprise visibility. That's both the backup policy and information access, assuming appropriate security credentials.

MM: Scanning the disk image is a far more elegant solution than tying up a laptop as it kind of thrashes back and forth for many minutes or hours to sync up.

JC: Right. I guess, Michael, my point is that if you're scanning the disk image for other purposes – for example, so that you can have a backup of it – repurposing that scan for other purposes is a less intrusive way to do it than doing a whole other scan.

MM: I couldn't agree more.

JC: So that's one notion of the *frontier scout*: the ability to run reports and have visibility in a centralized or quasi-centralized location.

The other piece is in addition to being a scout. It's to decide whether you need active protection from doing certain things. For example, someone has a document on their desktop – e-mailing it out to someone via webmail or a non-corporate mail environment. That's in the case where you don't necessarily even need a scout, but you need someone who's going to be able to do *active policy enforcement*.

I think that's still an emerging area. There are solutions out there that can handle some of that. But it's a question of making sure that they're applied surgically, so that it doesn't create a barrier for folks also getting legitimate work done. That's one piece.

The second piece is what about the Mac question.

MM: Haven't we been asking that question for 20 years? The only difference is, the answer keeps changing.

JC: Yes. You know, my first computer was an Apple II, and my second was a Mac. So I've certainly lived on both sides of that equation, personally. When you look at what solutions are there specifically for the Mac, I'd say there are fewer on the frontier side of things. But also from a connectivity standpoint, the more we're getting to open standards, acceptance of content-based storage systems – the more the folks are getting onboard with the notion that you don't need to collect and manage everything in one spot ... I'd say the outlook is more promising for the Mac world.

It's a question, really, of a division between who's making the investments to be on that platform – and also making sure that from an openness standpoint, solutions are not tied to a specific platform.

I'd also say that when you look at things like *Web 2.0* and the *Enterprise 2.0 spin on Web 2.0*, it's the notion of a services-based environment. One where you're able to consume and do work on different things from different places. That's a much more likely scenario for success from an open platform standpoint.

MM: As we begin to wrap up on our interview here, Jim, I'd like you to speak to what – to my mind – becomes a tsunami-like wave of change, specifically as it relates to compliance, storage management and so on.

I'm asking this question to bring us up to speed on *XBRL – the business reporting language*. SEC and IRS and Edgar and FASB and all of the regulatory bodies have all gotten together and said, 'Here's an XML standard for publishing financial information.' More specifically, a federal mandate now, as of – I think – 2010, to publish all the financial information in your annual reports, 10Ks, 10Qs and disclosures in XBRL-compliant formats.

That basically means that all your financial data – all the financial data of all publicly held firms are, as of 2010, going to share an explicit XML schema. What are therefore some of the implications of that?

JC: Michael, I'm sure you're familiar with the old joke of the great things about standards. 'There are so many to pick from.' XBRL is a great example of the great thing about a standard in a segment is when it can get to the point where it's something that gains adoption. Then it becomes a real standard.

MM: This is beyond a real standard. This is now a regulatory compliance requirement.

JC: That's where I was trying to go with it, Michael. It's one thing to be an open standard. It's another to try to come up with a licensed model that allows folks to use it. Beyond that, real adoption comes either out of necessity and market momentum, or because of mandate and requirements from a compliance standpoint.

Whatever one of those two drivers there is – and in the case of XBRL, it's really the mandate side of things – it can help tremendously specifically in terms of preservation, archiving, and information management. Then you can reduce the number of permutations you need to deal with, and you can focus on the value-add.

In my mind, this is very similar to where the industry ended up around *SEC compliance for*

messaging. Most systems were doing their SEC compliance with Internet-standard format as the back-end underlying format. That allowed for folks to switch from vendor to vendor if needed. It allowed for – on the discovery side – folks to focus on doing things in a standard way – to focus on the processing of the information rather than on the conversion of formats.

I think that whenever you get to a point where you're able to say, 'We've got to the point of stability for a subset of information in terms of format,' it allows you to focus on more of the important aspects, rather than engaging in a format war. Of course, the process to get there is sometimes painful. But in the end, I think it's worth the journey.

If going on the journey is a non-negotiable, all the more likely that it happens quickly.

MM: I'd like to ask you to speculate – in this XBRL standard land, basically the schema specifies how to characterize money, time, interest rates, financial ratios. It also specifies a uniform way of characterizing the identity of a company, of subsidiaries, officers, roles and responsibilities within an organization.

Doesn't this then create almost a normative model by which to start reverse engineering all of the metadata and the taxonomies so that it's now compliant with this XBRL standard?

Thereby almost imposing the uniform metadata structure on heretofore lots and lots of permutations and differences that required really complex technologies to normalize?

JC: I think it's a mixed answer.

I think for things around currency and financial reporting that there's a very high likelihood we'll get there. For things around solutions around organizations' enterprise directories and things like that, it's a question of how far an entity needs to go to meet the requirements for XBRL, relative to reporting.

I guess the thing I'm not certain of is whether XBRL from an organizational hierarchy standpoint becomes a bolt-on or subset or whether it becomes a dominant linkage point to an organization's directory. I think organizational directories' maintenance from an electronic

standpoint is one of the things that's a barrier to a lot of things. I don't think that XBRL is going to be the thing that pushes that over the finish line.

MM: Clearly that's not. I'm still in the process of noodling through how this event affects all the things that we know about it.

JC: One thing at least for me seems to be clear. In one respect, you could almost view it as a reporting engine or output function. But as it becomes more central to what needs to be done, the optimizations embed the XBRL capabilities internal to what a lot of software systems will do, relative to output.

MM: To finish up, would you have any summarizing comments in terms of what we've touched on?

JC: I'd just come back to some of the points we made early on around preservation. As you look at what you called the wave of changes that have come down – focusing on providing capability at scale is really important.

The notion of figuring out what the right usage models and patterns are is step one. Step two of that is making sure that they actually work at scale, and over a long time period.

For us, those are some of the things we focus on. The other thing is inconsistency, especially given the environment today, with the benefit of 20/20 hindsight, never looks good.

When you think about doing things consistently, at scale, over an extended time period – that's a tall order. Iron Mountain is in the business of helping our customers with some of these hard problems. In the end, this reduces their risk and strengthens their compliance posture. We've been doing this since 1951. It never goes out of style.

MM: Great. On behalf of the *Journal of DAM*, I want to thank you. I look forward to having a conversation with you in another year or two, to get reprised on new developments that we didn't know to talk about unless they'd become central to the whole life cycle of digital assets – with an emphasis on end-of-life cycle preservation.

JC: Thanks a lot. I really enjoyed our discussion.