GiantSteps
Media Technology Strategies

515 West End Ave., Suite 11A
New York NY 10024
212 712 9190
fax: 212 873 1351
http://www.giantstepsmts.com

Enterprise Content Integration

with the

Digital Object Identifier

A Business Case for Information Publishers

By **Bill Rosenblatt**

June 20, 2002

DOI: 10.1220/whitepaper5

Table of Contents

Executive Summary

Many of the current challenges of information businesses, particularly large, diversified ones, are addressed by digital asset management (DAM) solutions. Yet information companies have had difficulties adopting DAM solutions at the enterprise level, due to excessive costs, implementation time, risk, and other factors.

In this white paper, we discuss those difficulties and outline a new approach based on the concept of Enterprise Content Integration (ECI) and the Digital Object Identifier (DOI)[1]. We define an architecture for DOI-enabled ECI in large information providers and demonstrate the lower costs and ease of deployment as compared to an enterprise DAM architecture.

After the explanation of DOI-enabled ECI architecture, we examine four scenarios that demonstrate specific types of quantifiable business benefits realized by information providers who use DOI-enabled ECI. Each scenario reflects a single type of benefit, but in reality, any single information provider may well realize combinations of these benefits, as well as others described in this paper but not quantified via any specific examples. The four scenarios and financial benefits included here comprise the following:

- Cost avoidance of approximately $120,000 for a vertical market information publisher building a new cross-brand Web portal.
- Annual incremental revenue of $700,000 for a periodical publisher from being able to publish more books per year of repurposed periodical content.
- A 94% reduction in staff effort, representing a potential savings of over $400,000 per year, for a textbook publisher building Web sites to accompany textbooks.
- Over $1.2 Million in incremental revenue for a publisher of financial information from selling documents through third-party investment selection tools.

Information providers can use the scenarios as models for elements in their own business cases which, when combined, will easily justify the cost of ECI implementation.

Introduction

Challenges for information providers today

Information providers face a variety of challenges as they move to expand their businesses. They must develop larger numbers of new products each year in order to address new market opportunities, competitive threats, and new ways of reaching customers. These may include new product delivery modalities, distribution channels, and integration of content with tools that customers use to learn, perform their jobs, or get entertainment.

Large information companies, in particular those that comprise several brands and media types, have additional challenges of creating synergies among business units in order to

[1] "Digital Object Identifer," "DOI" and "DOI.ORG" are trademarks of the International DOI Foundation (IDF – www.doi.org), and are registered in the U.S. Patent and Trademark Office.

justify their existence as conglomerates. Such companies' annual reports promise cross-business-unit synergies to shareholders, and some of them have indeed delivered on synergies in such areas as back-office processes, advertising sales, and cross-brand marketing. But cross-business-unit product development and production is something that many large information companies can only do on an ad-hoc basis, without realizing any economies of scale.

Finally, all information businesses are constantly looking for ways to make their operations more efficient and eliminate unnecessary production resources across brands and media types.

Digital Asset Management addresses the challenges

These challenges have become well known over the past several years, and a technology-based solution paradigm, known variously as digital asset management (DAM), media asset management (MAM), digital libraries, and content management (CM), has emerged to address them. DAM systems are repositories of content that can be searched, browsed and accessed over a content provider's network by internal users and possibly external users such as business partners and customers.

A DAM system typically consists of: a database on a server, which stores content and metadata (information about content); client software, often delivered through a web browser, that provides browsing, searching and other features; and other functions that may include asset check-in/check-out, workflow management, tight integration with content creation tools, and interfaces to distribution technologies.

Content providers have been experimenting with and deploying DAM and related systems since the paradigm came into being in the mid-1990s. Although many agree that DAM systems, when properly deployed, can contribute significantly to the business goals described above, nevertheless truly successful DAM deployments at the enterprise level in diversified information companies have been few and far between. In this white paper, we examine a new approach to enterprise-wide DAM deployment. The new approach, called Enterprise Content Integration, should lead to DAM deployments that not only are more likely to deliver the promised business benefits, but can do so significantly more quickly and cheaply than the enterprise DAM approach that many diversified information companies have attempted to implement.

The Original Vision: enterprise DAM

Original enterprise DAM architecture

Just about every large content company has attempted to build a digital asset management system for the entire enterprise according to a simple yet compelling vision, which is illustrated in Figure 1.

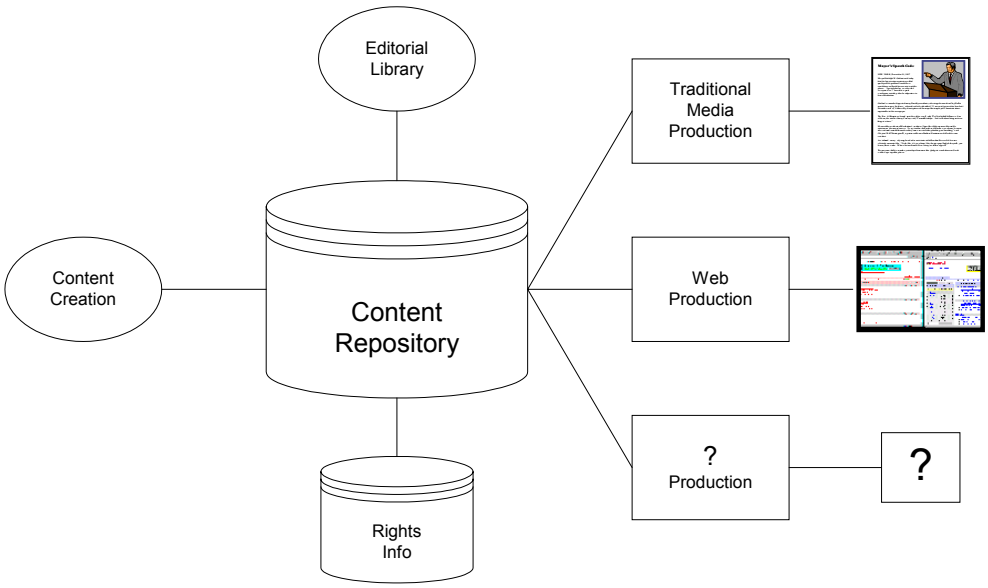


Figure 1: The original simple vision of enterprise DAM.

The original vision was predicated on the enterprise networks and client-server databases that were becoming common in the mid-1990s. It included a centralized content repository, which would contain content of all types, including text, images, page layouts, audio, video, etc. It would be accessible to everyone throughout the enterprise over the network. Content creation, editorial, and production people could access the system to browse content and search for it according to keywords and other criteria, thus allowing for easy repurposing. New content would be put into the system, along with metadata that described it. The system would be able to automatically feed various production and distribution processes, including prepress and the then-emerging Web. Information about the rights that the company had to alter and distribute the content would be held in a separate but linked database.

These systems were revolutionary, because they finally allowed information providers to manage content assets in a central, uniform way, just as any business manages physical assets and human resources. Content assets tended to be managed minimally – or not at all – on PCs, Macs, and local file servers, in drawers and file cabinets, and at outside service providers. With DAM, information providers could finally get control over their revenue-producing assets.

Brief history of enterprise DAM

Perhaps the first attempt to implement enterprise DAM for multiple media types at a large, diversified media company was the Times Mirror Digital Library System at Times Mirror Co., whose newspaper properties are now owned by Tribune Co. The project was started in 1994 and attracted some of the vendors of DAM precursors, such as text search engine vendors, image archives, newspaper editorial systems, and document management systems.

The first commercial DAM systems came out on the market around 1995, led at the high end by IBM's Digital Library, which Big Blue originally built for the Vatican. In 1997, after the Web had emerged as a commercial media distribution channel, the so-called content

management[2] paradigm came into being. Web content management took off much faster than content management for traditional media, because Web sites had to be built from scratch and thus could be designed around products like Vignette Corp.'s StoryServer instead of having to retrofit content management to existing editorial and production processes.

1999 saw the emergence of off-the-shelf products that could provide true enterprise-class DAM functionality across media types - products such as TEAMS from Thomson Technology Group (now Artesia), Cinebase (now eMotion MediaPartner), and North Plains Systems' TeleScope. It was also around this time that some of the large consulting and system integration firms began to develop competencies in DAM.

Unfortunately, the simplistic architecture of enterprise DAM systems, as implied in Figure 1, has led to very few successful implementations. Instead, many content companies have implemented various different content-storing systems at departmental or business-unit levels, with little or no integration among them.

The first thing that happened was that content companies created Web outlets for their content. But instead of implementing new systems that could publish to print and Web, content companies left their traditional production systems alone and built parallel systems for Web publishing; see Figure 2. In such situations, there was often a jury-rigged process for translating content intended for traditional media to HTML.

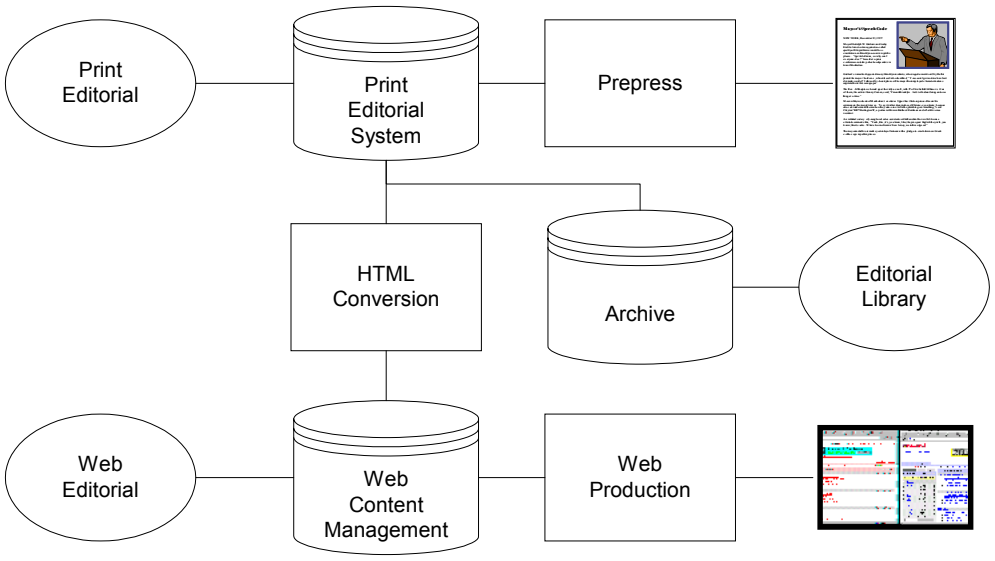


Figure 2: the rise of the Web led to a proliferation of standalone systems for different output media.

This proliferation of monolithic systems continued as media types and new product development efforts expanded. As Figure 3 suggests, some information providers have gotten to the point where they have separate systems for print (or other analog media)

[2] Although the term "content management" first applied to systems for managing a variety of media types, it came to specifically describe Web content management through the popularity of tools like Vignette StoryServer. "DAM" became most commonly used to describe content management for multiple digital media types. Film studios, broadcasters, and others who store large analog media assets but manage them digitally often use "MAM".

production, Web publishing, wireless distribution, secure distribution through digital rights management (DRM), syndication to other Web sites, and possibly others.

Furthermore, many media companies have product catalog databases that are connected to marketing and sales systems but not to production or distribution, as well as rights, permissions, and contract management systems that sit on (often several separate) standalone PCs in legal and rights clearance departments. Finally, the systems that handle the actual vending of the content, tracking and reporting of sales within the back office, etc. are entirely separate – often so separate that transactions that take place in various digital distribution channels are not represented in the traditional financial systems, so they cannot easily be accounted for in normal P&L reporting.

Each of the "drum" figures in Figure 3 (except for the product catalog, financial reporting, and rights/permissions/contracts systems) is a complete database of content in some format that is particular to an output type. There is little notion of a centralized repository of content that can be repurposed; today's processes are ad-hoc, incremental extensions of existing processes.

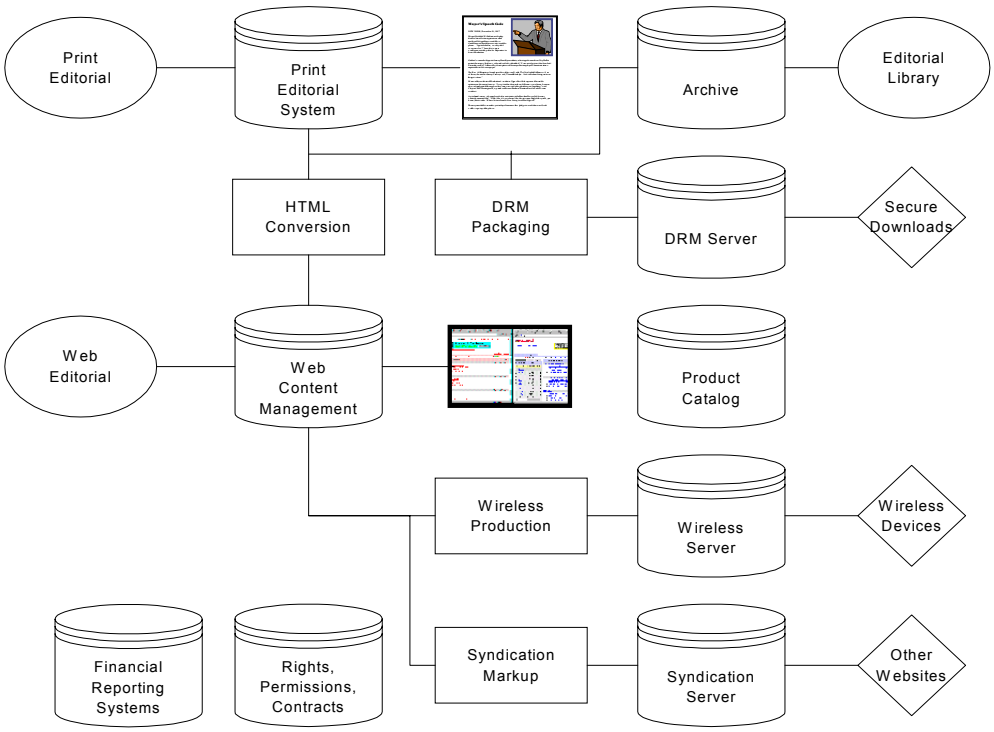


Figure 3: Today's large media companies often have many different systems for different types of production and distribution, with little integration among them.

What went wrong?

The industry has learned quite a lot since the early deployments before 2000. Typical enterprise-wide DAM implementations are very expensive – often $10 Million or more – and can take a very long time – 18 to 24 months – before any business benefits are seen.

Enterprise DAM implementations are also risky, because they can be disruptive to existing mission-critical business processes. Deploying a single DAM system across multiple production groups can force all production groups to modify their workflows to fit a set of capabilities that may not apply to their particular situations. In addition, a single enterprise-wide DAM implementation often implies a single metadata schema for all content throughout the enterprise, which is often impractical for organizations that handle different types of media (e.g., magazines, books, music, video, databases). Even for companies that only handle a single media type but are organized via autonomous, market-focused business units, this type of unification can be unacceptably disruptive to established processes, and for that reason, politically challenging.

Finally, enterprise DAM projects were attempted under assumptions that we now know to be dubious, such as:

- *Everyone should see all content*. Not only should it not be permissible for everyone to see all content, but in many cases, it is never necessary.

- *Digitizing legacy content is fast and cheap*. In many cases, deployment of DAM systems has depended on digitizing large collections of content that only exist in analog (physical) form, such as image libraries, audio clips, and even some text works. Such digitization has often turned out to take years and involve costly personnel and storage media – factors that rarely appeared in business cases for early DAM implementations.

- *Generating lots of quality, consistent metadata is trivial*. Metadata creation is the most significant bottleneck in all DAM implementations, especially in types of information providers that have little experience in metadata generation. Metadata creation takes time and is a skill that must be learned, not something that just anyone can do well; moreover, doing it properly requires setting both metadata standards throughout an organization and policies for adhering to those standards.

- *Content equals products*. It is true that media, publishing, and information companies all sell products that are based on content, but one of the key potential benefits of the digital revolution is to separate content in "raw" form from packaged, finished products. In today's world, one piece of content typically finds itself in many different finished products. DAM implementations should treat products differently from raw content – if they hold finished products at all.

A New Approach: Enterprise Content Integration with the Digital Object Identifier

An alternative to enterprise DAM

Given the many challenges to successful enterprise DAM deployment, what should large content providers do now? Technologists often have the temptation to tear everything up and start from scratch, and there are certainly many vendors out there who would encourage such an approach. While it is healthy for a content provider to think in terms of radical new enterprise content architectures, there is another approach that lets information companies live with what they have and acquire many of the benefits of enterprise DAM with far less cost, time, and risk.

The term Enterprise Content Integration (ECI) has been used to describe an approach similar to that described here[3]. ECI is an approach to digital asset management at the enterprise level that eschews creating a single, monolithic repository (as in Figure 1) in favor of adding a layer of functionality that integrates content stored in various types of systems at departmental or business unit levels.

The ECI approach satisfies a subset of the usual list of requirements that enterprise DAM deployments are intended to fulfill. Specifically, the requirements are these:

1. *Know what you have*: Users should be able to look at catalogs of content available throughout the enterprise.

2. *Find what you want*: Users should be able to search and browse for desired content items to be used or repurposed in products.

3. *Know where it exists*: Users should be able to easily see where a desired content item exists within the organization – in what business unit, on what system, etc.

4. *Be able to get it*: Users and production tools should be able to retrieve the desired content item for their own use.

Enterprise DAM systems typically are based on requirements that are out of scope in the ECI approach, such as:

- *Central asset store*: ECI does not include building a single, centralized repository for all digital assets.

- *True cross-media content creation*: In the ECI paradigm, content creation processes are largely left alone, meaning that content may be created specifically for a single output medium (e.g., print).

- *Single metadata scheme for entire enterprise*: Building a single repository normally entails creating a single metadata schema for all content. The ECI paradigm allows for multiple metadata schemas, as long as they are normalized to a common subset.

- *Process reengineering*: enterprise DAM deployment usually requires that content creation and production processes be redesigned to fit with the central repository. The ECI paradigm leaves these processes largely intact.

Enterprise Content Integration implementation architecture

To implement ECI, we first define a lightweight set of metadata attributes that can apply to all types of content in an organization. This enterprise-wide metadata scheme should use the Digital Object Identifier (DOI)[4] as a unique way of identifying each content item. Then for each content-storing system: if the system does not store any metadata (e.g., the

[3] Rosenblatt, Bill. Enterprise Content Integration: Next Step Beyond DAM? *The Seybold Report,* v. 1, no. 20, January 21, 2002, pp. 10-14.

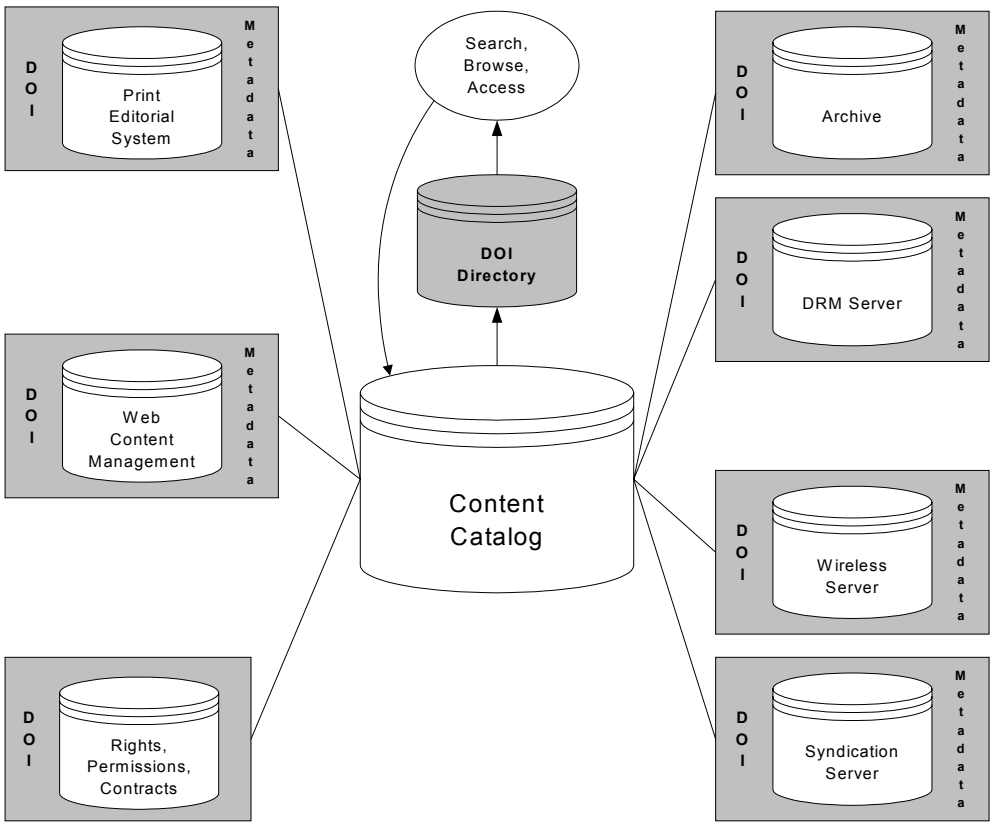
[4] The DOI is an identifier for content objects on the Internet, supported by a global routing system that links permanently and reliably to the object itself or to any related objects, information, or services. Unlike pointing to an object via a URL, where the link "breaks" if the object is moved to a different location, a DOI-based link can be re-pointed to different locations and/or to different services even if they change over time. For information about the DOI see www.doi.org/about_the_doi.html or www.contentdirections.com.

system is simply a file server), then metadata must be created for all content in that system that fits the lightweight schema. If the system does store metadata (e.g., an editorial, content management, or departmental DAM system), then its metadata must be mapped and normalized to the enterprise-wide lightweight metadata scheme.

After the metadata has been defined, functionality can be set up to collect metadata from every content-storing system, on demand or on a regularly scheduled basis, and store it in an enterprise content catalog. The content catalog should have a user interface that allows searching and browsing, thus satisfying the first two of the requirements above. DOI resolution through a DOI directory satisfies the third and fourth requirements: knowing where a particular content item lives and being able to retrieve it. The DOI Directory can be internal to the information company, or it can be the global distributed DOI Directory.

Figure 4 shows the basic architecture of a DOI-enabled ECI approach. Standard metadata should be assigned, and DOIs registered, for every content item in each content-storing system. DOIs must be unique. The Content Catalog is set up to "harvest" metadata from each content-storing system using interfaces appropriate to the individual content-storing systems, which could include reading XML data, executing SQL queries, parsing media file formats, or using the proprietary APIs of DAM, content management, editorial, or other types of systems. The "harvesting" functionality will store each content item's DOI as the location of the content.

In addition, the architecture shows a rights, permissions, and contract management system that is linked via DOIs to the content catalog. Ideally, this allows both human producers and automated distribution processes to check rights information on a content item to ensure that it can be used in the product or distributed over the distribution channel, and if necessary, to initiate any processes (such as permission letter, financial transaction generation, or legal review) that must take place in order to secure such rights.



• Figure 4: Enterprise Content Integration architecture using a DOI Directory and a content catalog.

The entry in the DOI directory for each content item will have at least one pointer associated with it: a URL that causes the item to be retrieved. If the system holding the content item is a file server, then a simple file URL suffices for this purpose. If the system holding the content item is a relational database, then the URL could point to a CGI script containing a SQL query, and the URL could contain parameters to feed to that CGI script. If the system holding the content item is a DAM, content management, or other type of proprietary system that has an API, then the URL could point to a CGI script that uses the API to retrieve the content item.

This architecture can also take advantage of the DOI's multi-linking capability, wherein a DOI can point to multiple URLs. For example, a DOI can link to multiple versions of a given content item, such as the QuarkXPress, PDF, XML, and HTML versions of a document, the TIFF, Photoshop, JPEG, and thumbnail versions of an image, or several different encodings of an audio or video clip.

In addition, multi-linking can be used to create additional pointers, either at first or over time on an ad-hoc, evolutionary basis, which lead to additional information or services relating to the item. In addition to the examples just given, which all involve the content itself and/or the original native application files that created the item, these additional multi-links can point to the rights information within the rights system, or to the licensing terms for various distributors, or to the system holding the author's contract terms.

DOIs offer several strategic benefits

Although it is possible to implement Enterprise Content Integration without DOIs, there are many benefits to using the DOI as an identifier and linking mechanism within the ECI architecture. Some of the most important benefits are:

- The DOI directory, as a level of indirection, acts as a buffer zone between the DOI and the actual source of the content, thereby ensuring that the identifier never becomes obsolete if the location of the content changes over time (e.g., from putting in new content-storing systems, file system reorganization, company reorganizations, acquisitions and spinoffs, etc.), and thereby making it easier and cheaper to update all content locations whenever they change.

- The DOI is open and flexible enough to be usable, as described above, with the widest possible varieties of content and content-storing systems, including those yet to be invented. The DOI syntax was designed so that it can subsume any existing type of identifier, including ISBNs for books, ISSNs for serials, SICIs for journal articles, CUSIPs for securities, SMPTE/EBU UMIDs for video clips, ISANs for MPEG-2 and MPEG-4 files, and so on.

- Although DOIs are useful as linking mechanisms within an information company (as we will see in the first three business case scenarios below), they can also be easily and quickly made available to the outside world, to facilitate content distribution to partners and customers (as shown in the fourth scenario below). This is done simply by registering an existing internal DOI with the global DOI Directory through a DOI Registration Agency, such as Content Directions Inc. The global DOI Directory is virtually unlimited in its scalability.

- Unlike some organizations' proprietary content identification and cataloging schemes, DOIs are based on standard technology (invented by Dr. Robert Kahn, co-inventor of the TCP/IP protocol on which the Internet is based) with a growing body of third-party vendor tools and support. Adopting DOI technology means eschewing the cost and risk of proprietary development and maintenance.

ECI approach requires less effort and expense

It is possible to compare the level of effort and cost required to implement a full enterprise DAM system and the lightweight, DOI-based approach described above. We do this by examining the major steps required to implement both approaches and looking at the relative costs of each step.

Here are the major steps required to implement a system using the DOI-based Enterprise Content Integration approach:

1. *Create a common layer of metadata standards*. This metadata schema can be extended at the departmental or business unit level to meet its individual needs, but there should be a common core across the enterprise.

2. *Acquire metadata creation skills*. Many types of information providers do not have people in house who are skilled in the art of consistent metadata creation. This is not a trivial task that can be assigned to production assistants without prior training or done without agreement on keyword taxonomies and other elements of metadata.

3. *Populate the common metadata set and register DOIs*. Some modifications to editorial and production processes must be made in order to create metadata for content according to (at a minimum) the common layer defined above. DOI registration can be done locally or to the global DOI Directory with software tools from Content Directions Inc.

4. *Convert content to cross-media format after authoring on an as-needed basis*. Do not implement cross-media content creation tools in place of existing content creation tools; instead, use technology and processes that convert existing content formats to cross-media format (e.g., XML).

5. *Build metadata catalog and metadata harvesting functionality*. This can be done through custom development and/or by using some available off-the-shelf tools.

6. *Integrate the catalog and DOI resolution with content creation and distribution tools.* Users should be able to retrieve content through the system directly into their favorite content creation tools, and automated distribution functions (e.g., DRM packagers or syndication servers) should be able to access content directly through the catalog.

Contrast this with the major steps necessary to create an enterprise DAM system:

1. *Create metadata standards for the entire enterprise*. It is very difficult to create metadata standards that span multiple media types and product lines, and are tractable in size.

2. *Acquire metadata creation skills*. This is a vital step (if necessary) for both approaches.

3. *Populate the large enterprise metadata set*. Populating a large enterprise-wide metadata set takes an inordinate amount of manual effort. Some technologies exist that can automate part, but by no means all, of this process.

4. *Migrate content creation to cross-media tools and processes*. This entails drastic reengineering of content creation processes and may limit production (e.g., page layout) flexibility for existing products.

5. *Build a central asset store, and migrate content from departmental systems to it*. This requires major server, database, and storage technology acquisition. Content migration is a nontrivial, time-consuming task. It also may take systems out of operation that are not off the books yet.

6. *Integrate the central asset store with content distribution tools*. Distribution functions should be able to extract content directly from the central asset store.

Table 1 shows the relative levels of effort involved in the two approaches. Costs are L (low) through VH (very high). Exact dollar amounts depend on the size of the organization, number of business units, number of content items, and many other factors. If we were to assign values of 1 for L through 4 for VH, we would find that the total of resources necessary to implement DOI-enabled ECI is 11, while the centralized enterprise DAM approach has a cost of 21 – roughly double the resources necessary.

DOI-enabled ECI	Primary Resource	Cost	Enterprise DAM	Primary Resource	Cost
Create common layer of metadata standards	people	L	Create metadata standards for entire enterprise	people	H
Acquire metadata creation skills	people	H	Acquire metadata creation skills	people	H
Populate lightweight common metadata set	people	L	Populate large enterprise metadata set	people	H
Convert to XML on back end as needed	tools	M	Migrate to XML-based content creation tools & processes	tools & people	VH
Build metadata catalog & harvesting functionality	tools	M	Build central asset store	tools	H
(Leave content where it is)	-	-	Migrate content to central asset store	tools & people	H
Integrate with content creation & distribution tools	tools	M	Integrate with content distribution tools	tools	M

Table 1: Relative costs of DOI-enabled Enterprise Content Integration vs. enterprise DAM.

As mentioned above, enterprise DAM implementations can cost upwards of $10 Million for a large, diversified content business, and they typically take 18-24 months before producing tangible enterprise-wide business benefits[5]. During that period of implementation, business units will have their operations disrupted for a significant period of time as they cut over to the new system.

In contrast, the approach of DOI-enabled ECI can save millions dollars and many months of implementation time; in fact, rough estimates of the up-front financial costs of each approach[6] lead to the conclusion that DOI-enabled ECI costs a quarter to a half as much to implement as enterprise DAM costs; thus, an organization for which enterprise DAM costs $10 Million could implement DOI-enabled ECI at a cost in the low Millions. Moreover, DOI-enabled ECI provides most of the benefits of enterprise DAM while minimizing disruption to ongoing operations.

Business Case Scenarios

In the remainder of this white paper, we present several scenarios that are representative of product development initiatives at publishing companies and other types of information providers.

Each of these scenarios shows an example of a different type of business benefit that can be derived from DOI-enabled Enterprise Content Integration. They are meant to be examples of the kinds of business case elements that are effective in cost-justifying any sort of content management implementation. Readers are invited to examine these scenarios, use them as models for their own organizations, and replicate or combine them as necessary to build complete business cases for enterprise-level content management projects.

[5] These figures assume multiple business units and include consulting costs and time for doing requirements analysis, software package selection, change management, and so on.
[6] Estimates available on request from the author.

Although we have described separate scenarios to illustrate the specific benefits of ECI with the DOI, it is reasonable to expect that companies will actually realize multiple benefits that accrue from several scenarios cumulatively over a period of a few years. Actual content management business cases are made on the basis of multiple product development and/or cost reduction initiatives rather than on single initiatives or "point products." If half a dozen such initiatives are identified, then the total financial benefits – as suggested by the numbers below – easily outweigh the costs of ECI implementation.

In each of these scenarios, we assume that editorial and production resources have fully loaded annual compensation of $75,000 and work 40-hour weeks.

1. Cost Avoidance in Cross-Brand Product Development

A common business driver for DAM deployment at the enterprise level is to develop products that span brands and business units. As implied in the Introduction, efficient cross-brand-unit product development is a major hurdle that diversified content companies must jump in order to fulfill their promises of synergies that lead to shareholder value.

Scenario

In this scenario, we imagine a typical vertical market information publisher.

Electronics World Group (EWG) is the dominant provider of information for the electronics industry. EWG has accumulated, in addition to its flagship *Electronics World* magazine, over a dozen businesses that produce information for the electronics industry. These include an assortment of monthly and weekly trade magazines, weekly and daily newsletters (15 total publications), and databases of electronics vendors and products.

EWG's management wants to aggregate all of its content into one Internet information portal, ElectronicsWorld.com, which will contain content from all of EWG's publications. Magazine and newsletter articles will be organized by subject area (e.g., semiconductors) and contain links to EWG's databases for products and vendors mentioned.

We assume that EWG magazines and newsletters are produced using separate publishing systems, such as Quark Publishing System for the magazines and a combination of Microsoft Word and Adobe Acrobat for the newsletters. Furthermore, we assume that although each magazine title currently has its own, independent Web site with magazine content converted to HTML, these sites will be phased out in favor of the ElectronicsWorld.com portal.

ElectronicsWorld.com will have some combination of Web publishing system and human production editors who will take content from the various sources, put it in appropriate places on the ElectronicsWorld.com site on a daily basis, and create links to the vendor and product databases by hand.

Business Benefits

Without any of the capabilities discussed above, ElectronicsWorld.com production editors would have to do the following to publish content to the site each day:

- Search source files across several EWG publications to find relevant content and manually group it by subjects that that correspond to areas of the ElectronicsWorld.com portal.

- Copy and paste articles source files into the Web publishing system for ElectronicsWorld.com.

- For each product or vendor mention in an article, search the appropriate database for the right entry and build a link into the Web article that executes the right database query.

We assume that ElectronicsWorld.com has 5 editorial sections by subject area. Each section has an average of 10 new articles every day, for a total of 50 articles per day. Each article mentions an average of 5 vendors or products, for a total of 250 links to the product and vendor databases per day.

We calculate the time it takes to update a the site on a daily basis as follows:

- A section editor searches several different magazines and newsletters for relevant content. We assume that each section editor has to search 10 of the 15 publications (because she knows that 5 of the 15 will not contain relevant content). Each search takes 5 minutes, and each article takes 6 minutes to retrieve, send (possibly by email), copy, and paste into the Web publishing system, for a total of 50 plus 60 or 110 minutes per section, or 550 minutes total per day.

- For each of the 5 vendors or products mentioned in each article, a link to one of EWG's databases must be constructed. We assume that section editors must manually search for the correct database entry for each link: advertisers pay for these links, therefore they must be 100% accurate, and automated tools will not provide acceptable accuracy in finding the correct entries. It takes 2 minutes to find the relevant entry and construct each link, for a total of 10 minutes per article or 100 minutes for 10 articles in each of the 5 sections, for a total of 500 minutes per day.

In total, section editors must each spend about 3.67 hours creating their sections of ElectronicsWorld.com each day, or a total of 18.33 staff hours to perform the above tasks for the entire site.

		Before	**After**
Task	Search and browse for relevant content	50 min.	5 min.
Tasks per day	5 sections	250 min.	25 min.
Task	Retrieve and enter articles into Web publishing system	6 min.	2 min.
Tasks per day	50 articles	300 min.	100 min.
Task	Find product/vendor database entries and create links to them	2 min.	1 min.
Tasks per day	250 links, 5 links/article	500 min.	250 min.
Total per day		18.33 hrs.	6.25 hrs.
Days per year	260	4766 hrs.	1625 hrs.
Annual cost avoidance			**$117,788**

Table 2: Business benefits in cross-brand product development scenario.

Table 2 summarizes the cost avoidance possible with DOI-enabled Enterprise Content Integration; Before and After columns note task time without and with DOI-enabled ECI. Here are specifics of how DOI-enabled ECI provides cost avoidance benefits in line with the four requirements listed on p. 9:

1. *Know what you have*: We assume that section editors are familiar with the subject matter of each of EWG's 15 publications, so there is no savings in discovering content from possibly unknown sources. Yet this additional savings does pertain in cases where the publisher wants to link outward to other parties' sources of information, if those sources have also been DOI-enabled.

2. *Find what you want*: With DOI-enabled ECI, production editors would search all news content for relevant articles through a single interface instead of searching 10 publications separately. This means that for each section, searching is done once (taking 5 minutes), and retrieval of articles into the Web publishing system takes 2 minutes each, for a total of 25 minutes instead of 110 per section, or 125 minutes per day. Database searches in this scenario would not take less time because the editor will know up front which database to search for which type of item (product or vendor).

3. *Know where it exists*: a DOI-enabled ECI capability would tell the production editor where each selected article and database entry lives, although this does not provide any significant time savings.

4. *Be able to get it*: a DOI-enabled ECI capability allows a production editor to retrieve the content easily from a single place. It also allows her to trivially create an actionable link to the content on the ElectronicsWorld.com Web site. This reduces the 2 minutes per link (to locate the relevant database entry and construct a link) to the simple task of cutting and pasting DOI links directly from the search results into the web publishing system, a task that should take half the time. The result is a savings of a minute per link or 250 minutes per day.

As shown in Table 2, DOI-enabled ECI saves a total of about two-thirds the effort required to perform the relevant tasks. It is a savings of about 12 hours per day, or over 3100 hours (equivalent to more than 1.5 headcount) saved during a 260-day year. This translates to **cost avoidance of approximately $120,000 per year in a vertical market information company with 15 publications and two databases**.

Additional Scenario Applicability

Here are some other situations where the architecture presented here could lead to cost avoidance in cross-brand product development:

- An **educational publisher** building a unified distance learning offering that combines content from different brand units.

- A **financial services company** building a financial information portal for its customers that culls content from various sources, both internal and third party.

- A **diversified publishing company** building a Homeland Security Web site that culls content from various book, magazine, and professional information business units[7].

2. Scalable Product Development through Repurposing

As also mentioned in the Introduction, information providers face the challenge of needing to launch more and more new products each year, in order to keep up with the proliferation of delivery vehicles, new customers, and competitive threats. In order to meet these challenges, content companies must find ways to do more scalable product development – that is, develop capabilities of developing and launching more products per year, with a minimum of incremental resources, by making production more efficient and time-to-market faster.

Scenario

This scenario involves a major metropolitan daily newspaper or weekly news magazine.

SportsWeek is a weekly sports news magazine. Its marketing executives find that they can create incremental revenue by producing soft-cover books that summarize important sports events with articles and photos, and selling them as "impulse items" on newsstands, but only if they are produced and made available before interest in the event wanes. For example, they feel that they can sell 10,000 copies of a book (at $10 apiece, for $100,000 in gross revenue) that summarizes the NCAA basketball tournament, but only if it appears on newsstands within two weeks of the final game. *SportsWeek* has put out such books in the past, but the time and effort to produce them is too great to do as many of them as management would like.

We assume that *SportsWeek* pages are laid out in a standard magazine editorial system like Quark Publishing System. Article texts are kept in separate files in Quark CopyDesk or similar format. High-resolution photos are kept on a file server, and some of the photos that *SportsWeek* will want to use for the book products will not necessarily have appeared in the print magazines; furthermore, some of the photos will be from freelancers or agencies who may or may not have granted reproduction rights to *SportsWeek*. Books will be laid out using QuarkXPress, Adobe InDesign, or some similar page layout tool.

Business Benefits

We assume that each book contains an average of 50 stories and 100 photos. Without DOI-enabled ECI capabilities, the production editor would have to do the following:

- Search and browse CopyDesk files in the editorial file server for relevant articles.
- Search and browse images (potentially thousands) in the image file server for relevant photos.
- Retrieve articles from the editorial file server and copy them into the book layout system.
- Clear rights to use selected photos, and for those photos for which rights are too expensive, time-consuming, or impossible to clear, select alternates.

[7] See doi.contentdirections.com/homeland for an example of this.

- Retrieve photos for use in the book layout.

Table 3 summarizes the savings and incremental revenue in this scenario.

		Before	After
Task	Search and browse image files	2 min.	2 min.
Tasks per book		1000	200
Total image search		33 hrs.	6.7 hrs.
Task	Search and browse article texts		
Tasks per book	50	8 hrs.	1 hr.
Task	Retrieve texts and images into book layout	4 min.	2 min.
Tasks per book	150	10 hrs.	5 hrs.
Task	Clear rights for images	8 hrs.	1 hr.
Total		59 hrs	13.7 hrs.
Total savings per book			**$1,699**
Revenue per book	$10		
Products per year		3	10
Annual Incremental Revenue			**$700,000**

Table 3: Business benefits in scalable product development scenario.

Here are specifics of how DOI-enabled ECI provides time savings benefits, in line with the four requirements listed on p. 9, which lead to scalable product development:

1. *Know what you have*: We assume that the production editor is generally familiar with recent *SportsWeek* content.

2. *Find what you want*: With a search capability that spans both article texts and images, the editor can easily search and browse content instead of having to page through file after file to choose relevant items. The searching capabilities can dramatically reduce the number of files that the editor would have to examine. For image files, the number of browses could reasonably be reduced from 1000 to 200, and at 2 minutes per image, that is a savings of 26 hours. For article texts, the editor could search all articles at once rather than having to look through each issue's files, which may be in separate directories on the file server (or on more than one file server). This could reasonably be expected to save a day of effort.

3. *Know where it exists*: We assume that the editor knows how to access the file servers for article texts and image files; therefore there is no savings realized from seeing the file locations in the search results.

4. *Be able to get it*: The most important savings in this area for this scenario is the ease of looking up rights information. By examining rights metadata in the search results, or by using an image's DOI to link to a rights database, the editor can instantly check the rights that the publisher has to reuse the photo in the book and select an alternative if the rights are unavailable, too expensive, or take too long to obtain. Without such information, the editor would have to create a list of desired photos and check their rights on a separate system (most likely a standalone PC) or through a separate department, a task that could easily take a day.

In addition, the editor can save time per item by being able to search and retrieve article texts and images directly into the layout tool instead of having to import them by browsing files. A reduction in time from 4 to 2 minutes per item here results in a time savings of 5 hours of effort.

The total savings in time to select items for the book and insert them into layouts is 5.6 working days, or just over a week, for a production cost savings per book of about $1700. Of course, additional time will be necessary to produce special content for the book, such as front matter and cover layouts.

However, far more important than the savings is the ability to produce more of these books in a timely fashion. It is not simply a matter of shaving production time off the schedule; rather, it is a matter of getting the books out quickly enough so that people will be interested in buying them. DAM capabilities can bring production time down to a point where it will be worth doing many more of these products – say, 10 per year instead of 3. 7 additional books, each of which sells 10,000 copies at $10 apiece, will bring **$700,000 in incremental revenue per year for a weekly periodical publisher**.

Additional Scenario Applicability

Here are some other situations where the architecture presented here could lead to scalable product development through content repurposing:

- A **vertical market information publisher** or **large market research firm** producing customized information services to be site-licensed to large customers and delivered onto their intranets.

- A **management consulting firm** or **investment bank** assembling collections of reports and white papers into soft-cover books that are sold or given away to prospective clients as marketing tools.

- A **professional training company** putting together custom curricula for courses given at large corporate customers.

3. Cost Reduction on Existing Production Processes

Information businesses are always looking to reduce costs of existing production processes. This is particularly important in situations where production costs rise without corresponding increases in revenue, due to market forces such as customers' demands for product features or competitive threats.

Scenario

This scenario involves a publisher of college textbooks.

Higher education publishers are in a competitive "arms race" with respect to the Web sites they offer as companions to their textbooks. Such Web sites are inducements to college faculty, or administrators at state schools, to adopt one publisher's textbook over another's. In addition to features like bulletin boards, professor-led chats, and posted course syllabi, publishers often want to leverage other content they produce by linking to it

as related material. The rationale is that if a publisher has more related content available, the Web site can be more compelling.

Diploma Press, a leading higher ed publisher, wants to be able to create "related content" sections on Web sites for each chapter of its 100 textbooks in relevant fields, such as the social sciences and humanities. Assume that each textbook has an average of 15 chapters and is updated once a year. For each chapter, Diploma Press will have an average of 10 related content items that it can find and link to on the textbook Web site. That means that a total of 15,000 related content items must be linked to textbook Web sites across all of Diploma Press each year.

Business Benefits

Table 4 summarizes the potential cost savings in this scenario. Diploma Press uses a suitable Web publishing system to produce each textbook Web site. Here are specifics of how DOI-enabled ECI provides cost reduction benefits on existing production processes, according to the four requirements listed on p. 9:

		Before	After
Task	Discover and identify sources of potentially relevant content for each textbook throughout the enterprise	80 hrs.	1 hr.
Task	Find and identify specific content items to be linked for each chapter	2.5 hrs.	0.25 hrs.
Chapters per textbook	15	37.5 hrs.	3.75 hrs.
Task	Enter info about each item into spreadsheet	1 min.	0
Items per textbook	150	1.25 hrs.	0
Task	Create links to content items and place on Web page	3 min.	1 min.
Links per textbook	150	7.5 hrs.	2.5 hrs.
Textbooks per year	100		
Total		12,625 hrs.	725 hrs.
Annual cost savings			**$446,250**

Table 4: Business benefits in production cost reduction scenario.

1. *Know what you have*: The Web site producer would have to search for relevant items through content stored in a variety of places, ranging from those that are easy to find from those that aren't. It could easily take two weeks of phone calls (counting time taken to return those phone calls) to discover relevant content from throughout the enterprise for each entire textbook. DOI-enabled ECI reduces this task to insignificance. This is by far the largest time-saver: total savings is 200 staff weeks (at 40 hours/week) per year.

2. *Find what you want*: Once the relevant content is "discovered," specific desired items must be found by sifting through various files, servers, and publishing systems. These searches would presumably be done separately for each textbook chapter. DOI-enabled ECI reduces this from several searches to one single search for each chapter. If we assume that the number of searches reduces from 10 to 1 and each search takes 15 minutes, the savings for 15 chapters is 33.75 hours or, for 100 textbooks, over 84 staff weeks per year.

3. *Know where it exists*: a single search capability would eliminate the need for producers to keep separate track of which content item lives where, which they might do using spreadsheets. Assuming it takes half a minute to enter an item into a spreadsheet, it would take 75 minutes to enter each textbook's 150 related content items. Total savings from the elimination of this task across 100 textbooks is 125 staff hours or about 3 staff weeks per year.

4. *Be able to get it*: Producers can create links to the relevant content directly from search results, because each item has a DOI (or can have one created on the fly). Therefore creating links from the textbook Web site would be trivial, as discussed in Scenario 1 on p.15. Without DOI-enabled ECI, it could take an average of 3 minutes to create a link to each content item (depending on where the item lives), but with the capability, the time would be reduced to 1 minute. For 150 links in each textbook, the savings of 2 minutes per link translates to 5 hours, or for 100 textbooks, 500 hours or 12.5 staff weeks per year.

Total reduction in staff time from all these tasks is greater than an order of magnitude. Time is reduced from 12,625 staff hours (316 staff weeks) to 725 staff hours (18 staff weeks), **a 94% reduction, representing a potential savings of over $400,000 per year for a college textbook publisher**, which includes some reclaimed staff "down time" that could be used for other purposes**.**

Additional Scenario Applicability

Here are some other situations where the architecture presented here could lead to cost reduction on existing production tasks:

- A **medical publisher** saving time and money by locating and reusing images from throughout the organization instead of commissioning new ones.

- A **large journal publisher** inserting DOIs for referenced articles into online versions of its journals, where the references are to journals that the publisher also owns[8].

4. Increasing Market Share and Revenue through Customer Tool Integration

In markets for professional information, more and more customers are using Internet-enabled tools to do their jobs. Tools become more desirable when they incorporate industry-related information. This presents an opportunity for information providers who can integrate their content with standard tools used by their customers; it also presents the threat of marginalizing those information providers whose content is not easy to use with customers' tools. If customers must choose between content that's easy to access from their favorite tools and content that must be accessed separately, they will choose the former, sometimes even when the integrated content is merely "good enough."

Scenario

This final scenario is different from the preceding ones in two ways. First, whereas the previous scenarios involved production tasks internal to an information provider, this one involves distribution of content to customers. Second, the primary business benefit is in

[8] Several journal publishers participating in CrossRef have implemented this; see www.crossref.org.

increased revenue and market share rather than in production efficiencies; therefore, the four requirements used in the preceding scenarios are not very relevant.

This scenario involves a financial information publisher.

EquityLine is an independent producer of equity analysis and research. EquityLine produces analysis reports on 5000 companies and industry sectors on a quarterly basis. EquityLine's production operation is divided into autonomous groups by industry sector, and industry sector groups get redefined on occasion. Most of its products are hardcopy reports that can be made into PDF files.

Allegiant Investments is a large money management company that contains (in addition to a retail brokerage, mutual funds, and other lines of business) the business unit Allegiant Benefit Administration Co. (ABAC), which administers retirement plans for employers. A market leader in benefits administration, ABAC manages 401(k) investment plans for 20% of the market, or 83,200 employers with 8.16 million employees who participate in 401(k) plans[9]. Allegiant's plan management services include a Web site that it makes available on customers' intranets, called AllegiaNet, in which customers' employees can check on their 401(k) plans and change their investment allocations. Employers representing half of the ABAC's 401(k) plan participants, or 4.08 million employees, provide AllegiaNet access on their intranets.

Assume that the Securities and Exchange Commission has agreed to allow 401(k) plans to contain individual equities in addition to mutual funds. Now, in addition to information on mutual funds (of the sort provided by Morningstar and others), 401(k) participants will want information on individual stocks. This leads to a new marketing opportunity for EquityLine, which wishes to integrate its research reports with ABAC so that they can be sold to ABAC's customers' employees while they are doing stock research on the AllegiaNet site. It needs to do this in a way that creates minimum development effort for ABAC, in order to make ABAC's partnership with EquityLine more attractive than partnership with EquityLine's competitors. EquityLine will sell company reports individually for $10 each and pay ABAC a commission of 25% or $2.50 each.

Business Benefits

EquityLine can establish a DOI-enabled ECI architecture like the one described in this white paper. They can create DOIs in the global DOI Directory that point to the latest research document for each company that they cover. These DOIs can be in a standard format that ABAC's Web developers can follow without having to consult EquityLine for each company; for example, each DOI can contain the company's ticker symbol.

With this scheme, ABAC can produce web pages for AllegiaNet that embed DOIs for each company, which point to EquityLine documents. They can be guaranteed that when a user clicks on the appropriate button on AllegiaNet, he or she will immediately get an offer to buy the latest version of the relevant report, no matter in which of EquityLine's systems it resides or if EquityLine reorganizes its internal content repositories or production processes.

Without DOI-based asset management, EquityLine would have to create a separate repository of documents that AllegiaNet users would retrieve. The alternative would be to publish URLs to files on individual industry sector groups' servers. However, because industry sector groups are redefined from time to time, and companies are occasionally

[9] 2002 estimates; source: SpectremGroup.

reassigned to different industry sectors, those URLs would become broken; this is not a fixable problem, since those URLs are embedded into the AllegiaNet HTML code. Using DOIs gives EquityLine the best of both worlds: no need to copy files to a central repository, and links to files on individual workgroups' servers that will not break.

The DOI-based approach is more efficient than the central file server approach, because it is more efficient to create or change DOIs than it is to move files – especially 5000 of them. Both approaches could be done automatically by batch processes, so the savings would be in computer processing rather than labor. A more interesting cost benefit is that with the DOI-based approach, there is no need to create the separate repository for ABAC access, which might save a one-time cost of $100,000 in hardware, software, and system development.

However, the much more important benefit is that of revenue from increased sales of EquityLine reports, which is shown in Table 6. It is reasonable to assume that 2% of employees whose 401(k) plans are managed by ABAC (equivalent to typical direct-marketing response rates) will purchase one or more EquityLine reports during the first year of availability on AllegiaNet. Assume further that the average purchase is 2 reports. That makes a total of about 47,000 reports purchased, representing **over $1.2 Million in incremental revenue for the financial information publisher** after sales commissions. Moreover, those customers would be more likely to keep buying EquityLine reports than to switch to a competitive research provider. **ABAC also profits from this business to the tune of over $400,000 in commissions**.

Total 401(k) plans in U.S.	40,800,000[10]
ABAC market share	20%
Percent with AllegiaNet access	50%
Total employees with AllegiaNet access	4,080,000
Price per report	$10
Take rate	2%
Reports per employee per year	2
Total reports purchased	163,200
Total revenue	$1,632,000
Sales commission	25%
Net revenue to EquityLine	**$1,224,000**
Net commission to ABAC	**$408,000**

Table 6: Business benefits in content/tool integration scenario.

Additional Scenario Applicability

Here are some other situations where the architecture presented here could lead to increased revenue and market share through integration with customer tools:

- For an **electronic parts manufacturer** or **publisher of electronic parts information**, integrating illustrated parts catalogs into CAD tools used by circuit designers.

- For an **educational publisher**, offering educational content on demand through a third-party Internet-based distance learning service.

[10] 2002 estimates, source: SpectremGroup.

- For an **aircraft parts manufacturer**, integrating part specifications into aircraft makers' online illustrated parts catalogs for maintenance technicians.

Conclusion

We have shown how many of the benefits of digital asset management for large information companies can be achieved more easily and at less cost through a new approach and architecture called Enterprise Content Integration, augmented with the use of Digital Object Identifiers (DOIs). We defined the ECI architecture, showed how it differs from the more traditional DAM approach, and showed how it can be implemented with one-quarter to one-half of the financial and personnel resources necessary for an enterprise DAM implementation.

Through several example business case scenarios, we have demonstrated how DOI-enabled ECI leads to these types of business benefits:

- Cost avoidance in cross-brand product development.
- Scalable product development through content repurposing.
- Cost reduction on existing production processes.
- Increasing market share and revenue through customer tool integration.

In sum, we have shown how Enterprise Content Integration can be a driver of scalable product development, faster time to market, more efficient production, and easier product and service customization for information providers.

About the Author

Bill Rosenblatt, president of GiantSteps Media Technology Strategies, has 20 years of experience in technology architecture, business development, and marketing; publishing; new media; and online education. He has been a business development executive at a leading technology vendor, an IT executive at major publishing companies, and chief technology officer of an e-learning startup. He has expertise in digital media technologies such as content management, digital rights management, streaming media, and publishing systems. Bill is the author of several books, including *Digital Rights Management: Business and Technology* (John Wiley & Sons, 2001), and he is the publisher of the newsletter DRM Watch (www.drmwatch.com).

Thanks to **Lee Greenhouse** of Greenhouse Associates (www.greenhousegrows.com) for his contributions.

About GiantSteps Media Technology Strategies

GiantSteps Media Technology Strategies is a management consultancy that helps its clients achieve growth through business strategy and technology architecture expertise. Clients have included book and periodical publishing companies, news, entertainment, and professional information providers, and digital media technology vendors ranging from early-stage startups to Global 500 firms.

Contact:
phone: +1 (212) 956 1045
email: info@giantstepsmts.com
Web: www.giantstepsmts.com

White paper commissioned by Content Directions, Inc.

About Content Directions, Inc.

Content Directions, Inc. is a Registration Agency and Internet services firm dedicated exclusively to helping companies implement and profit from the Digital Object Identifier (DOI). This includes both registering DOIs for a company's digital content and consulting on how the DOI can increase revenues and cut costs, both internally and throughout the supply chain.

Contact:
phone: +1 (212) 792-1847
email: info@contentdirections.com
Web: www.contentdirections.com

Additional copies of this white paper, and related information, may be obtained by clicking on the white paper's DOI: 10.1220/whitepaper5.