

Enterprise Content Management

with Microsoft
SharePoint

Enterprise Content Management with Microsoft SharePoint

Christopher D. Riley
Shadrach White

Copyright © 2013 by Shadrach White and Chris Riley

All rights reserved. No part of the contents of this book may be reproduced or transmitted in any form or by any means without the written permission of the publisher.

ISBN: 978-0-7356-7782-1

1 2 3 4 5 6 7 8 9 LSI 8 7 6 5 4 3

Printed and bound in the United States of America.

Microsoft Press books are available through booksellers and distributors worldwide. If you need support related to this book, email Microsoft Press Book Support at mspinput@microsoft.com. Please tell us what you think of this book at <http://www.microsoft.com/learning/booksurvey>.

Microsoft and the trademarks listed at <http://www.microsoft.com/about/legal/en/us/IntellectualProperty/Trademarks/EN-US.aspx> are trademarks of the Microsoft group of companies. All other marks are property of their respective owners.

The example companies, organizations, products, domain names, email addresses, logos, people, places, and events depicted herein are fictitious. No association with any real company, organization, product, domain name, email address, logo, person, place, or event is intended or should be inferred.

This book expresses the author's views and opinions. The information contained in this book is provided without any express, statutory, or implied warranties. Neither the authors, Microsoft Corporation, nor its resellers, or distributors will be held liable for any damages caused or alleged to be caused either directly or indirectly by this book.

Acquisitions and Developmental Editor: Kenyon Brown

Production Editor: Christopher Hearse

Editorial Production: nSight, Inc.

Technical Reviewer: Jeff Shuey

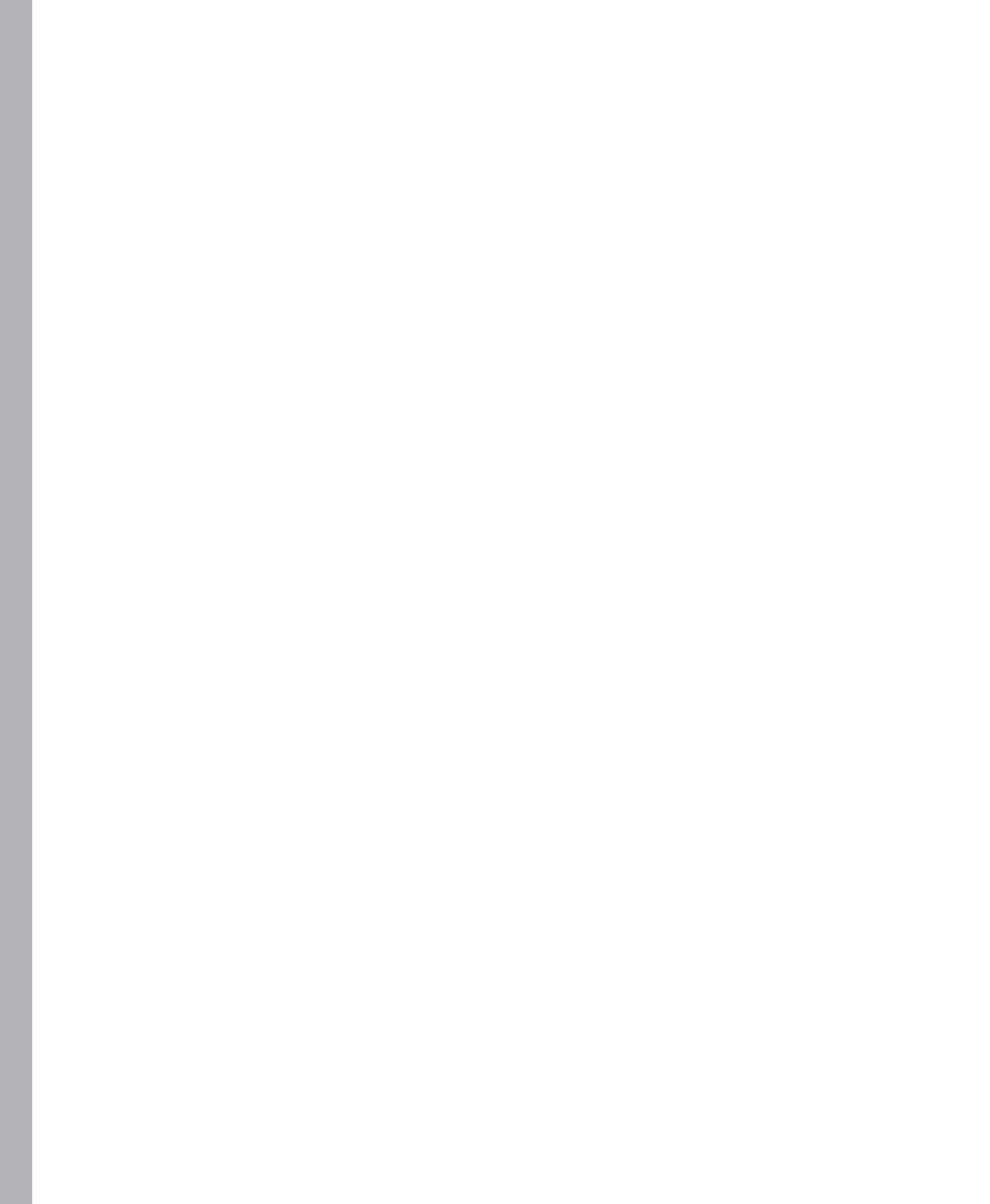
Cover Design: Twist Creative • Seattle and Joel Panchot

Cover Composition: Karen Montgomery

Illustrator: Rebecca Demarest

Contents at a glance

	<i>Introduction</i>	<i>xi</i>
CHAPTER 1	ECM defined	1
CHAPTER 2	ECM stack: content in	35
CHAPTER 3	ECM stack: content control	69
CHAPTER 4	Cases in point	91
CHAPTER 5	Building an ECM team	133
CHAPTER 6	User adoption	149
CHAPTER 7	ECM planning guide	163
CHAPTER 8	Records management	191
CHAPTER 9	eDiscovery	221
CHAPTER 10	Extending SharePoint 2013 ECM solutions	243
CHAPTER 11	Tools and final thoughts	259
	<i>Index</i>	<i>265</i>
	<i>About the authors</i>	<i>277</i>



Contents

<i>Introduction</i>	<i>xi</i>
Chapter 1 ECM defined	1
What is Enterprise Content Management?	1
The ECM stack	5
Capture	7
File upload	8
Microsoft Office documents	9
Native SharePoint documents	10
Electronic form capture	10
Document scanning	11
Content streams	11
Store	12
Information Architecture	13
Versioning	16
Transformation	16
Manage	17
Records management	17
Security and access	18
Policy	19
Change control	20
Deliver	20
Search	21
Editing and viewing	21
Publishing	22

What do you think of this book? We want to hear from you!

Microsoft is interested in hearing your feedback so we can continually improve our books and learning resources for you. To participate in a brief online survey, please visit:

microsoft.com/learning/booksurvey

Process	22
Workflow	23
Business Process Management (BPM)	23
Business Intelligence and BigData	23
eDiscovery	25
Preserve	25
Reformat	26
Compression	26
Why use ECM?	27
Proactive driver	28
Reactive driver	29
How can you use information to make better decisions?	29
Return on investment	31
Who does ECM target?	31
Building expectations	32
Next steps	33

Chapter 2 ECM stack: content in 35

Building a solid foundation	35
Capture	36
File upload	36
Microsoft Office	39
Native SharePoint documents	41
Electronic forms	43
Document scanning	43
Store	47
Physical Storage	47
Logical storage/Information Architecture	49
Web applications and site collections	52
Libraries	53
Metadata	55
Document ID	57
Taxonomy and folksonomy	58

Process	61
Content routing	63
Disposition workflow	63
Three-state workflow	64
Conditional formatting	65
Workflow	66
Next steps.	68

Chapter 3 ECM stack: content control 69

Management of content	69
Change opposition or support	70
Who manages content?	71
Security	73
Repository	73
Document	77
Document management	79
Delivery of content	81
Consistency	82
Browsing and navigation	83
Site contents	86
Search	86
Viewing	87
Preservation	88
Next steps.	90

Chapter 4 Cases in point 91

Deployment assumptions	91
Managed metadata—taxonomy	96
Content types	109
Shared Information Architecture	117
Small scale	125
Large scale	129
Next steps.	131

Chapter 5	Building an ECM team	133
	Don't go it alone	133
	Time and conflict	135
	Team selection	137
	ECM team roles and responsibilities	139
	Team culture	140
	Team communication	141
	Project management	141
	Subject matter expert	142
	Technical team	143
	Quality control	144
	Pre-mortem	145
	Be a practitioner as well as an implementer	146
	Next steps	147
Chapter 6	User adoption	149
	Least common denominator	149
	Preparing the organization	150
	Encourage behavior	152
	The super user	154
	The community	154
	The experts	155
	The change manager	156
	Branding	156
	Bad for adoption	158
	Enforcing the plan	160
	Next steps	161
Chapter 7	ECM planning guide	163
	Documentation	163
	Configuration blueprint	164
	Source of truth	165
	Information Architecture	166

Site and library architecture	166
Content types	172
Taxonomy	177
Content governance	180
Next steps	189
Chapter 8 Records management	191
Principles and life cycle	191
Business drivers	193
Retention schedule	195
Records management features in SharePoint	198
Records center vs. in-place records management	215
Records management processes in SharePoint	217
Next steps	219
Chapter 9 eDiscovery	221
Holds	221
Isolating content	223
Litigation support	223
eDiscovery processes	225
Office 365 consideration	227
Implementing eDiscovery in SharePoint	228
Exporting content	237
Notification	240
Next steps	242
Chapter 10 Extending SharePoint 2013 ECM solutions	243
Office 365	243
Data security	246
Bandwidth and accessibility	247
Third-party services and tools	248
Backup and recovery	248
Business intelligence	249

Business process management	249
Content enrichment	250
Remote BLOB storage	250
Governance and security	251
Integration with LOB	251
Records management	252
Document imaging	253
Social	253
General considerations	254
Systems integrators	256
Next steps	258

Chapter 11 Tools and final thoughts 259

Tools	259
CloudShare	260
SharePoint community	260
Conclusion	262
<i>Index</i>	265

What do you think of this book? We want to hear from you!
Microsoft is interested in hearing your feedback so we can continually improve our books and learning resources for you. To participate in a brief online survey, please visit:

microsoft.com/learning/booksurvey

Introduction

In this book, you will learn why Enterprise Content Management (ECM) is important and how it can be implemented utilizing the SharePoint platform. When you have completed reading the book, you will have the comfort level to know how to implement ECM inside of SharePoint and to understand why you are doing it. This will also help you bridge the gaps in communication between technology and business needs that exist in most organizations.

As you read the book, you will find that there is more emphasis placed on ECM principles than there is on SharePoint 2013. This is intentional, because as you will soon learn, trying to achieve ECM by simply turning on and configuring SharePoint features is completely the wrong approach. Until you fully understand ECM, you don't stand a chance of making SharePoint's features useful to deploy an ECM solution.

Our research, and more importantly, our practical experience have shown that the difference between successful ECM projects and disasters always points to a people or planning problem. Very often, business users are demanding an ECM solution and expect IT departments to implement it. Without a clear understanding of why they are implementing an ECM solution, IT takes the approach of setting up a SharePoint farm, enabling most, if not all, available features, and leaving the rest to the user. Or, it's taken one step further, and they attempt to configure content management functionality such as versioning and managed metadata, without the guidance from the knowledge workers. This is the definition of "knowing enough to be dangerous."

This results in a SharePoint farm configured, standard features enabled, and users given access. Left to their own devices, users will often blame IT for not seeing what they are looking for, and IT will be stuck without a clue about where to make changes.

Let's hope that at this point the organization gets a clue and does a reset on the project. If not, one of two things can happen: first, the users will not use the system and default to what they did before, and in a year the deployment will be chalked up to failure; or second, the system does get adopted but in a rapid way that mirrors all the mistakes that occur from the use of shared network drives. This is called *SharePoint sprawl* and results in site proliferation and/or the *webification* of already poorly managed shared drives.

The catalyst for these problems is a language gap between IT and business users. When an outside ECM consultant asks you to do an inventory of how knowledge

workers operate within your organization, it seems like a mundane and overtly simple question.

However, interestingly enough, when you approach a knowledge worker, requesting a complete description of how they do their job, often the results are a trickle of ideas about the particular tasks a knowledge worker completes in a day, but not really how they are executed. When tasks become routine, their details disappear. This vanishing act results in oversimplification on the part of the knowledge workers of what it takes to manage these tasks. The result is that the IT team has no clearer picture of how content is consumed in the organization than the knowledge workers do, yet everyone's expectations for automation are extremely high.

It is comforting to know that there are very few aspects of ECM that can't be automated, streamlined, and improved with SharePoint 2013. But until an organization knows how the technology fits into their specific modes of operation, it's impossible to go from zero to a successful SharePoint ECM solution.

We are suggesting that planning is the key ingredient to get right early on in ECM projects. But planning usually fails. This is most often because organizations are too nearsighted to do it, because the wrong people are asked to implement ECM, and because the communication between the users and implementers is poor.

Who this book is for

Whether you are a business analyst, IT manager, decision maker, or knowledge worker, by the end of this book you will all be on the same page, speaking the same language, and able to push forward with a proper ECM project. To that end, this book will be part technical and part philosophical. The methodologies of ECM will be followed up with technical examples of how this is accomplished in ECM. This mapping is the exact link that is too often missing in most projects. IT gets the technical SharePoint pieces, and knowledge workers, whether they can state it or not, know the methods of working with the content that SharePoint will be processing and managing.

In addition, we have included a CloudShare demo environment that contains the Information Architecture (IA) discussed extensively in this book.

If you are a business user, you might be tempted to skip the technical bits; please don't. The cursory knowledge of the SharePoint terms and implementation will help you communicate clearly, understand the reasons why SharePoint is being implemented a specific way, and ultimately enable you to get what you want out of the ECM solution.

If you are technical, you might be tempted to skip ahead to the technical pieces and address only those topics that relate to your current project. If so, this is a red flag that your ECM project has already failed and that you are not looking at the solution holistically.

The use cases and labs are meant to help get you started and are in no way designed as a step-by-step guide for building a complete solution. We highly recommend that you read the early chapters first so that you can understand the concepts and best practices for designing and building an ECM solution. Otherwise, you will likely put the technical aspects of SharePoint ahead of the operational business and people aspects of a successful ECM Project.

Ideally, every ECM project stakeholder in your organization will read this book. The project sponsor, power users, IT manager, business analyst, legal counsel, and records management teams will find this as glue to guide the successful execution of the project.

Assumptions about you

The basic assumption of this book is that your organizations needs content management and that you are a stakeholder in getting a content management system deployed. The second primary assumption is that you already own, or will soon own, SharePoint and that your goal is to use SharePoint as the platform for building the required ECM solution.

There are several versions of SharePoint. While most of what is discussed in this book will be relevant for all versions of SharePoint 2010 and SharePoint 2013, both on-premises and in Office 365, the specifics will focus on SharePoint 2013 on-premises standard version or higher. The methodology and process will be relevant whichever version of SharePoint you use. However, if you have SharePoint foundation versions, you will not be able to implement many of the functions required.

Organization of this book

Chapter 1, "ECM defined," begins with a structured definition of ECM. We believe that this will help everyone in the organization use the same nomenclature and concepts. These definitions rely heavily on existing bodies of work that have been established to create a common understanding for ECM professionals to communicate, design, and deploy ECM solutions in a standard and methodical way.

As we move to Chapter 2, “ECM stack: content in,” we cover all the aspects of ECM that pertain to the proper capture and storage of content in SharePoint. Next, in Chapter 3, “ECM stack: content control,” we cover all the processes associated with properly managing that content throughout its life cycle so that it can provide the most usefulness to the organization. These two chapters will cover all the ECM basics that will require the bulk of your attention and the overall time needed for the planning, implementation, and deployment of your ECM project. These are equally important parts of the ECM stack, but they cover very different subject areas and bodies of knowledge. Therefore, first we discuss how to capture, store, and process content, and then we dive into managing, delivering, and preserving content using SharePoint as the core of your ECM solution.

Now that we have established the same definitions and broken your ECM solution into two distinct stacks, we move on to Chapter 4, “Cases in point.” The sections included in this chapter are designed to provide examples that you can point to for best practices as you design your ECM solution. Business processes are never exactly the same for every organization; every Accounts Payable department has unique aspects in how they process work. They might all pay bills, but they do so using a variety of policies, approval methods, financial applications, and Generally Accepted Accounting Principles (GAAP). Note the word “Generally”: these aren’t laws; they are just guidelines. The same goes for other areas that you are planning to address by deploying SharePoint content management technologies and ECM policies and principles.

As we stated earlier, building a successful ECM solution requires a focus on people. In Chapter 5, “Building an ECM team,” we help you identify the right people for your project. To do this, we have provided guidance to help you define specific roles and responsibilities and embrace the need for professional project management. We recommend strongly that you have a PMI certified individual responsible for managing the project from beginning to end. It is also important to identify the technical team members and outline specific subject matter experts who will be responsible for delivering on all the technical requirements outlined in a detailed project plan and architecture design. Finally, we touch on quality control; this will help you understand the importance of unit-testing every deliverable.

Regardless of how well the project is planned, managed, and tested, it will never achieve the desired results without the sections we include in Chapter 6, “User adoption.” How many times have you heard someone say, “I have no idea why I do this; IT said that’s how it works”? Involving users to provide feedback about usability and helping you to understanding how to listen, not just take notes, will be the one thing you point back to and say that’s what made the difference. We will discuss how to incorporate change management practices and create motivation for users. In the end, it might

come down to your ability to lead and change user behavior that will enable the ECM solution to manage content easily.

To help you kick your ECM project into high gear, in Chapter 7, “ECM planning guide,” we include a complete outline of our recommended IA. You will see that we come back to IA many times during the book; this should illustrate why it is so important. We will cover the importance of governance for the project and the use of the platform as it relates not only to SharePoint but also to content that is used throughout your organization.

The principles covered in Chapter 8, “Records management,” and Chapter 9, “eDiscovery,” might not be familiar to everyone. However, they are extremely important to understand and, if possible, to include in your ECM solution. Many organizations find out how important these features and the policies needed to implement them effectively are only after a negative litigation, audit, or regulatory event happens.

We round out the book in Chapter 10, “Extending SharePoint 2013 ECM solutions,” and Chapter 11, “Tools and final thoughts,” by covering some additional areas of interest, including Office 365 and SharePoint user groups. We have also included access to a CloudShare development environment that is a preconfigured SharePoint 2013 environment with the various features and IA discussed in the book, implemented with sample documents.

The purpose of the CloudShare environment is twofold. First, it’s to allow the reader to have hands-on access to a SharePoint farm to actually play with the features discussed in the book. The second reason is to foster the best practice of having a sandboxed SharePoint farm when you do any planning, testing, and implementation work. The reader can take the development environment and use it to expand their own ECM design, demonstrate to various stakeholders how certain features will work, or customize it to be more similar to their specific requirements.

CloudShare offers a time-limited trial, with subscriptions starting at a low monthly cost. We recommend that you, at minimum, utilize the trial so that you have access to the virtual environments at your leisure while you read this book.

Acknowledgments

We would like to thank the following people for helping us. Their support and contributions are greatly appreciated.

Shad

The time spent writing this book was substantial and occurred mostly on nights and weekends. It meant sacrificing some pretty sweet powder days at Crystal Mountain with my son Cade or missing out on playing a game of Pokemon cards with my other son Jaxson. Whenever I would retreat to my office, I could hear the family activities occurring without me. Then a knock at the door with a hot cup of coffee, snack, or request to come eat a meal with the family would remind me how supportive my wife and boys were during this effort. I was rarely alone in my office, usually accompanied by either our dogs Luci and Lily at my feet or Chris Riley, who was working via Skype right alongside me. I would like to thank my family for supporting my decision to co-author this book. I had started a new business just a year before the opportunity to write the book came up, so they had already become used to Dad working seven days a week, including late nights and long weekends. My wife Michele is my best friend, and she keeps me honest about most everything I do. My sons are my heart and soul; I cannot thank them enough for letting Daddy focus and work on this book, mostly without interruption.

Chris Riley is downright brilliant, I hope everyone who reads this book has an opportunity to meet him and spend some time getting to know him. He hasn't come by any of his knowledge or core values easily; he has worked tirelessly to achieve success and failure. The results are a person who has amassed a great body of technical knowledge and personal character. He asked me to be his co-author for this book, and I will always be grateful for the opportunity.

I would like to thank my parents Margaret and James White for teaching me that honesty and hard work are the core elements to success. My parents worked hard every day to create a stable and loving home for my sister Serena Kraft and me. My sister is a successful entrepreneur in Alaska, and we both acknowledge that it was core values more than anything that helped us achieve success in life.

Last but certainly not least, I would like to thank the following people who helped me in so many ways both during the process of writing the book and helping to support efforts that allowed me to write it. In addition, I have included individuals who I have worked with during my career that have had a lasting impact and helped shape my opinions about proper design, execution, and teamwork. These people are ECM professionals in their own right, and I am lucky to have worked with them: Kenyon Brown, Christopher Hearse, Jeff Shuey, Cullen Hower, Dennis Brooke, Mathias Eichler, Jeff Doyle, Bob Stellick, Robert Latham, John Dougherty, Al Senzamici, Phong Hoang, Ryan Keller, and Richard Norrell.

– Shadrach White

Chris

I first have to acknowledge the ECM gene for this book. Without the insatiable drive I have to organize vast amounts of information in a way that makes me consume information better and faster, I would never have jumped into ECM because—let's face it—it's a bit boring. Next I have to thank Shad. I've known Shad for many years now; we were introduced as industry peers but quickly became friends. We connect on the fact that the world, despite most Independent Software Vendor (ISV) assertions, is completely behind in technology adoption. He realized, as I have, that companies, although the technology is available, are just not adopting solutions in an effective way. We share the belief that there is no need to further complicate things. Instead, let's make people successful. We also share a drive to help. We know what is possible. Not only because we have implemented it, but because we live it. Shad wants everyone to know this possibility.

One of the reasons I was able to meet Shad was due to a rock-hard industry that is fighting the daily battle of steering organizations correctly. That organization is AIIM. AIIM has been my surrogate industry parent. They have been the source of vast amounts of my applied knowledge in ECM as well as strong relationships that have catapulted my experience and career.

Another big chunk of my drive came from people I've never met but wish I had. These are the figureheads: James Gleick, Richard Feynman, Guy Kawasaki, Ray Kurzweil, and quite substantially, Steve Jobs. While they have not all been directly involved in ECM, they have contributed to my "style," which is critical in my approach to all technical projects.

And finally, I have to thank my wife, Lauren. She sees how I obsessively organize files on my computer and allows me to speak about what to her might seem like gibberish—about the importance of information, avoiding information glut, and the desire to push the world forward into using that great technology that already exists. She is the records manager of my brain and forces me to do what I need to keep my mind from getting cluttered.

– Christopher Riley

Support & feedback

The following sections provide information on errata, book support, feedback, and contact information.

Errata

We've made every effort to ensure the accuracy of this book and its companion content. Any errors that have been reported since this book was published are listed at:

<http://aka.ms/S2013ECM/errata>

If you find an error that is not already listed, you can report it to us through the same page.

If you need additional support, email Microsoft Press Book Support at *mspinput@microsoft.com*.

Please note that product support for Microsoft software is not offered through the addresses above.

We want to hear from you

At Microsoft Press, your satisfaction is our top priority, and your feedback our most valuable asset. Please tell us what you think of this book at:

<http://aka.ms/tellpress>

The survey is short, and we read every one of your comments and ideas. Thanks in advance for your input!

Stay in touch

Let's keep the conversation going! We're on Twitter: *<http://twitter.com/MicrosoftPress>*.

ECM defined

In this chapter, we will lay out the fundamental building blocks of ECM. This will help you understand at a high level what you will need to include as part of a comprehensive strategy for managing content in SharePoint. Almost anyone can create a server farm and then create a site collection and start adding content to a document library. Typically and unfortunately, that's how most SharePoint projects begin to fail. What you will gain from the following sections is a much broader scope of ECM to help you prevent this from happening.

What is Enterprise Content Management?

Enterprise Content Management (ECM) is not a technology, and SharePoint is not an out-of-the-box ECM solution. These two statements contradict what some have thought to be fact and the marketing slicks and PowerPoint presentations so many of us have seen. You might be asking yourself, "What gives, we bought into SharePoint so that we could have an ECM Solution?"

ECM is a set of practices, processes, and methodology that make the technology morph into the most effective way to store, secure, and consume content. SharePoint is a platform, a grab bag of features, and technology that can be molded into a fantastic ECM solution.

ECM is also not moving what has been done in shared drives to a web-based modern platform. Consider how shared drives evolve, how much file duplication and organizational chaos is typically found in shared drives accessed by a community of users. It is extremely hard to find anything resembling structure. Ask yourself the following questions:

- Does your company have a documented File Naming Convention policy?
- Do the Shared Drives you work with follow a common Directory naming structure?
- Can you easily navigate another department or workgroups Directory structure and find a file?

Most will answer no to these questions, and when asked in a group setting these questions usually provoke eye rolling, scoffs, and jeers. The organization and planning of Shared Drives happens in real-time during content creation. This is often impounded by busy workdays, staff changes, and different personal organizational strategies. Ultimately, this approach is not considered a solution for proper content management.

One must acknowledge that unplanned ECM is the process of taking this same structure and lack of planning and modernizing it. Simply doing a one-for-one transposition of the Share Drive into SharePoint results in an even more rapid use of a very bad system.

ECM is the opposite of this; it requires up-front planning and practical application of information architecture. It assumes a standardization of proven methods for capturing, naming, and storing content. In the past, this required a lot of effort and thought on the part of the user. But more and more, technology is taking over for the unfriendly pieces.

Q & A

Q: How many ways can you organize unstructured information?

A: Four—chronologically, alphabetically, numerically, and geographically. These four approaches are compounded by the many ways people use, format, and duplicate them.

In most cases, the information stored in an ECM system is used to support the primary corporate data found in traditional line-of-business (LOB) systems, such as contract management, human resources management (HRMS), customer relationship management (CRM), or enterprise resource planning (ERP) systems. You can commonly find “ECM solutions” with content not attached to LOB applications. Usually, this is a sign of bad ECM, but there can be exceptions. An increasing number of ECM systems are now being used to drive complex LOB activities because people often end up looking for an important document or record to answer a question or to complete a business transaction.

It often takes organizations several failed attempts to realize that without proper ECM no platform can succeed as a content management system. And without a platform with the right set of features, and end-user ease-of-use ECM can't succeed. The core feature sets of an ECM solution are as follows:

- Storage of documents
- Document viewing
- Document editing
- Document security
- Metadata model
- Versioning of documents
- Check-in/check-out of documents

SharePoint today lends itself very nicely as an ECM solution, but that has not always been the case. SharePoint has evolved to provide all the necessary components of an ECM solution, capable of facilitating the complete life cycle of content management. Prior to SharePoint 2010, you could not implement a structured system that would be considered a suitable tool for large scale ECM.

Starting with SharePoint 2007 MOSS, core ECM functionality outlined above was brought into the platform. While SharePoint 2007 could be molded into a reasonable ECM solution for small business, it missed some core functionality in the following areas:

- eDiscovery
- Electronic forms
- Records declarations
- Consistent metadata models
- Business process management
- User and service level audit logs
- Compound documents

With SharePoint 2010, this functionality was introduced, and SharePoint was finally a full-blown ECM solution capable of delivering at almost every level. SharePoint 2010 was still lacking in a few areas, such as more robust business process management, document auditing, information architecture, and capture for document imaging. In these cases, third-party solutions appeared out of the Microsoft Partner community and took the 90 percent to 100 percent. We will highlight some of possibilities for widely used third-party solutions that you will want to evaluate, depending on your specific project requirements.



Note In this book we will focus on ECM related features. While SharePoint can be used for many information management use cases we will not be including or highlighting areas like Business Intelligence, Web Content Management, or Social Collaboration.

With the release of SharePoint 2013, the following additional features are included:

- Disposition of sites
- Federated eDiscovery

These are two desperately needed ECM tools we will discuss later. SharePoint 2013 also brings some new challenges. The new user interface is better for user adoption but trickier from a governance standpoint. For example, the new “Share” feature, when not properly governed, will cause uncontrolled propagation of site security. Remember SharePoint sprawl, mentioned in the Introduction? The rapid adoption of the Share feature without governance can aggravate sprawl. We will discuss this in greater detail later.

Now that we know that SharePoint 2010 and, more importantly, SharePoint 2013 offer the complete feature set of ECM and then some, we can discuss how organizations can leverage SharePoint as a complete ECM Solution.

A very common misperception about ECM in SharePoint exists with organizations that believe that, by using team sites and collaboration features or by uploading content from shared drives, they are

performing ECM. This simply is not the case. These are manifestations of the same bad content habits in a more modern way. This points to one of the primary reasons that ECM projects in SharePoint fail.

Top five reasons ECM in SharePoint fails

- Not understanding that ECM is a methodology, and SharePoint is a platform
- Jumping into adoption with no project planning
- Trying to build an organization-wide ECM solution versus a quick win
- Not balancing technology, business, and user requirements
- Stakeholders speaking different languages of the business, such as technical versus operational

Speaking the same language is a common issue and is the trickiest to identify and mind-bogglingly hard to address. Very often, knowledge workers who request the need for technology for more efficient document processes do not possess the knowledge to ask the right questions of the IT professionals who are responsible for delivering a solution.

To mitigate this, many organizations are utilizing the role of business analyst to bridge the gap between user requirements, IT implementation, and operational benefits to the organization. This role can be justified, but return on investment (ROI) can be hard to measure in terms of hard dollar costs versus savings. The ROI for a position such as business analyst is best determined by looking at the soft costs and opportunity costs of getting it right the first time.

Business Analyst or Information Architect

In most cases, the role of the person who manages the communication between IT and the rest of the business is called a Business Analyst. Traditionally this position focused solely on analyzing the effectiveness of a specific business process and the systems that are used to facilitate that process. There is a difference between analyzing and documenting the outcomes of a business process and providing the upfront planning and guidance for the ECM solution you are designing. If you have seen the Business Analyst role expand to include these types of activities we recommend that you intentionally separate them and assign a new role of Information Architect.

Too often, organizations look only at raw capital expense put into deploying ECM, which is the cost of employee time, professional services, hardware costs, and software licensing. They don't factor in what it will take in terms of change management, user adoption, and the implementation of new policies and amendments. In some cases, the deployment of a SharePoint ECM solution can require changes to job function or description for certain aspects of department operations for individual roles. These are just some of the areas where a business analyst adds tremendous value.

A good Information Architect

- Works with all parties and speaks both IT and Business Operations language
- Helps mitigate misunderstandings between departments
- Navigates a variety of personalities and political turf issues
- Understands the principles of project management

As you can see, the communication gap is not a trivial issue. It requires admission from both the end-users and IT that they are not communicating and perhaps don't know how the organization operates today. And it even sometimes requires an individual just to manage the requirements gathering and communication.

We have defined at a high level what ECM is, why SharePoint is an ECM solution, and the problems organizations face when taking the SharePoint platform and configuring it to deliver a complete ECM solution. Before we go further let's dig into the details of the ECM components.

In this section of the book, we will look at the definitions of each component of the ECM stack, and in later chapters, we will examine the full details about each component's use and implementation in SharePoint.

The ECM stack

We started this chapter telling you that ECM is not a technology, but rather a methodology. This implies that to execute on those methodologies there must be some collection of technology that can do it. The ECM stack is where we stop talking about ECM as a single comprehensive object and start breaking it into its pieces. It's in the pieces that we align the methods to technology.

The ECM Stack includes all the components of an ECM solution. It is sometimes referred to as the document life cycle. The document life cycle is defined as all possible stages that a document can encounter in its life. The primary stages and components are shown in Figure 1-1.

First comes the Capture stage, which puts the content into the system. The next stage, Store, focuses on storing content, which is predominantly achieved automatically with technology. This is the proper committing of the captured content to the system and is both the logical and physical storage of the content and associated metadata. These two stages feed the ECM system with appropriate content. These stages are considered by most knowledge workers to be the least beneficial to their operation, but without them, knowledge workers have nothing to execute on. It's in the Manage, Deliver, and Process stages where the real value of the content becomes obvious.

We have found that in user adoption it's a lot easier to convince users to engage in the management, delivery, and processing of content than in the Capture and Store stages. However, without

good capture and storage of content, effective management, delivery, and processing can't happen. Later we will discuss strategies for user adoption.

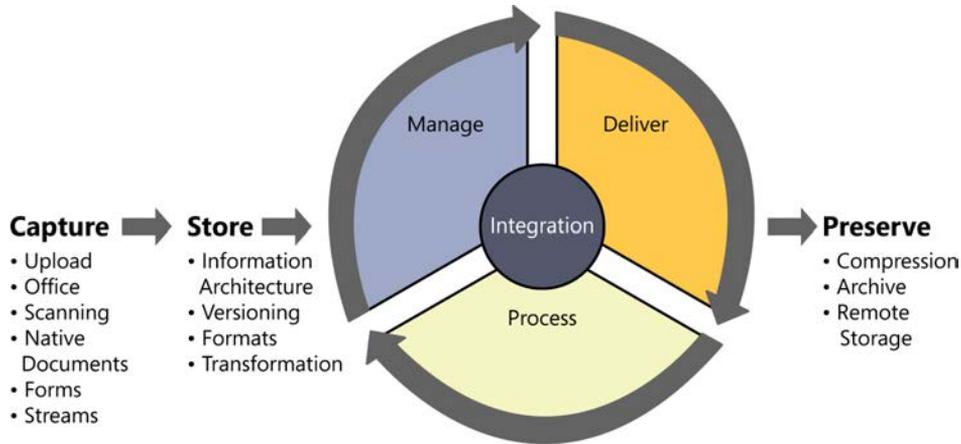


FIGURE 1-1 ECM stages.

The goal is to spend most planning effort on the quality of the Capture, Store, and Manage stages so that very little effort is required for the Delivery, Process, and Preserve stages. Arguably, the lion shares of the planning for ECM are in the Capture, Store, and Manage stages, while the Delivery, Process, and Preserve stages should be nearly effortless, with good content being put into SharePoint.

We have laid out the document life cycle in Figure 1-2. The diagram is laid out to signify two additional aspects. First, on an X-axis, we have outlined features in each of the document life cycle stages and listed them in order of the most commonly used features in SharePoint to the least commonly used. Second, on a Y-axis, we have ordered each document life cycle according to the amount of planning and configuration effort that should be placed on each stage.

It's important to think of these stages, beginning with Capture and ending with Preserve, as having "downstream" effects. What happens, the good and bad, while capturing content directly impacts the quality in the Deliver and Preserve stages. If content is captured and identified inconsistently or if a minimalist approach is taken to applying metadata, the ability to find the content will be greatly diminished.

While the greatest benefits to the majority of users is found by searching and finding content in the Deliver stage, the understanding and adherence to the "downstream" concept will ensure that you take capture seriously. The proper attention to detail in each stage will ensure successful results for knowledge workers when they are searching, processing, or managing content. We will define each of these document life cycle stages in this chapter.

Ultimately, these components align to a feature or combination of features in SharePoint. These technologies/features/aspects combine to make the overall ECM solution. Some deployments can have additional components, and at the later stages of the ECM stack, for example, some organizations can omit preservation and eDiscovery. This inclusion of these later stages is usually driven by

specific business needs or regulatory compliance requirements. We find that, for the most part, all organizations will or should deploy some aspect of all these stages.

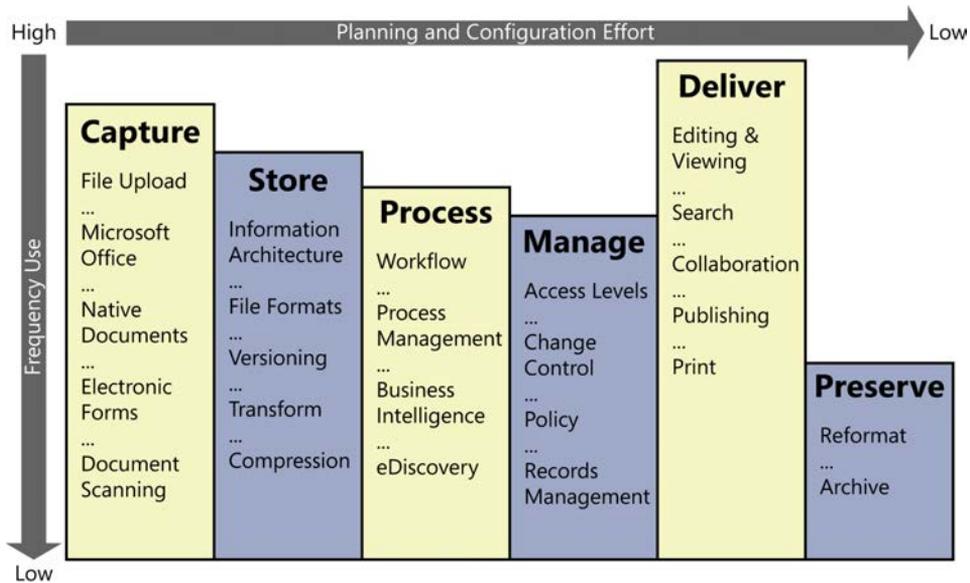


FIGURE 1-2 ECM life cycle.

As we said earlier, each SharePoint component aligns to a feature or combination of features. In later sections of this book, we will explain the alignment and provide examples. Now let's look at each stage in more detail.

Capture

Think of Capture as grabbing information: grabbing it from existing content stored elsewhere, grabbing it from the minds of its curators, or grabbing it from one format and transforming it into another. We capture content constantly without knowing it. Even the process of writing this book is a form of "born-digital" capture.

Capture is the process of preparing, collecting, and indexing content before being stored in an ECM Platform. Capture into SharePoint can happen in the following six distinct ways, ordered by most common to least common:

1. File upload
2. Microsoft documents
3. Document capture
4. Natively created SharePoint documents

5. Electronic forms
6. Data streams

Many tools, forms, formats

During the authoring of this book, we used online document collaboration, email, Microsoft Office, and Skype. These can all be considered forms of Capture.

Without the Capture stage, content does not get into the system. Often this stage is transparent to users because they do it so frequently that they do not realize they are doing it.

This can cause problems in the planning stages of ECM deployment. As depicted in Figure 1-2, Capture is the second-most used stage in ECM, and requires the most planning. We all recognize that this is step one, so failure here results in failure in any downstream stages.

As one of the primary goals for this book, we want to make sure that your ECM team is speaking the same language, so let's look at the specific types of content capture so that we are all on the same page.

From the field

One time I was sitting with a client telling them how great document imaging would be for him. He peered at me from behind a stack of papers on his desk, with filing cabinets at both sides of him, saying, "This technology is cool, but we don't use paper anymore." - Chris

File upload

File upload is the most common way users contribute content to SharePoint. Here we are mostly referring to ad hoc user-driven file upload, but the mechanism can also be used in bulk either by a user or by an unattended script.

Users are most accustomed to browsing to a local file location, selecting the file, and uploading it to a designated space in SharePoint. It is one of the slowest capture processes, but it usually results in the greatest quality of content and metadata.

Primarily performed as a singular activity, based on immediate needs to share or manage a document collaboratively with other users, the file upload method is rarely used to bulk load content into SharePoint. Although users might attempt to do so, it is not recommended.

Document uploading can happen by both browsing to a local folder or network directory and selecting a document(s) to upload. In addition, users can simply drag files onto the document library web interface.



Note There is a SharePoint user interface limitation of 100 items for bulk operations in the web interface. You can use scripting tools such as Windows PowerShell to perform bulk-loading operations for larger volumes of content.

Organizations can elect to automate the initial load of content into SharePoint, leveraging either Windows PowerShell scripts or migration tools. In later chapters, we will explain both the risks and the benefits of this approach.

One key element of the manual document upload is the process from the users' perspective. Where are they finding the content? My Documents? A Shared drive? This could highlight some issues that you certainly don't want transferred over. What are they uploading? Certain file formats don't belong in ECM, so do you allow your users to upload anything they have access to? What metadata must they complete for a successful upload? We will address these issues in detail in the next chapter.

Microsoft Office documents

Office documents are a born-digital method of capture. These documents have never known a physical existence. This comes with a lot of advantages. This is when a user creates a document in the Microsoft Office suite (PowerPoint, Word, Excel, and so on), which supports the saving of documents directly into an accessible document library.

The advantages of the born-digital method are improved capture accuracy and, most importantly, ease of use. In this mode of operation, SharePoint libraries are "save as" locations for content. Using Office as a client application to contribute content to SharePoint is also a solution that most end-users embrace, because it's similar to how they have always worked with content.

This encourages knowledge workers to adopt habits leading them away from the storage of content in "My Documents" and "Shared Drives," which is one step closer to better capture. However, this approach works best when knowledge workers are located within the corporate firewall that is connected to the farm 24/7. Lapses in connectivity can cause problems with upload, versioning, and, more painfully, user adoption.

Proper user training can help with this, and SharePoint "Workspaces" is another alternative. Workspaces is a client Office application that allows content saved in a specific location to be automatically uploaded to SharePoint. Conversely, it allows the viewing of SharePoint Libraries as a folder on your desktop. This is very similar to "Explorer view" and, for the purposes of this book, is considered the same tool.

While Workspaces and "Explorer view" can be used for project portals and personal file repositories, they should not be used for ECM. From a technical point of view, the functionality in Workspaces actually allows the bypassing of proper ECM functionality such as managed meta data columns. From a strategic point of view, if you offer users Workspaces, they are encouraged to live in a My Documents type structure and not a proper ECM. Give the users Workspaces, and you will never be able to take it away.

While many Office documents are manually created, we recommend coaching users to create documents in Office that are saved directly to SharePoint Libraries or, even better, to create content by using built-in Office Web Applications.

Native SharePoint documents

The second type of born digital document is the native SharePoint document. Native SharePoint documents are those documents created without the use of an external client. They very often are Office documents with the addition of blogs, wikis, and pages. For documents, this is done using a feature in SharePoint called Office Web Applications OWA.



Note Office Web Applications is an add-on in SharePoint 2010 and SharePoint 2013 that allows users to create content in-browser instead of using a client application. It requires Office CALs, and it has some limited functionality compared to client applications. In most cases, it is all a typical knowledge worker needs.

SharePoint now supports the authoring of documents directly in the browser. The majority of the required functionality found in Microsoft Office client applications can be found inside the browser, making content capture of 90 percent of documents very easy to perform directly in SharePoint. Today, this is not the most commonly used form of capture; however, it is the future and the desired method to guide users toward.

As we look to the future of ECM, natively born-digital documents will take over, shortening the time to capture, increasing the accuracy of content capture and its metadata, and being more acceptable to end-users. The challenge of this approach is to make sure that an organization is current in browser and operating system support.

We now move to types of Capture that are more use-case specific and sometimes omitted by organizations.

Electronic form capture

Electronic form submission is one way to get user content into SharePoint. The process enables individuals to complete online forms often referred to as e-forms, and the results of form submission are saved directly to SharePoint. The person completing the form very often will not have access to the form submission content nor necessarily be a named user in SharePoint.

The preferred approach to electronic form submissions in SharePoint is the Microsoft product called InfoPath, because it is a tightly integrated electronic form solution. However, it is also possible to leverage third-party or custom e-form solutions. We will include references to additional resources that can be used for e-form solutions in SharePoint later in this book.



Note Because individuals who submit forms are most often not the consumers of content in this capture scenario, security is a primary concern. Often, it's important that submissions are not accessible to submitters.

One of the most neglected aspects of form capture is form design. Without good electronic form design, the information entered into the form rapidly diminishes in quality. Because electronic forms are one of the best ways to capture structured information and metadata, we encourage organizations to spend considerable time on the usability, presentation, and transformation of their forms.

For purposes of this book, discussion of the specifics of e-forms and InfoPath will be omitted. Rather, forms will be highlighted as a way to get content into SharePoint.

Document scanning

Often, documents are not born digital. They begin as physical entities, but they need to be a part of the ECM system just as critically as born-digital documents. The process of capturing this content is called document imaging or document scanning.

This process occurs in three primary ways: ad hoc, departmental/distributed, or production capture. The process also can include image-only capture or capture with intelligent conversion technology.

The owner of the content at their client workstation performs ad hoc capture with an attached document scanner. Departmental/distributed capture is an extension of this, but shared document imaging stations are usually shared among several knowledge works. Finally, production capture focuses on centralizing all capture operations, controlling the indexing, and performing advanced processing and data collection functions in a predictable manner.

How organizations pick one type of capture over another depends on the document maturity of an organization or their own structure. Later we will go into detail about how the various types of capture can be incorporated into SharePoint. While SharePoint does not have natively built-in scanning functionality, there are several ways to incorporate ad hoc and departmental scanning. Production scanning is often considered as a process outside of SharePoint.

Content streams

Sometimes content is not created by a user but rather by another system. "Integration" is a rather broad term that describes connecting disparate systems. For purposes of ECM, we will define a specific type of integration where only content and metadata are transferred from outside SharePoint into SharePoint and refer to it as *content streams*.

A content stream can come from any electronic source and is defined as the ingestion or publishing of metadata and content via some standard such as Windows PowerShell, REST, XML, and RSS. An example would be the consumption of news feeds inside of a SharePoint page or the publishing of RSS feeds for a list to be consumed by another system.

The functions are broad, covering many different methods for integrating the capture of content in SharePoint. The most important thing to understand is the necessity and governance of such integrations and how it relates to the other forms of capture. For example, surfacing documents that live in another content system impacts the way an organization will plan their ECM environment and train their users.

After content is captured, it has to be properly stored. This moves us to the Store process.

Store

Storage is not just the writing of a document's content to a list or library. It also refers to all aspects of that document, including its security, its history, and its metadata. The following pieces of document storage are listed in the order that they should be implemented and used:

1. Information Architecture
2. Formats
3. Versioning
4. Transformation

Storage is the physical and logical storage of content and associated metadata. The distinction is important. Some might consider storage as equivalent to file shares and databases. The danger in this is that in ECM we add a strong logical component to storage in the form of metadata. It's this metadata that makes downstream ECM processes possible. Without it, you end up neglecting information architecture in favor of a new representation of the old shared drive paradigm or forgetting that the physical file has to be saved somewhere as well. This results both in no planning for scalability and in hitting a wall at some content storage limit.

It's not just Storage

Logical Storage: The metadata model and references that SharePoint uses to find the content—for example, a hyperlink.

Physical Storage: The actual location of the content (technically, the BLOB object)—that is, the combination of server, database, and storage device or server used.

In SharePoint, the physical storage happens in Microsoft SQL databases. These databases are referred to as *content databases*. This is the physical location of the BLOB object that makes up the documents content. Metadata is stored in a separate location and then linked to the content BLOB.

Information Architecture is the logical storage of content. This includes web applications, site collections, sites, list, libraries, and content types. This includes their configuration, number, and relationship to each other.

Arguably, information architecture is one of the most critical components to successful ECM and is where an organization will spend the majority of its time planning.

SharePoint also supports configuration of Remote BLOB (Binary Large Objects) Storage (RBS). We will talk more about this in Chapter 2.

RBS can facilitate a measure of scalability for specific use-cases that involve large individual file sizes and/or high volumes of objects. The configuration of RBS can be targeted to specific SharePoint Web Apps, Site Collections, or individual sites.

An example of this would be large engineering vector files and high-volume document imaging solutions. In most use-cases, RBS is not necessary, and storing the objects in the database has the advantage of providing one source from a backup and disaster recovery perspective.

To continue the theme of making sure that your ECM team is speaking the same language, let's look at the specific aspects of a standard SharePoint information architecture so that we are all on the same page.

Information Architecture

Often, Information Architecture is confused with one of its individual components, such as folders or site collections, rather than looking at it holistically. We find that while many organizations implement Information Architecture organically, very few know what it is. We can visualize all aspects of Information Architecture in Figure 1-3, showing that this is the logical location of content in SharePoint. The figure shows, in a hierarchical order, all aspects that define the logical storage of a document and how its metadata is represented. In later chapters, we will define each of these and their use in detail.

The growing trend is to make the repository portion of Information Architecture as flat as possible and the metadata portion as comprehensive as possible. Later in the book, we will talk about the theoretically ideal Information Architecture and align it with practical implementations of SharePoint. We will come up with guidelines for designing your Information Architecture.

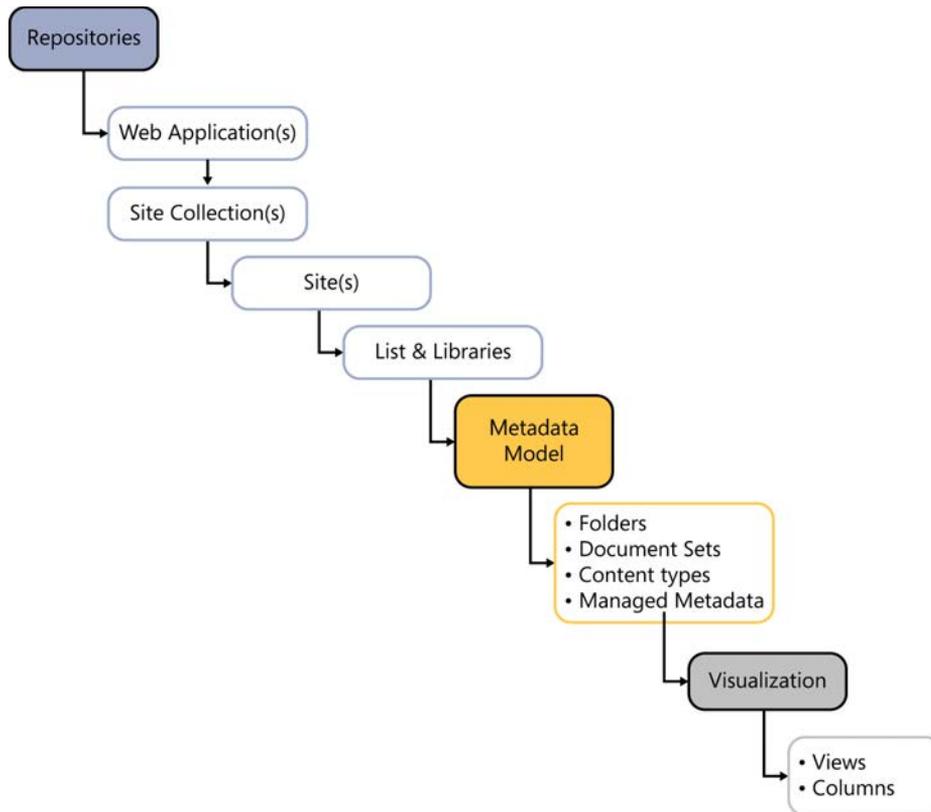


FIGURE 1-3 Information Architecture.



Note What you might not realize is that a folder is just a piece of metadata. To file systems, a folder appears as an extension of the file name. What folders do is offer a single point of view for content, which is too limited because it doesn't provide flexibility in terms of user search. Additionally, in situations that require support for large volumes of content, folders have inherent rendering limitations. By incorporating a flatter information architecture with more metadata, users can slice and dice content according to several variables at the same time, changing that single one-dimensional point of view into a multidimensional, flexible way to browse and search for content.

Q & A

Q: Are folders bad ECM?

A: The trend is to abolish folders as the exclusive method for managing content.

We dare to suggest that Information Architecture is not only the place where you will spend most of your time in planning a good ECM solution, but also a place where you can have fun. For those who strive to be organized, Information Architecture is where it happens.

In contrast, the use of proper metadata in content types instead of folders allows a user to slice and dice content on any number of combinations of information they want. Metadata provides a more flexible means of organizing and reorganizing content on the fly. Surfacing content in visual folder layouts is acceptable, but it must be used in conjunction with metadata and search.

From the field

File formats

One time, I had a client struggling with certain files in SharePoint. I found out that the user was trying to apply ECM logic to .dll and .exe files. I had to stop this practice immediately. - Shad

Every file enters SharePoint as a file format. Most commonly, when we talk about ECM, we are talking about Word documents (.docx), text files (.txt), and compressed portable files (.pdf). A great but often dangerous aspect of SharePoint is that *anything* can be content in SharePoint. As long as there is an electronic way to represent an object, it can live in SharePoint. This includes .dll files, .zip files, and so on. We could even invent a file type (.you) and place it in SharePoint.

Because this is true, the types of files a user can contribute to SharePoint should be controlled. Consider the following:

- Are the files you are storing supported for native SharePoint viewing?
- Do you have the proper management tools (iFilters, Viewers) to support non-native access?

We recommend that ECM planning teams determine approved file formats that are permitted, and the fewer the better. This reduces the overhead of managing the files and opens the doors for more possibilities.

For example, if an organization could ensure that all content contributed to SharePoint consisted of Microsoft Office documents, the organization would then be able to consider the features in SharePoint that allow for the automatic formatting of office documents with barcodes. If the organization can't ensure this, enabling this feature will result in inconsistent content consumption and use.

We will share suggested approaches to file formats and configuration alternatives later in the book. Even in organizations with a variety of specialized content, you can limit the format types to a small subset by using proper conversion and transformation technologies.

Versioning

As documents come in to the right location, in the right format, they often need to be versioned.

Versioning is the process of storing earlier versions of a document with their associated time and date stamps. The earlier versions might be used to revert back to previous version, for comparison across versions and for sharing one version while a more current version is still being edited. This is not to be confused with tracking changes in a Word document; tracking changes is a function that facilitates the creation of versions, but it is not a version control system. In SharePoint, versioning can happen automatically. When a document is saved, it is possible to save a major version or a minor version. The versioning number system is what determines whether it's major or minor. A minor version is everything to the right of the dot or decimal, and the major version is everything on the left side. For example, a document with version 3.5 has a major version of 3 and a minor version of 5.

Because major and minor versioning is usually used only when an organization publishes documents—for example, to an extranet or partner portal—the major version represents the published version while minor versions remain unpublished and continually edited. So while editors are working on version 3.5 and on to 4, the people who have access only to published versions see version 3. When versions are saved, the editor has the ability to add comments to help identify what changed in the version. Some versioning happens upon document upload and is automatic, based on filenames that already exist.

There are two dangers of versioning. The first is blindly enabled versioning, which could result in the overwriting of files without the users knowing. The second and more serious danger of versioning is the impact it has on content databases by adding a separate copy of the document and additional comment metadata. This can cause a site collection to get too large too fast.

One of the huge benefits of the SharePoint Platform is that anything and everything can be content. Besides file size limitations per file, there is no limitation on the type of file you can upload to SharePoint. However, there are best practices for choosing which file formats to support as a policy. The considerations are based on how functional the files are, how they can be viewed or edited, and whether they pose any security risks.

The most common formats used are MS Office documents and PDF files. But it's not uncommon to see media files, CAD files, and other proprietary file formats. Another part of file formats in SharePoint is a mechanism called iFilter. The iFilter is what makes content useful to SharePoint. It exposes the internal content of the document in a way that SharePoint can index and search on it. It also is necessary for third-party products that run within SharePoint to visualize, edit, or otherwise work on the content.

Transformation

As we indicated in the "From the field" sidebar on formatting, transformation, while not always a consideration, often becomes an important consideration when it comes to the types of file formats you allow and the desired formats to have in the ECM system.

Transformation, also referred to as *conversion*, is the process of taking an original file format and converting or transforming it into another. The transformation process often is just a format change, but it can also include other processes, such as optical character recognition, natural language processing, translation, and other types of content manipulation that make the resulting document more useful. SharePoint has some built-in conversion functionality for Office documents and hooks to incorporate other transformation processes.

You now have the documents captured, properly stored in a content database in the right filing location according to your Information Architecture. You have them in the right format with versioning capabilities. Guess what? This is starting to look like a real ECM system. Now let's work on managing the content that has found its way into SharePoint.

Manage

The Manage portion of Enterprise Content Management comprises all aspects around governance of the system. This includes informal and formal policies for users, the requirements for how content is captured, the requirements for how content is stored and secured, what is involved in records management, how and when content is deleted or consumed, and finally, but most ignored, how the system grows.

Governance is defined as bringing together all the elements necessary to facilitate the long-term preservation, accessibility, and disposition of content. This is first accomplished by establishing Organizational Policies and Procedures and often includes authoring and review of a document reviewed and approved by the Legal Department and then formally adopted by the Board of Directors. After this is completed, it can be used as a guideline for implementing a records management plan that includes a formal definition of records, how long they are to be kept, how they are disposed of, who has the authority, and what to do in case of litigation. In conjunction with well-defined Information Architecture, a complete governance plan can help ensure that the user adoption is high and that the extensibility of the ECM solution is straightforward

Governance includes the following elements:

- Records management
- Security and access
- Policies
- Change control

Records management

Records management, like its broader parent, ECM, is a practice and methodology. However, records management requires a far more strict set of principles. Records management includes the following disciplines:

- Records series
- Records declaration
- Retention schedules

A record is a stamp in content, time, location, and metadata for a document. When a document is declared as a record, its content will not change; the metadata, such as last modified data, will not change; and its logical and physical storage location will not change. Very often, there is an additional step that occurs during records declaration, which is a reassignment of security so that individuals who don't have authority cannot access the records.

In the past, SharePoint records management was limited only to record centers. A records center automatically declares all documents as records when they are saved in that site collection designated as a records center. However, now with SharePoint 2013, records declaration can happen in a records center or in any other library automatically, via a workflow, or with manual in-place records management.

Records and non-records alike might be subject to retention schedules. A *retention schedule* is a listing of all document types and their associated life spans. For example, a contract might be deleted five years after the termination date. Retention schedules are required in compliance-driven industries but are not common in smaller business. However, the use of retention schedules is an excellent discipline for any organization and can be a cheat sheet for implementing ECM in SharePoint.

Records management, even when implemented, does not apply to all content. The relevant content is determined by the retention schedule and is usually critical to business operation or contains potential risk/value associated with compliance or litigation. Organizations, even without records managers or records management policies, can choose to learn from the strict organization principles so that they can be better prepared for future compliance restrictions or litigation. The retention schedule also determines which elements of metadata are required or not. This particular practice of deciding what metadata is required or not, while mentioned here in records management, is mandatory for all organizations.



Note Sometimes the enforcement of mandatory metadata is referred to as “management by the red asterisk” because of the symbol placed next to mandatory metadata fields. The plus of this strategy is that you get better metadata when content is added; the minus is that users add less content to avoid it. There is a balance to be achieved.

In your organization, it's not the content you know about that can hurt you; it's the content that you don't know about.

Security and access

Another huge benefit to SharePoint is its ability to manage security at nearly any level. What you learned above about Information Architecture is an absolutely critical element in security considerations. The various site collections, sites, and libraries will all have separate security considerations.

Access levels are among those considerations. Who has access to what? Similar to Information Architecture, the hierarchy of security can spiral out of control. Security and access levels have a direct one-to-one relationship. For example, even though item level security can be achieved in SharePoint, the maintenance and risk of such a policy is high. Also, as with Information Architecture, the trend is to make security envelopes at the top level and flatten and widen the repository where security restrictions are applied. For example, instead of having a site collection for all departments and a site below it per department, organizations are making a site collection per department, with security at the site collection level instead of the site level.

SharePoint has the ability to show to users only content they have access to and also block users out of a repository completely, if necessary. This process is called security trimming and is a tremendous tool that you can use to help protect content and support compliance initiatives.

Security is one way to enforce governance, but it does not solve everything.



Note *Security trimming* is the process of showing only the content that the currently logged in user has access to. For example, if you do a search for documents with the term “security” in it, SharePoint will first find all, for example, 100 documents that have the word “security” in it. But because the user logged in has security access to only 20 of those documents, SharePoint will show only the 20 documents the user has access to. This is true in libraries as well.

Policy

There are certain elements to governance that can be implemented with technology, some can be implemented only with rules, and others could be accomplished with either technology or written rule, so a decision must be made. For example, “management by the red asterisks” is the process of making certain metadata fields required for document upload. But this can encourage users not to use the system when over used, so it might be better for an organization to declare to its knowledge workers which metadata is required and which is not.

A policy is a written rule on how to use the system. What is written in a policy is one thing, but how the policy is implemented is far more important. Most policies will also need to have specific procedures and, in some cases, user training to make sure policies are understood. Policy is usually set at an executive, board, or steering committee level. The department management and knowledge workers familiar with the business process generally develop procedures. A policy that is simply published without documented procedures, training, and, ultimately, enforcement is rarely effective. Therefore, part of a policy system must include the enforcement system. What we mean when we say this is that if you are going to make the policy, you have to take action when it’s broken or the policy’s value is nullified.

Organizations might want to believe they can do without policy and utilize technology enforcement or rely on the better judgment of their users. This will ultimately lead to great adoption in the wrong way or to no user adoption at all. Not considering the policy system can result in ECM failure.

Change control

When an organization decides to take on a project like SharePoint, the requirements will slowly evolve with time. They actually start to evolve as soon as the project starts. However, if your organization is similar to others and plans for all adaptations along the way, you will be crippled by planning and ultimately finish nothing. There has to be a system that prepares for changes in requirements, technical environment, and business environment so that these changes do not halt the deployment of a system, prevent its adoption, or prevent its extensibility into future solutions for an organization.

Change control is the process of managing this change. In any system, the life of the system can be impacted heavily by change in organization structure, staff, or even just focus. Change control is the tool used to mitigate these negative impacts on the system so that they don't accumulate to reduce the life of the system. In ECM, change control starts at the project kickoff and lasts throughout the life of the system on to the next one. It defines the roles, what happens when a change to the system is requested, and who is responsible for the longevity of the system.

By its very definition, ECM is global in scope, so most business systems and processes have an impact on other departments in one way or another. Truly understanding the impact that poor change management can have is usually felt when a system goes down or a process fails. This happens for a variety of reasons, and we have all been there when things go bad due to a random change made without planning, approval, and documentation.

From the field

Specific staff such as legal, records managers, clerks, and content managers usually take on the management aspects of ECM. Ideally, when managed and implemented well, taking on these responsibilities has a low impact on its users, with the right amount of control to ensure the success of the system. Users do not like being managed, but they will respect a system that is consistent and always up and running in a proper way. They will also respect the fact that not just the user has to be concerned about their content; the organization has to be concerned with the integrity of the entire catalog of documents contributed by all users. - Shad

Deliver

The content delivery stage is the process of enhancing content with new information or consuming the content it already has. This includes editing of existing documents, changing of metadata, and sharing of the content with other users. The components of the Deliver stage are as follows:

- Search
- Editing and viewing
- Publishing

Users love the delivery portion of content management. It's where they start to see the value of capture, storage, and management. The biggest value of content comes when it's used effectively in a decision-making processes. To do so efficiently, the user interface and functions need to be created to facilitate the rapid viewing and editing of documents.

These tools should be fast and effective. The knowledge worker should burn as little time as possible getting to the desired content and spend the bulk of time reviewing and/or editing the content. The primary tools for delivery are search, editing, viewing, print, publishing, and collaboration.

Search

Search is the processes of using keywords or Boolean logic to locate content and information. The process of search is to issue a query and to review a set of results to determine an appropriate item. While the actual search query feature is fairly basic, the tools that isolate the correct piece of content or, even further, the correct page in the correct content are very complex. Such features include best bets, search refiners, and relevance.

Search can be both a positive and negative indicator. The more often that users find what they are looking for will imply that the Information Architecture is perfect. Also, the fact that you can search across the enterprise is empowering and reduces the time it takes for you to get to content. Therefore, search and Information Architecture are intimately tied and share joint planning. They are so tied, in fact, that they share components. Facets, also referred to as *refiners*, are components of Information Architecture as well.

The goal of search is to get users to content faster. Search can start out with very basic principles but quickly evolves to topics such as thumbnails, best bets, in-page relevance, and so on. Whatever the cool components of search, the end game is always the same: get the user to the right content with the fewest clicks.



Note The approach we will stress is to put more energy into Information Architecture and encourage the use of other tools rather than search. But when search is needed, keep it simple until a specific requirement arises.

After a user finds the document, either via search or browsing, they have to be able to view and edit it.

Editing and viewing

While users don't often realize where they spend their time or why they do what they do, a simple study will prove that users spend most their day in some sort of viewer. A web browser is a viewer, Outlook is a viewer for email clients, and SharePoint is a viewer for content. An amazing amount of time is spent reading and consuming content as compared to creating it. Most job functions spend the majority of their time consuming rather than creating, while other specific audiences, such as

technical writers, business analysts, and publishers, spend a lot of time creating and editing. For this reason, the ability to do this seamlessly is critical.

With the exception for records and archived documents, the majority of content is living. It gets edited and viewed on a regular basis. SharePoint has built in viewers for Office documents. This allows you to work with the documents in the manner in which you are accustomed. For other varied file formats, a different type of editor might be required. It's important in the planning process to understand how content will be consumed, edited, and repurposed.

You need to understand that viewing and editing ties heavily to the formats discussed in the Store stage of ECM. The fewer types of documents, the easier it is for organizations to creating great editing and viewing tools for them. The Office suite has the clear advantage of having essentially bundled viewing and editing capabilities with SharePoint, either with client applications or with Office Web Apps. Documents such as PDF, which is predominantly designed for viewing and not editing, have special considerations when it comes to ease of access. For example, do you allow users to open the PDFs in the browser or client application, or do they have to download to their local machine first?

While the aspects of viewing and editing seem obvious, its considerations are not, which is why it's an essential component to ECM planning.

Publishing

Publishing, which contains aspects of both search and viewing, is the process of pushing content or allowing content to be pulled, by individuals who are not necessarily the curators, for viewing purposes only.

Publishing incorporates versioning in the storage stage, formatting, and ease of access either by targeted search or great Information Architecture. Usually, publishing also includes portals that are branded with basic themes or other more complex branding to make it a great landing page for content consumption. These portals are called intranets or extranets. Content on intranets and extranets is usually read only and comes from another location within the ECM system. This book will discuss in detail creating and managing content for publishing, or rather, the locations from which content is pushed. It will not cover in detail the configuration and branding of such intranet and extranet portals.

Process

Content is not just viewed on an ad hoc basis, or edited, which is essentially another type of capture. Hopefully, it is incorporated into a decision-making process or line-of-business process. The problem is that these types of processes are usually designated for structured information in SharePoint, such as lists. Documents have an additional element of complexity because their content is unstructured, which means incorporating them into any process requires special consideration about their associated metadata.

Process is taking what has been stored in content and incorporating it into another line-of-business activity. These processes are often unmanned and automatic. However, a manually driven process still falls into this aspect of ECM. The elements of process are as follows:

- Workflow
- Business process management
- Business intelligence
- eDiscovery

Workflow

The most common example of process is referred to as *approval workflows*. This is the process of routing a content item or transaction through a series of predefined steps for approval between different layers of management. This is very common in Human Resources, Finance, and Procurement.

For purposes of this book, workflow describes a process that contains one or more states or steps, incorporates user and system tasks, and is routed in a single direction. The purpose of workflows is to take metadata and content and turn them into action. A SharePoint workflow requires an existing business process to be well understood and defined. In most cases, work just gets done, but the flow of how that is accomplished is not documented or fully understood by all knowledge workers. To facilitate that process from beginning to end, there are user driven activities, management decisions, and transactional exceptions that need to be reviewed and completed. Workflows are very high-value items to automate, but most organizations underestimate the amount of due diligence required to prepare for a workflow automation project.

In ECM, the considerations around workflow are not just the content that is moved around via workflow but the steps and considerations of the content flow itself.

Business Process Management (BPM)

Business Process Management looks very similar to workflow, but it differs in that it allows for multidirectional processes, ability to version processes, and change control for processes. It is said that workflow is available out-of-the-box with SharePoint, but BPM usually comes via third-party solutions. The biggest technical difference is not just the workflows that can be created but also the management of those workflows.

In later chapters, workflow and BPM will be discussed together with the assumption that the user is performing out-of-the-box SharePoint functionality.

Business Intelligence and BigData

After content has been captured and processed, gaining insights from the content is a great way to take their value even further. Business intelligence (BI) is a broad category of technology that extracts greater value from content.

The bridge between BI and BigData is very strong. Therefore, we have lumped them together. Both increase the value of the data and assume some sort of structure and good metadata. Arguably, BI is a subset of BigData, although BigData implies manipulation of large datasets, whereas BI could be big or small. We illustrate this in Figure 1-4 by showing content actualization, starting with the most complex and highest cost and moving toward the most commonly used to achieve tangible business value.

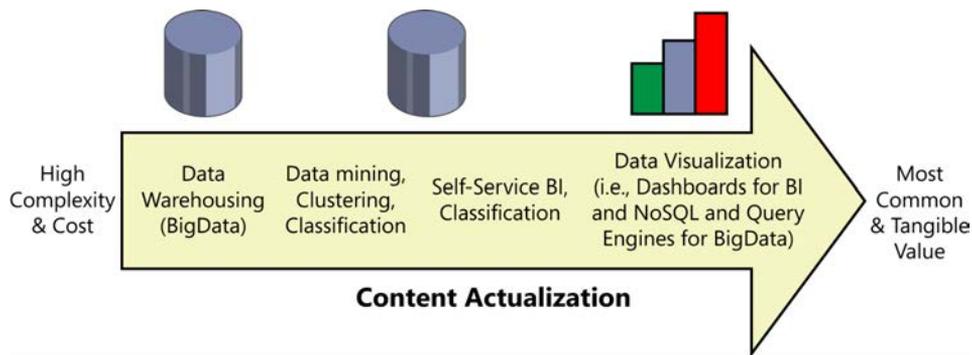


FIGURE 1-4 Content actualization.

BI really falls into the following three types:

- **Dashboards and Key Performance Indicators** This is most often what people are referring to when they say BI. In SharePoint, this is Performance Point and Conditional Formatting.
- **User-driven BI** This requires some data expertise and is usually performed using Excel and PowerPivot.
- **Data mining** Data mining is the intelligent extraction of value, and in SharePoint, it relies on a third-party tool.

Today, most people refer to BI as the ability to visualize large amounts of data in a graphical way, such as a graph. These are referred to as Dashboards or Key Performance Indicators (KPIs). They are fed by structured content, but unstructured content can be incorporated by using technologies such as Natural Language Processing, Text Analytics, Auto-Classification, or incorporating search into the analysis.

User-driven Business intelligence in SharePoint manifests itself as Excel spreadsheets and, more popularly, the PowerPivot add-in for Excel and SharePoint that allows a user to manipulate data in three dimensions or cubes.

Data mining is the most advanced and expensive use of BI. SharePoint can certainly be a source of information for advanced data mining tools but does not natively have data mining tools.

Arguably, BigData is just another definition of, or a broader definition of, a business system that includes and implies larger data sets and the use of a new database approach called NoSQL. Again, SharePoint does not have native BigData support, but it can certainly be used as a source for BigData third-party tools.

eDiscovery

A very specific type of processing of content is called eDiscovery. This particular process is one that organizations hope to avoid. Not only is the process of audits, litigation, and so on painful to the organization, the cost of such processes when not planned for is phenomenal.



Note An organization of 250–500 employees with 2–3 TB of data will pay in excess of \$300,000 to cover the cost of just eDiscovery associated with litigation. Organizations well prepared for eDiscovery not only reduce the cost when they are faced with a matter; they improve the overall organization of their ECM system. A *matter* is the subject of legal action, compliance, or content audit. *Discovery* is the process of gathering content associated with the subject.

When planned for, this cost can be dramatically reduced. And the process of planning for eDiscovery, fortunately, is the same as all aspects discussed in search and Information Architecture. Why? Because eDiscovery is essentially search combined with records management.

eDiscovery seems like some abstract term, but it really talks about an advanced form of search. In a later chapter, we will go into more details about what eDiscovery is, how it's used, why it's used, and how it works in SharePoint. eDiscovery is the identification, isolation, and locking of any content that pertains to a matter. The most common instance of a “matter” is litigation. However, a matter could relate to the Freedom of Information Act, content audits, and so on. After eDiscovery is run and content is identified, it must be isolated and separated from content not associated with the matter. The relevant content must be held as a record so that it is not updated or modified, which would make its value null.

It's safe to say that delivery and process are the areas we all enjoy the most when it comes to consuming content. It's also the area where the greatest investments and advances in technology are happening. The goal of ECM is to get users to spend more time in the Process stage and less in the Capture, Store, and Preserve stages of ECM.

Preserve

Content, just like everything else, has a shelf life. Most of the time, content that expires is deleted. This is beneficial to organizations from a compliance and legal perspective and as a part of ECM, but there is some content that is long lived. This type of content, when not consumed regularly, just takes up needless space in an ECM system.

Content preservation is about taking that active content generated as part of ECM and moving it to a location in a format that can be accessed, although infrequently, in the foreseeable future. The important aspect of preservation is storage. In this context, storage refers to both the physical storage of the content and the format that the content will be stored in.

Most organizations prefer to use the high-availability content databases for active and vital content only. For archive content, the preference is to see its metadata but not allow the content to take up space in the content databases. This can be done by using tools such as remote blob storage (RBS).



Note Before active content can become inactive archive content, it must be reformatted.

Reformat

It is very common to reformat content when it's preserved. Ten years ago, the trend was to convert content to microfiche, but today it's PDF/A. The purpose of reformatting is to ensure that it's not editable and to ensure that it can be viewed in the future. The reformat process also includes the purging of unnecessary versions and metadata. In addition to reformatting, many organizations also choose to compress their content.



Note PDF/A is a special PDF format designed for long-time archival of content. It is an open standard that helps ensure the quality of the content and the ability to retrieve the content in any viewer designed to read the format. In addition to the content, it contains standards from metadata.

Reformatting also has the consideration of viewing. How can you be sure that the content you have preserved can be viewed in the future? This is where librarians and content preservationists excel.

Compression

One major consideration for reformatting is the size of files. There is a process of compression that is often considered for reducing the size of files prior to final preservation. The problem is that it's rare to find compression technology that does not alter the content of a document during the compression process. Lossy content, or content that loses quality each time it is converted and/or compressed, is similar to the result of taking a picture of a picture. Over time, information is lost, and the possibilities for editing, consuming, and repurposing content in the future diminish with each iteration.

Compression is just one type of transformation process that is commonly used in ECM. This process reduces the file size of a document without loss of integrity of the content. With graphics, compression can reduce the quality of graphics, but the content of the documents remains. Compression becomes significantly important when discussing ways to save space and archival processes.

Overall, preservation is not all too common, and as the cost of storage reduces and technology for managing content improves, it becomes less and less common. For that reason, preservation will be referred to in this book as a possible end to a document's life in SharePoint, using such tools as RBS and SendTo locations.

Why use ECM?

Now that we have covered all the aspects of ECM, let's talk about the "why" and "who." Specifically, why should your company, department, or team implement ECM? Who will be responsible for implementing ECM, and who will use and benefit from it? The remainder of the book will then cover the "how."

Do not mistake the question of why you should pursue ECM with anything your organization has in terms of current SharePoint implementation. Our goal for "why" you will implement ECM is a question that assumes the perfect hypothetical ECM system, a system where a knowledge worker's effort to capture and categorize content is minimal but the amount of metadata capture is high, and the cost of finding and consuming the content is very low.

These goals are often at odds with each other. For example, expecting a user to enter content into SharePoint in the perfect way means that at time of capture they have to put in additional effort. By not getting the content captured in an ideal manner, the findability of content suffers from long search times and getting, at best, duplicate content and, at worst, the wrong content. In Figure 1-5, we show the contrast in effort between getting content into SharePoint (Capture) and retrieving or finding content (Consumption). Ultimately, you want to automate the process of capturing content to provide enough structured and validated metadata to make the consumption of the content relatively easy.



Note *Findability* is a measure of the effort it takes to locate a document. Content with greater findability enables a user who intends to locate a document to do so with the least amount of effort. Findability can happen at the platform level with tools that improve the quality of all searches, and at the content level with better metadata. You will sometimes hear the term *putability*, which, similar to the term *findability*, is the measure of how easy it is for a user to contribute high-quality, easily findable documents inside of SharePoint.

It was said earlier that it is possible for organizations to plan forever. So it's important to note that all aspects of ECM come with a balance. The emphasis is unique to each organization and its current environment. While it's frustrating, it's a fact of ECM life, and it needs to be accepted.

Therefore, for the sake of this section, we will assume that the ideal ECM solution exists. Your organization's motivating factors for implementing ECM will come from two primary drivers: reactive or proactive.

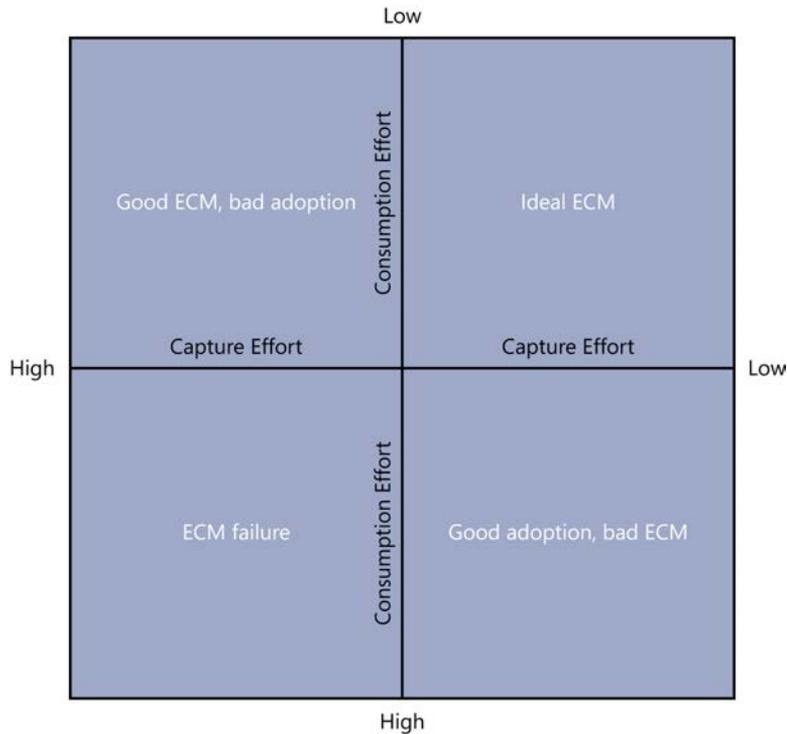


FIGURE 1-5 Capture and consumption.

Proactive driver

It's rare for knowledge workers without the skill set of content management to be interested in ECM without a reactive driver. However, you might have departments or managers that promote a culture that values organization. And from the perspective of being organized, they embrace the methodologies of ECM. Ironically, it's also these departments and this culture that resist change, because they very often have a well-established system for filling content and are not much interested in taking that away.

Another area where organizations are proactive is when ECM has a direct impact on operational costs. This is most often where content touches a line-of-business application. This can be found in health care, insurance, accounting and finance, banking, and so on—that is, organizations where everything has a process and there is content involved. The benefit of these organizations is that they usually already have a retention schedule and content that varies very little in topic, so building an ECM solution is very simple.

And finally, there are the select few, like the book's authors, who love the idea of getting content under control so that greater value can be attained. It is our goal that some readers of this book will finish in this category. Now let's talk about those reactive drivers which are a more common catalyst for ECM.

Reactive driver

Reactive organizations have faced some negative event where the cost of poor ECM was evident. The most common such events are litigation and compliance, where a company was the defendant or claimant in a lawsuit. I'm sure that you can spot something in your organization where you can't help but laugh at the way things are done. At some point, these areas of operation, you hope, are addressed with new approaches.

The second most common is compliance driven. A good example is if your organization operates in an industry that is compliance-heavy, like health care, insurance, or banking. These industries deal with a tremendous amount of regulatory burden. In fact, many of the legislative actions taken to create regulations like SOX (Sarbanes-Oxley), HIPAA (Health Insurance Portability and Accountability Act), and Dodd-Frank (Wall Street Reform and Consumer Protection Act) contain very specific provisions that can be complied with by using components of an ECM solution.

The benefit of the later reactive scenario is that very often, in such cases, the implementation of ECM is very clearly defined, and it's more a matter of conformity. Most often, these organizations, due to their history in the industry, already have a good sense of what needs to be done, and their business challenge is adoption rather than implementation.

For the reactive scenario in litigation, these organizations very often were not aware of what was wrong until it hit them. The process of getting content associated with this litigation is called eDiscovery. Later in this book, we talk in more detail about eDiscovery. The cost associated with eDiscovery is phenomenal. Therefore, to be better prepared and to mitigate these costs in the future, organizations are deploying proper ECM solutions and participating in what is called early discovery.

The reactive scenario is not what the world of ECM hopes companies are using as their primary driver for implementing ECM. The hope is that organizations are thinking about what it takes to best actualize content, get the most value from content, and to make their knowledge workers as effective as they can.

This scenario is very rare and usually combined with one of the above reactive scenarios. There are a handful of organizations who believe that if they better capture, store, and deliver content, their costs of working with content will be reduced, the effectiveness of their staff increased, and finally, that they can move that content into more valuable decision-making processes, such as workflow and BigData.

How can you use information to make better decisions?

Information is just bits of data; it's stored everywhere, and access to the information is abundant. Putting information in context provides the ability to acquire knowledge. Knowing when and where to use the knowledge gained from the information is wisdom. Putting wisdom into action is what differentiates individual performance and enables streamlined decision-making. As shown in Figure 1-6, each step represents a greater degree of meaning derived from content. Departments, workgroups,

and knowledge workers who can harness the Information, Knowledge, and Wisdom steps will help an organization to perform at a high level.



FIGURE 1-6 Information steps.

Businesses generally make the mistake of thinking they are managing information by having hardware, software, and communications. In the majority of organizations, information is stuck in silos, duplicated without control and communicated in an inconsistent manner. A well-architected ECM solution can enable the enterprise to overcome the chaos and achieve remarkable results.

It is important to understand the difference between structured and unstructured content.

- **Structured content** A simple example is the color of red. This can be seen as a keyword value pair and can be used to create lists, tables, and relational context. This type of content is easy to get value from but is harder to create.
- **Unstructured content** This book is an example of unstructured content. This type of content amounts to 80 percent of all content and is easy to create.

The justification for ECM is to help knowledge workers create structure out of unstructured content. This helps them become more efficient by empowering them with content in the right context at the exact moment they need it. In today's world, the volume of information is exploding at an exponential rate. Knowledge workers at every level of the enterprise are adopting a myriad of devices and technologies. Enterprises understand the value of information; we can all agree that the right information can mean the difference between success and failure. The only way the enterprise can get to the stated benefits of BigData, business intelligence, process efficiency, and business automation is to design, implement, and manage an ECM solution that focuses on helping knowledge workers access information successfully.

What is a *flexible content structure*? This is the ability for anyone to create, name, and store content without policies or requirements for organizing content in a meaningful way. This is what enterprise and knowledge workers have been allowed to do since the rise of the personal computer.

As long as the flexible content structure world exists, you need ECM. Most people think that because they are creating content they are managing it. This is blatantly false, and the end result is nothing short of chaos. Users have no idea where to find content; they spend most of the time searching and sifting through irrelevant and duplicative bits of information that was never stored with any thought of how to find the information later.

Return on investment

The primary drivers that create a positive economic impact on the enterprise are revenue growth and profitability. A secondary driver, like increased productivity, is more difficult to measure but nevertheless creates a substantial return on the investment for a well-structured ECM platform. We know that knowledge workers spend 5 to 15 percent of their time reading information, but up to 50 percent looking for it. The ability for a user to quickly find the information they are looking for is important. Even more important is finding that information in context of the business activities or systems that they already use.

To achieve even larger productivity gains, you have to move beyond just finding information and toward automating the collection, distribution, and processing of the information. ECM platforms provide a distinct advantage for business process improvement by automating common processes and reducing or eliminating redundant activities. Process re-engineering is a fundamental part of implementing a successful ECM platform. This goes beyond the technology and requires a commitment to change management and feedback loops to help determine where inefficiencies exist. Some of the changes will be procedural in nature and might require changes in human behavior, which is always more difficult than configuring the technology.

The duplication of content is pervasive. In a world where storage is cheap and nearly everything has a synchronization option, it's important to think beyond just ease of access. At some point, the original content has to be discarded as it reaches the end of its life cycle. Your ability to ensure that content that needs to be deleted is actually deleted can provide the enterprise with cost avoidance by reducing the risk of having discoverable content in the eventual case of litigation.

Who does ECM target?

ECM targets different people for a variety of reasons and use cases. ECM is a set of disciplines that are used to guide the development of Information Architecture and governance.

A successful ECM platform supports the management of information for operational, transactional, and regulatory purposes. It provides access to the unstructured information assets that complete the record of a transaction, project, person, or activity. The information stored in an ECM platform is supportive in nature to all other systems used in the enterprise, putting structure around this information to control how it is used, who has access to it, and what processes are enabled to leverage it. As a document nears the end of its life cycle, the ECM platform is used to discard information to achieve compliance and reduce risk.

Everyone in the organization benefits from ECM. The key players that are targeted are broken down into those who implement and manage the SharePoint platform and those who access the information stored in SharePoint. Table 1-1 describes how to segment the roles and responsibilities of each of the individuals targeted.

TABLE 1-1 Roles and responsibilities of key individuals

Users	Technical	Management	Governance
Knowledge Workers	IT Operations	Executive	Records
	Developers	Departmental	Legal
	Business Analyst	Project	Procurement

Building expectations

SharePoint offers an amazing set of tools and functions that can be used to deploy a myriad of business solutions. To focus and give you tools that you can use from day one, we are providing an Information Architecture section. In this section, we will discuss the standard building blocks of a SharePoint ECM solution, namely Site Collections, Document Libraries, Content Types, and the Records Center. When you understand what these terms mean and how they work together, you can incorporate the control that every organization wants and needs. Governance can too often be unbalanced, creating either a system that is clunky and difficult to use or one that is too open and ends up being a digital landfill. This is where a well-defined taxonomy that is neither vague, overly granular, nor verbose will help everyone follow the same rules. We will look at types, best practices, and a real use case of taxonomy in action using the Term Store.

SharePoint 2013 is a very powerful on-premise server based solution that is also available as an off-premise Cloud solution. Primarily targeted at small and medium sized organizations, Office 365 takes minutes to sign up for but in the range of a few hours to be fully configured. Administrators will have immediate access to settings for provisioned services. The core Services, Lync and Exchange, are available in minutes, but initial configuration of a multitenant SharePoint site takes around two hours to complete. We will discuss the ins and outs of the configuration options and introduce the concept of the Office Store.

All versions of SharePoint have benefited from a large, talented, and enthusiastic partner community. SharePoint 2013 continues to provide great opportunities for Microsoft Business Partners and Independent Software Vendors to build complementary technologies and vertical solutions that extend the platform. We will outline some vendors and applications that you should evaluate during your project planning and requirements definition. Knowing what technologies are available and how they can help to make your SharePoint ECM solution better will be a real advantage.

The legal and regulatory aspects of ECM from FOIA to SOX can seem like an overwhelming bag of acronyms. To help you navigate what is relevant and what is not, we will introduce the most common laws and practices that you should consider and how to address them in SharePoint. Litigation is a pervasive part of the business landscape, and being able to support eDiscovery requirements in the event of a lawsuit is critical. We will cover the technical aspects of SharePoint that are necessary for performing legal holds, controlling full fidelity of content, and supporting chain of custody.

Next steps

We have covered the difference between SharePoint as a technology platform and ECM as a methodology. In Chapter 2, we will dig deeper into the ECM Stack and uncover in detail the complexities of each stage of the content life cycle. The sections will follow the same structure but include actionable steps you can take to get comfortable planning for ECM in SharePoint. As you're reading through each section, you should be gathering pertinent information about your specific environment and creating a high-level roadmap for implementation.

At this point, you should begin identifying the stakeholders for the ECM solution you will be designing. You need to separate the functions of SharePoint from the information architecture that your organization will use going forward. Identify your existing content repositories both inside and outside of SharePoint, and determine where they meet or fall short of the principles we have outlined in this chapter.

ECM stack: content control

After content is in the system, it's time to put it to work. You might think this is the easy part. Functionally, it is much easier than the creation and capture portions of the ECM life cycle, but strategically, it's very difficult. We have found many organizations that are several years past an initial SharePoint deployment looking back and realizing that the adoption of this technology has not really improved a business process. In many cases, it has either become a web version of a shared drive silo or a collection of mismatched sites that aren't easily navigated by users.

As we mentioned in Chapter 1, "ECM defined," an ideal EMC solution is a balance between the effort to capture content and the effort to consume or find content. The preceding chapter was all about getting content into SharePoint; this chapter is all about consumption and accessing the content you need in an intuitive and structured manner. After content is stored in SharePoint, you have to pay particular attention to how content is managed, who manages that content, and how the content will be delivered and consumed. Last but not least, you need to determine how the content completes its life cycle, through disposition or long-term preservation.

Management of content

When we talk about the management of content, we are not just referring to who is in charge of SharePoint. Yes, we will discuss how to make sure your content is being treated properly. But we are also talking about how to manage its value, the people who use it, and setting proper expectations.

Two extremes prevail when organizations start implementing SharePoint. It either is seldom used, or it's used prolifically without governance. Being aware of both of these opposing outcomes will guide your ECM team's planning and help determine where to focus their primary efforts. You can start by answering the following four questions:

1. Is your organization one that embraces new technology or resists?
2. Does your organization process work at a macro or micro level?
3. Are shared drives the primary source of information management today?
4. What other systems or processes are used to manage content today?

Being honest about whether or not your organization can embrace new technology is an important question. In general, younger organizations are more adept and open to change, adopting new

technologies, and improving processes. This is also true about people, so being realistic about who, where, and how you will implement and be successful with change is vitally important. Be aware of signals that people are resistant to change. These signals can come in the form of availability for meetings, a consistent focus on exception processing during design meetings, and the use of phrases like *"If it isn't broke, don't fix it"* or *"This is how we have always done it."*

It's important to work within the confines of how your organization adopts and addresses change. Specifically, in question 2, we ask whether your organization processes work at a micro or macro level. What we mean is this: How are decisions made, and who can effect change in work processes most effectively? In some organizations, groups and departments are given a fair amount of autonomy to get work done in the most efficient manner, using the tools and techniques that suit them best. In other more structured organizations, everything is mandated from the top down and there is a great deal of command and control.

Depending on your situation, you need to know how and where to start scoping the management aspects of content in your SharePoint ECM solution. This will also help you determine what individuals need to be on the ECM team, which we will cover in Chapter 5, "Building an ECM team."

Change opposition or support

We implied earlier that ECM is as much, if not more, of a people problem as it is a technology problem. Although we use common nomenclature for business processes, like "expense approval," and for departments, like "accounting," no two organizations are alike. Because of this, it's important to use principles and methods to guide your ECM solution design and not necessarily defined steps or configuration outlines.

We have found that there are companies who love new technology and, as an aspect of the culture, cannot wait to get their hands on it. These types of organizations or groups will be very excited to start using SharePoint and will be your early adopters. You can leverage their enthusiasm, but you will need to contain it, because this is usually the beginning of how viral use of SharePoint starts. Without Information Architecture (IA) or governance, these users will adopt bad habits and the ECM solution will turn out poorly. When this group establishes a bad habit, it's hard to stop it.

You might find that the organization in its entirety or in specific groups is very used to process and control. These control-conscious groups are welcoming and normally not opposed to systematic ways to organize information. If you look at their shared drives, they tend to be very organized already, which we hinted could help with IA planning we cover in Chapter 4, "Cases in point." While this group will be great for adopting SharePoint for ECM in the correct way, they will be hard to get moving, and they tend to not like change.

The other interesting element in this group is the nature of their content; it is typically more transactional and repetitive, as opposed to unstructured information. Most organizations have one or more of these types of groups. Some common examples are in finance, engineering, operations, legal, and human resources. Also, many industries are very used to structured processes and governance such as healthcare, insurance, and financial sectors.

Both types of groups, if already using shared drives as the vast majority of the content world does, believe that they have a system for organizing content and that it's just fine. Whether a knowledge worker admits it or not, their slice of the shared drive pie is a dumping ground for content. Sometimes it's organized, and at other times it was created while on a conference call. The net result is a system that they know because they built it, but one that does not serve the overall organization well. To understand how people might want to use your SharePoint ECM solution, it's important to understand how content is being stored in these shared drives today.

How users currently manage their content is a good indication of how they will want to manage it in SharePoint. Habits are very hard to change, and if a person has been doing a particular job function for many years, they are used to adopting workarounds to get the job done. At the first sign of any struggle, a user may revert to old habits, such as looking for a workaround or complaining that the new system just doesn't work. This is compounded by daily stress and workload that make the burden of using a fancy ECM solution even higher.

The four questions we asked at the beginning of this chapter were not outlined to establish any specific components of your implementation. Rather, they are meant to generate introspection so that you will begin to understand where your problem areas will be. You should use this knowledge to identify what groups will provide opposition or support for your ECM team's planning of a successful implementation and extension of SharePoint. The next step is to use this information to address the specific areas of the Management portion of the content life cycle. The following outlines the primary areas to consider for who manages the content. Please keep in mind that the management of content is not the same thing as who manages the security or access levels to the content.

- Security
- Document types/retention schedule
- Document audits
- Managing bad habits
- Policy creation and implementation

Who manages content?

As a best practice, we recommend that the individual(s) in charge of managing content in SharePoint should not be the same people who maintain SharePoint, typically the IT department. In fact, depending on how large your deployment is, you can and should have many people who are responsible for managing content. The more people who are active managers of content, who understand the principles and benefits of your IA, the better chance you have of maintaining control of your information.

SharePoint as the ECM platform is the responsibility of the IT department, from the server and storage infrastructure to the farm installation and configuration. Their involvement or, rather, responsibility should stop at the site collection level. This is typically a blurry line, where IT responsibility

ends, and your IA and general ECM governance begins. Below the site collection, an individual(s) who has a vested interest in the success of ECM in your organization should manage your ECM solution.

In many organizations, this person is the Records Manager, Content Manager, or Information Architect. Some small organizations don't have the luxury of such a position for budgetary reasons. Even some larger organizations haven't identified the need for this specific position. There are a variety of reasons why this might be the case—for example, there is little regulatory oversight or the organization has never been party to litigation. In such cases, you have to consider appointing a team that is responsible for the content or, if the organization is large enough, individuals in each functional unit, often referred to as *Super Users*.

This is challenging in many organizations because both IT and the Content Manager or Super Users have, or should have, a stake in the ECM solutions performance. For example, it should be a defined part of their regular responsibilities or tied to a Managed Business Objective (MBO). In some cases, this can create challenges between IT and the Content Managers because they often do not have the technical know-how to take over at the site collection level.



Note Managed Business Objective (MBO) is a tool used by many organizations to establish goals above and beyond personnel job description. Monetary incentives usually accompany achieved MBOs to help acknowledge the extra effort required to meet the objectives. For example, a portion of the quarterly employee bonus can be based on MBO. Including MBOs for the effective use of the ECM platform is one technique that organizations are using to encourage proper adoption without the use of technology.

The first problem is usually addressed by creating synergy between Content Managers and the IT staff. Both should be on the *ECM Committee*, which we will talk about in Chapter 5, “Building an ECM team,” and both should be bound by the common goal to make the ECM solution work, both technically and operationally. In most organizations, this works out well, but it also always helps to draw a clear line in the sand stating that IT is responsible for everything up to site collections and the overall performance and uptime of the farm and that Content Managers are responsible for everything below site collections.



Note Specific roles and responsibilities should be part of a well-defined governance plan and formalized documentation. Having a person(s) assigned to the role of Records/Content Manager or Information Architect is the best scenario and can help further define the organization's commitment to ECM.

The second problem is much harder to address, and organizations can really address this issue only when they identify and recruit the required staff well, before SharePoint is selected as a platform. Many organizations luck out and can identify users in each function who aren't in IT but are technically savvy and eager to be a part of this exciting new technology initiative. As for officially titled Records Managers and Content Managers, it is good to make sure that they have sufficient

SharePoint experience and training. A rule of thumb is that it takes about one year with hands-on experience with SharePoint to really begin as an administrator.

Security

If you have ever attended a SharePoint event, you know that one of the most common topics or common subjects within a given topic or track is security. This is both good and bad. The good news is that there are many individuals out there who know a lot about SharePoint security. The bad news is that the number of sessions at any conference on a particular topic is a gauge of how complex that topic can be.

Our approach is different. When it comes to ECM security, the best principle is that less is more, so keep it simple. This is one of the many areas of building an ECM solution in SharePoint that could result in planning paralysis.

From the field

Over-architecting security or any portion of SharePoint can result in a solution that is difficult to manage and use. During a recent project I worked on, it was clear that the security team was more interested in creating a complex security hierarchy that only they could understand and manage, this caused the project to stall. It finally came down to usability of the site collection, and the business unit was able to convince the CIO to override the security team. In the end, the security model established was good enough and the ability to get content in and then manage and use it was excellent. - Shad

We know and can recommend that simple is better because indirectly, while you are designing a fantastic IA (covered in Chapter 2, “ECM stack: content in”), you are determining security. That’s right: security users and groups are very similar to IA. And fortunately, IA in SharePoint indirectly helps determine your security structure. So with a well-designed IA, part of the security work is already done for you. Now let’s cover the two basic types of security in SharePoint: repository level and document.

Repository

Repository level security is the shared responsibility of the Content Manager(s) and IT. The repository level security should map to functional or departmental security groups in your organization’s active directory. Active directory most often is where users and access levels are assigned. Typically, users are added to groups based on their department or function. Most organizations have this well established, which also means it cannot be changed. Although every organization is different and therefore every active directory forest, tree, and domain structure is different, the goal is to have these high-level functional groups align with the SharePoint repositories.

As we explained in Chapter 2, repositories are both the logical and physical location of content. The top-level repository is the web application, also known as the site collection. The next level is the site, and the final level is the library. The number of web applications is roughly determined by the size of the organization and the amount of content. While it's certainly possible to have multiple content databases, in this book we will assume that there is one content database per web application.

This means that the site collection is the first point, or envelope, that you can apply security to. For ECM, three general principles to follow for security are as follows:

1. Never assign user level security.
2. Never assign security lower than the site.
3. Never break inheritance lower than the site.

In general, one of the biggest challenges of security is the administration of it. Alarm bells go off in every organization when the security topic comes up; multiple people want to take ownership, and frankly, it becomes overthought.

When you think about it in simple terms, it's rare to find an organization where everyone in a department or group can't see all documents. Of course, there are exceptions; for example, Executive Management. However, the proposed approaches for IA that we outline in Chapter 4, "Cases in point," will show how this is addressed by having a specific site collection or site for the organization's leadership groups.

Therefore, the primary consideration is the groups, what the groups are applied to, and cross-functional sharing.

The ideal scenario

Based on our experience, we propose that each group gets assigned their portion of the IA. For example, the Human Resources security group gets assigned to the HR site collection or, for small organizations, to just a single site. The Engineering security group gets assigned to the Engineering site collection for large organizations or to a single site for small organizations. The manager of both of these functions should belong to a separate security group called Managers, and that group should be assigned to the Managers site collection for large organizations or to a single site for small organizations.

Our goal is that you begin to see a pattern emerging that will become very clear in Chapter 4 as you work through the practical examples. You should also be able to guide each group as they build out their portions of the IA by using the examples and templates we provide in this book.

This will also allow you to adhere to principle number 1, which is to never assign user-level security. With this method, you should never have to apply ad hoc security to individual users in the organization. The primary reason that we do not want you to do this is because managing it is very problematic, if not impossible. Without a third-party management tool, it's impossible to know where these ad hoc users have been assigned. It is also very difficult to make sure, when security is no longer needed or when the user is no longer with the company, that the user can be or is removed.

Team site

Cross-functional projects/repositories needing many users to be involved should be created in a separate site collection for teams. In this site collection, the root-level security is the entire organization, and the team leader will solely determine the individual team site security in an ad hoc manner.

This seems like an exception to the rule, and it is. But the concept is still within the principles of ECM. The result of the project—that is, the individual team site—is the document that belongs to ECM. Therefore, the team site itself is a sort of record. The team site collection is an ECM system for all projects, and because projects have short lives, the project and all activity within that project live and die there.

The biggest challenge with this approach is to make sure that team sites are deleted when the project is over and, if the result of the project was a document, that the final document ends up in the approach function in the organization. For this, in SharePoint 2013, we recommend using the new site disposition functionality, and in SharePoint 2010, we recommend looking at third-party tools or relying on the team leaders, via policy, to clean up after themselves.

This approach addresses the need for cross-functional work, keeps the principles of ECM in their designated functions, and makes IA and ECM management flexible. Figure 3-1 shows an example of a project team site used as a cross-functional repository for the Projects team site.

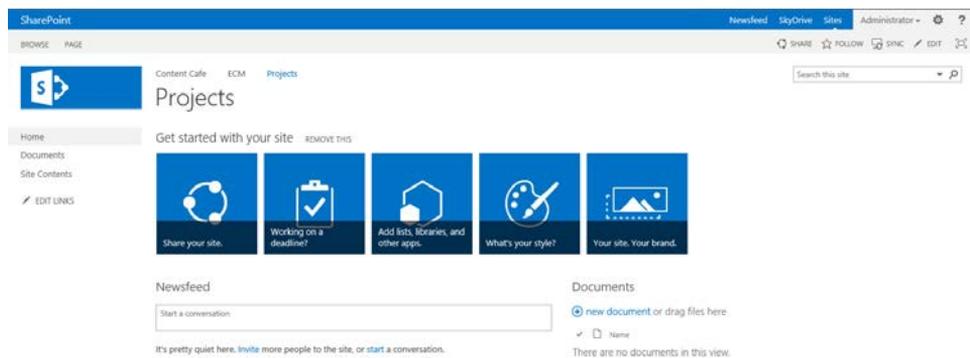


FIGURE 3-1 Projects team site.

The final two principles we established for repository security were never to break inheritance below the site level and never to assign groups lower than site level. Both are established because of the fact that security is very hard to manage, and as soon as management is lost, it's a rapid downward spiral in security and sprawl issues. This is especially true with the new share functionalities present in SharePoint 2013. Therefore, to keep security very clear and administration clean, we recommend that security be applied only at the lowest repository envelope, which would be the site collection for large or document-heavy organizations and a single site for smaller organizations. After this has been enabled on these repositories, it should not be broken out any lower.

It is even possible to get item-level security, and while it sounds fun, it's not. We have yet to find a fully justified use case for when item level security is right.

Site administration and recycling

In addition to repository security, the following SharePoint-specific security considerations will be important to understand and plan for when implementing your ECM solution:

- Site collection
- Site administrators
- Recycle bin

Who has access to administration functions on the site and, more importantly, the site collection is an important consideration, primarily because of the risk it could impose on the stability of SharePoint. A site collection and site administrator have the ability to modify security, enable and disable features, navigate to terms sets, modify content types, and so on. Without malicious intent, it's very easy to change something small, such as a content type, and harm the user's ability to use the system.

The process we recommend is to have no more than three site collection and site administrators. If at all possible, assign only site collection roles, and no ad hoc site administrators. Your ECM solution, when established, should not change often, so the need for a site administrator should be minimal. The group of three should include two Super Users or Content Managers and an individual from IT—most likely the farm administrator. The purpose of three is redundancy and accountability. In the governance section of Chapter 7, "ECM planning guide," we will talk about tracking these users and changes.



Note These three individuals don't need to be 100 percent sure of all the functionality available in site settings, but they should be aware of the risk of changing things suddenly and have an appropriate level of cautiousness when making any changes.

From the field

Every time I open up SharePoint Site Settings, it takes about five seconds before I know where I am, and I will be delayed trying to remember where that pesky link I want to use is located. There I am, stuck on the site settings page looking for "Term Store Management." From an administration level, SharePoint can be daunting to navigate even for the most experienced. - Chris

The recycle bin in SharePoint has special considerations at the site and site collection level. The recycle bin is where documents go when they have been deleted. By default, recycle bins are set up in two stages. This means that there are two recycle bins: the site recycle bin and the site collection recycle bin. The settings for this are found on the web applications settings page in central administration, as shown in Figure 3-2. It's possible to turn off the second stage recycle bin, and it's also possible to set a purge date for when the recycle bin is cleared: by default, 30 days.

Recycle Bin

Specify whether the Recycle Bins of all of the sites in this web application are turned on. Turning off the Recycle Bins will empty all the Recycle Bins in the web application.

The second stage Recycle Bin stores items that end users have deleted from their Recycle Bin for easier restore if needed. [Learn about configuring the Recycle Bin.](#)

Recycle Bin Status:
 On Off

Delete items in the Recycle Bin:
 After days
 Never

Second stage Recycle Bin:
 Add percent of live site quota for second stage deleted items.
 Off

FIGURE 3-2 Recycle Bin settings.

This method is useful because it's not uncommon for users to delete a document in the site and, on day 31, realize that they should not have. The consideration that needs to be made is who can restore, empty, and administer the recycle bin? Typically, the same three site collection administrators we outlined are the same who have access to administer the recycle bin. There are also two separate types of bins at each level: the end-user recycle bin and the administrators recycle bin.



Note There are unique scenarios for some companies where even the site collection administrators should not see the content of deleted items. In this scenario, if acceptable, we recommend enforcing with policy, due to the small number of site collection administrators. Otherwise, a custom solution is required to modify the security of the recycle bin.

It is also good to plan the limit size of recycle bins and to have a policy for when, if ever, they are emptied separate from the defined purge period. Ideally, organizations will establish a time period and adhere strictly to this rule because, as we will see in Chapter 8, "Records management," following your own policies is critical for remaining compliant.

Document

We have established security to the repository, so now it's time to consider what a user can and cannot do with the documents that are stored in the repositories.

On a site and library level, you can manage different types of security groups. These groups, separate from those in Active Directory, determine what can be done on a document level. For example, you can decide which users can edit documents and which users can only read documents. These settings should not be viewed as tools for producing content; rather, this is the security that happens on the repository. They should be viewed as a way to protect the content of documents from unfortunate editing by the wrong people.

From the field

I'm never surprised when I audit a SharePoint environment and I see a user belonging to multiple security groups, custom security levels on every site, individual users assigned to a site, and more document-level security options than I could ever imagine. With SharePoint, keeping it simple is the only way to succeed. Fortunately, in ECM, it's structured enough that there is really no reason to break this rule or the principles we outlined. - Chris

You will find many approaches and views on this. Unfortunately, most organizations mimic security levels in their active directory. This is OK, but it usually results in too many groups to manage. Then there is the approach to consider just the types of activities that can happen to a document. For example, a user could perform any one of the following activities:

- **Design a document** Provide the ability to view, create, delete, approve, edit, and customize.
- **Edit a document** Make modifications to existing documents and delete them.
- **Contribute to a document** View, create, update, and delete existing and new documents.
- **Read a document** View and download existing documents.
- **View only** This is similar to read, but you cannot download.
- **Moderate a document** View, add, update, and delete items.

There are a few special actions that are reserved for custom solutions and web services that we are not covering in this book. All of these options distill to the ability to create, delete, modify, read, download, approve, and customize.

Because we have isolated document-level security as a way to protect the content of documents, we have included a decision tree in Figure 3-3 to illustrate a process of defining document-level security.

Even between experts, this approach can be quite controversial. The reason for such a radical approach is to maintain the ease of administration that SharePoint needs to have so as to ensure that the platform is adopted and extended. We believe that most organizations will have the requirement for the minimum level content security groups, which are Full Control and Read-Only.



Note The most important thing in your ECM solution is to make the approach the same for all sites, and make sure that you also consider how hard it will be to administer.

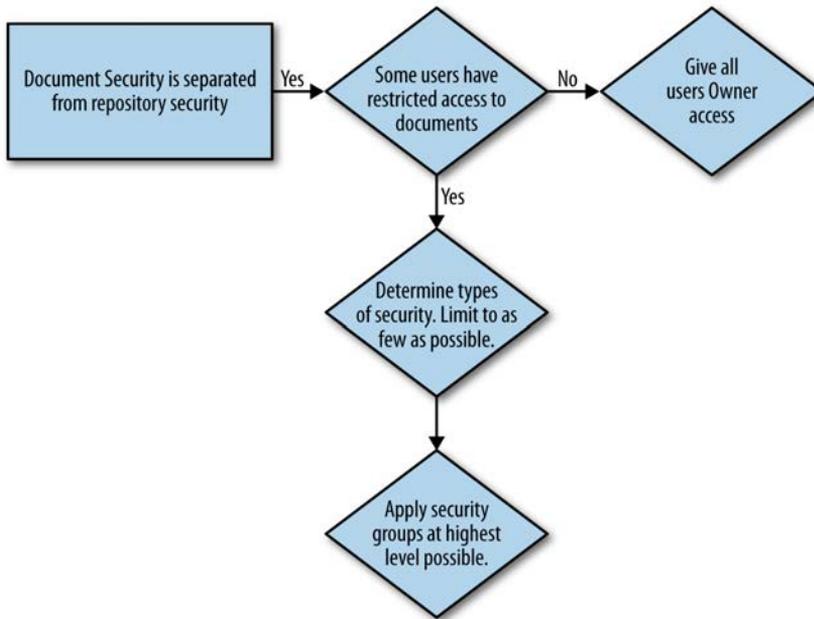


FIGURE 3-3 Document security decision tree.

Document management

Also at the content protection level are the check in/out functions and the versioning functions. Checking out a document is a very important yet simple tool in ECM that ensures that as a document is being edited by a user, the editing process is not impacted by another user who views the document.

From the field

When we were writing this book, we strictly adhered to the check in/out principle. There was not a time where we were working on chapters and the documents were not checked out. In the event Chris or I wanted to view progress, we would not impact each other's updates. - Shad

If you use default settings in SharePoint, it is up to the user to check out documents and to make sure to check them back in, at which point they should also add comments. Because it's at the user's discretion, you will find that when they own the content, they will tend to use this feature, because all users are sensitive to the additional work that could be created if someone else saves a version while they are editing. However, it's not guaranteed, so we recommend that you enforce that whenever a document is being edited, it is automatically checked out. As shown in Figure 3-4, you can do this on

a library level in library settings and by selecting versioning settings. You can select Yes for the option Require Documents To Be Checked Out Before They Can Be Edited?

Settings › Versioning Settings

Content Approval
Specify whether new items or changes to existing items should remain in a draft state until they have been approved. [Learn about requiring approval.](#)

Require content approval for submitted items?
 Yes No

Document Version History
Specify whether a version is created each time you edit a file in this document library. [Learn about versions.](#)

Create a version each time you edit a file in this document library?
 No versioning
 Create major versions
Example: 1, 2, 3, 4
 Create major and minor (draft) versions
Example: 1.0, 1.1, 1.2, 2.0

Optionally limit the number of versions to retain:
 Keep the following number of major versions:

 Keep drafts for the following number of major versions:

Draft Item Security
Drafts are minor versions or items which have not been approved. Specify which users should be able to view drafts in this document library. [Learn about specifying who can view and edit drafts.](#)

Who should see draft items in this document library?
 Any user who can read items
 Only users who can edit items
 Only users who can approve items (and the author of the item)

Require Check Out
Specify whether users must check out documents before making changes in this document library. [Learn about requiring check out.](#)

Require documents to be checked out before they can be edited?
 Yes No

FIGURE 3-4 Versioning settings.

A side effect of this important feature is users forgetting to check documents back in. This should first be addressed with training, and eventually there will be enough social disruption around popular documents that users will make it a habit. Administrators do have the ability to force checked-in documents. We have even seen some unique customizations that run on a daily basis to find checked-out items and remind users to check them back in via email. This is another people and technology problem that is better to address with people first.



Note Custom solutions add a variable to SharePoint that could impact extensibility, migration, and stability of the system. From time to time, we talk about custom solutions. By default, we recommend always using out-of-the-box features to ensure the life of your platform and that the ability to migrate to future versions is as easy as possible. When we talk about these recommendations, we are only hinting at some interesting possibilities; they should be considered carefully.

While in the versioning settings, we have something else to look at here. The next important consideration in protecting content, and a very common ECM feature is versioning. There are two types of versions, as we noted in earlier chapters. The first type is *major versions*, and the second is *minor versions*. A major version is everything to the left of the ".", and a minor version is everything after. For example, if a document has version 3-2, it has major version of three and a minor version of 2. Both sides of the decimal can be essentially infinite. Therefore, version 3-125, the 125th minor version of major version 3 of this document, is possible but not advised.

The first consideration is whether you should use versioning. In our experience, it's hard to find an ECM scenario where versioning is not a necessary feature. This helps make sure that the activity associated with the edits to a document is well contained and that you can revert to old versions if you need to. We did a study that shows that about 1 in every 20 documents has a need to be reverted back to a previous version. Only if documents are read-only, if you have technical limitations on your content database, or if you are advised by legal counsel (because every version is admissible content and available for eDiscovery) should you consider not using the feature.

When an organization has decided to use versioning, it must consider whether or not to implement minor versions. It's rare for an organization to have a need for minor versions unless they have implemented publishing. Publishing is the process of taking major versions of content and publishing them to other users in the same site or to intranet/extranet sites. This allows major version of documents to be published while the creator of the document can continually modify working versions. In most cases, organizations need only major versions. Unless you have strong documented reasons for doing otherwise, we recommend sticking with major versions only.

The final item to consider for content security is draft item security. It is possible to limit which users or groups can see draft versions of documents. In ECM, we feel that any setting other than "Any User who can read items" is dangerous. Does your organization use drafts and, if so, why? We find that most organizations, unless utilizing publishing features in SharePoint, don't need to use the draft feature in most cases. If you are utilizing drafts, because this is such a subtle feature, it's very difficult to document whether or not there is a special security consideration. As a result, if multiple editors are collaborating on a document, with some having draft privileges and others without, the problem of lost content will not be easily identified. It might be necessary at times for people without proper edit security to know about the existence of draft content. For these reasons, we recommend that, when using drafts, you should allow everyone to see the draft versions that exist.

We have outlined all the considerations for managing document repository security, ensuring integrity of the content, and special considerations as it relates to working with content in ECM. Now that your users have captured the content, your ECM solution is managing it because of the IA the ECM team put in place, and the users are putting the content to use. In the next section, we will explore the best practices associated with productive use of content.

Delivery of content

We have covered a lot of ground so far, but none as important to the people in your organization as the delivery of useful content. Content is useful when it is delivered to you and others in a familiar and consistent manner. The beauty of consistency is getting what you expect, whether it's your favorite restaurant, a solid relationship, or relevant content. When you don't get the experience you expect, especially if it is foreign to your common practices, whether it be in daily life or in business, you're going to shy away from that experience.

In this section, we will build on the Capture and Manage aspects of ECM by emphasizing a strong method for delivering content to your users. You want this to be a fast and effective experience for

them, getting them the content they are looking for in a consistent manner. This is very important for user adoption, which is the ultimate factor in achieving success for your SharePoint ECM solution. Searching, finding, and consuming content is where organizations get beyond all the parts of ECM meant solely for the input and management of content and into actually using it for daily activities. It is where the mass of unstructured content will begin to meet all the back office line-of-business systems and processes used in daily operations. We recommend that you adhere to the following three elements:

- 1. Consistency** Leverage your IA.
- 2. Focus** For example, focus on the rule and not the exception.
- 3. Users** Involve them early and often.

Consistency

The benefits of being consistent in your SharePoint deployment have been made clear throughout the book regarding consistency in how sites and IA are set up per department and consistency of content types and libraries. The same is true when you think about the ways you will put content in the hands of users for editing and consuming. The benefits to remaining consistent will be measured in the ease of maintenance, decreased help desk support, happy users, and the ability to expand the ECM solution in repeatable ways that can truly benefit the organization. This will happen by planning for and incorporating consistent file formats based on need, consistent views and viewing, search functions, and web app updates.

From the field

During the final go-live week of a recent project, the document preview feature for SharePoint was identified as useful by a key stakeholder but it had not been included in the initial design for document library searching. This required Office Web Apps and some additional configuration. The hands-on user training had not included this new feature, but it was included in the final rollout. The help desk was immediately answering questions related to the install of the plugin required for preview; some users liked it and others didn't. Ultimately, the feature was turned off to reduce remedial training efforts. - Shad

Consistency of updates means that when you change the system you change it in repeatable and expected ways. End users fear change; this is a truth that will not go away. You should avoid changing the system frequently or in ways that alter the familiarity or steps required to interact with your SharePoint ECM solution. When you make changes, they should be regular periodic small updates, and any large changes should involve a prior notification. A fair amount of selling the change should be done with the users long before the change is made. Most updates to a farm that are visible to the user relate to consumption of content, which is why change is considered as a topic here.

From the field

I was on a short duration project recently, and on the first day of the project we spent a lot of time on library views and how beneficial they are, how they can replace a folder, and so on. What I did not do was tell my client at the same time how dangerous they were. The morning of day two I went into the office of an excited client. He wanted to show me what he had created the night before. To my horror, he showed me several libraries where he created not one, not two, but no less than seven views per library. They were all great views, but not a single user would know what they are for. When I asked him, "Did you do this on the production farm?" He enthusiastically said yes. This was a rude awakening to some users, who found new and unfamiliar ways of accessing content. - Chris

To interact with content, users have to access it first. Access to content happens in two ways: by browsing or by search. Unfortunately, we find that the current trend is "throw it in the bucket and search for it." In a traditional line-of-business operation, this is not extremely effective. It is ineffective because search is the most subjective way of accessing content. In search, the burden is on the user to know the proper terms and format of a search query to get the document they seek. We would like to pose this question for you: If a user knows this much about the document, should they need to search? It's rare to find documents with enough content, or content in the right place, for search to be a highly effective method of accessing content. A single search is usually a 50/50 event, but after multiple searches, the user gets the feel of results and can dial in the results they are looking for much faster.

Therefore, we encourage organizations to consider search as the alternative when browsing does not work, rather than as the initial approach. In content browsing, a user drills down into the document they know they need, and because you have designed a good IA, they will get there very quickly. Indirectly, the more users need to search, the greater indication of poor IA planning.

Browsing and navigation

Browsing for content starts at the web application level. Users have to first identify where the document repository content lives and then the logical location of that content. Fortunately, we have explained why good IA is going to help you make sure that users spend most of their time in a single web application, making the need for extended navigation irrelevant.

The three main ways to navigate content are as follows:

1. Top link bar
2. Quick launch
3. Tree view

The first two are very common, and the last more common than it should be, because it implies problems with IA.

The top link bar, as shown in Figure 3-5, is the navigation at the top of the site. This navigation is generally reserved for physical repositories or subsites. It can be customized to link to any location via URL, but we recommend isolating to other web applications on the farm and subsites within the current site collection. You might also consider linking here to common resources that every employee in the organization can benefit from. As stated earlier, the key here is consistency. This is true for all navigation. Be consistent in how you name and order headings in navigation. Also be consistent about what the headings are. Navigation methods that are regularly changing will result in users finding another way to browse for content.

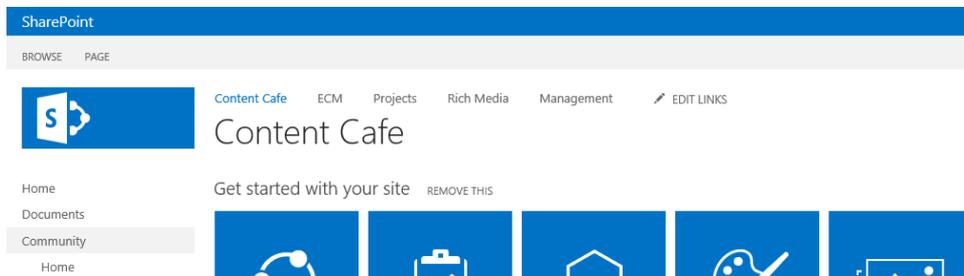


FIGURE 3-5 Top link bar.

The quick launch pane is on the left side of the browser. It is typically used to show all lists and libraries. Our recommendation is that quick launch be used *only* for listing libraries and lists in ECM sites, with the exception of cross-functional site links such as extranet/intranet quick access—that is, locations that you know a user will frequent. You can do this by creating a new navigation link, as shown in Figure 3-6. Navigation links can link to locations in the farm or locations outside the farm that are accessible and referenced by a URL.

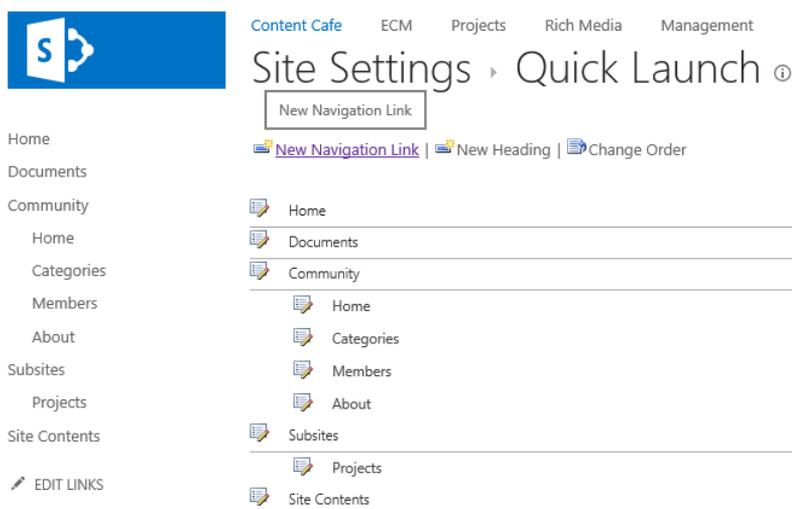


FIGURE 3-6 Quick launch links.

All the links in the quick launch are grouped by headings. Usually, headings are limited to Home, Libraries, Lists, Tasks, and Calendar items. In ECM, we generally do not see the Task or Calendar item in sites; these are reserved for the team sites discussed previously in the context of the Manage section, so these headings are often removed in favor of just Libraries and Lists. Because headings are also linkable, it's possible to have the headings themselves be the logical storage—for example, the “docs” library instead of “docs” being under “libraries.” The benefit of this is streamlined navigation, but it becomes troublesome if you mix varied types such as lists and libraries. The primary recommendation, as stated earlier, is be consistent and keep your navigation simple. As a best-practices approach to simplicity, we recommend that your SharePoint ECM solution include only headings that link to libraries such as “Documents,” “Rich Media,” and “Email.”

Tree view is an additional feature that can be enabled on SharePoint 2013 and is a way to visualize the relationship between logical and physical storage to its parent logical and physical storage. Essentially, when you turn tree view on, you get a new web part, as shown in Figure 3-7.

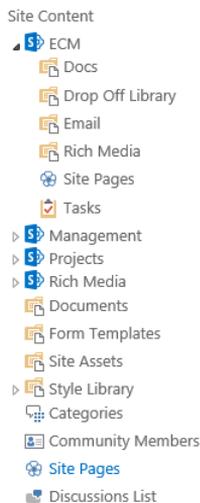


FIGURE 3-7 Tree view.

Another huge aspect of content view is library views, another tool that is either used not enough or in excess.

Many organizations are not aware of the extent that views in SharePoint can be modified. You can change columns by removing and adding which are visible. You can group content based on metadata, and you can create views that essentially filter content. Organizations that tend to use this in excess will have upwards of 10 views for any one library. This is an indication of bad IA. Organizations that are using the out-of-the-box views are using the tool too little and not leveraging the power of sorting by columns and filtering. These organizations will tend to be the same ones that use folders to organize their content.

There is no magic bullet on views, but there are a few guidelines. Every column shown should provide value to the user in the following ways:

1. If a column is useful only for some automation process or special type of user, it should be hidden or a part of a special view just for those types of users.
2. Columns should never exceed the right side of the page on a typical browser. Columns that are not visible in a single window are not usable.
3. Grouping is very powerful, but use this feature with caution and provide adequate user training as to the purpose of grouping. Grouping is used to aid a user who is browsing for content, while lists are used for more robust processes.
4. Limit the types of users who can create views, and make views consistent in all similar libraries across the entire SharePoint farm.

Site contents

Next on the navigation list is the viewing of site contents. In SharePoint 2010, this is called "View All content," and in SharePoint 2013, it's called "Site Contents." Asking one to "Click View Contents" is a common phrase, but we feel a crutch. It should not be required for the typical ECM user to navigate to lists/libraries or contents that are not otherwise accessible via the quick launch. Typically, if this is being used, it's a training problem, usually the result of someone far more experienced with SharePoint teaching how they use SharePoint, instead of how an end user should be finding and viewing content. Unfortunately, it is very common for a SharePoint site to have so many places to navigate from that an entire separate window to show all the contents is required. Both situations are not sustainable for widespread user adoption or simple browsing. We recommend keeping this tool a little secret, to be used by your super users only.

When IA fails or bad content somehow gets added to the system, the only option for accessing that content is search. Full-text search enables an index to be created from content stored in the body of a document. The goal is to get the user to the content with as few clicks as possible. With good consistency of content across the farm stored in a well-designed IA, a user is more familiar with content, and users can find that content based on keywords, where it's located in the farm, and other metadata searching.

Search

We have all experienced light and overloaded search results or getting too few results or too many. To eliminate the problem of having too few results, you want to make sure that indexing across the farm is adequate and that all the content in the farm has the appropriate iFilter activated so that it can be indexed. We can eliminate the problem of having too many results by filtering results using Boolean logic search terms, if/and/else/equal to, or search refiners.



Note Now search in SharePoint has evolved to include Boolean logic. For example, I can make a query for documents where the title starts with SharePoint but has the word ECM in the body. It would look something like this: ("title:SharePoint*" AND "ECM").

Because Boolean searching is not a common tool, something more is needed. Users are often in too much of a hurry or not advanced enough to use Boolean logic. One way you can help them get to their results faster is by using search refiners.

Search refiners are generated automatically based on a document's structured metadata. When you do a search for documents that start with "SharePoint" and contain the word "ECM" and you get too many results, you can refine your results by knowing simple things about the document, such as the author or when it was created.

Refiners can be taken even further, by including keywords and tags. Specific terms from the body of the document can also be generated on the left portion of a search panel, as well as their hierarchy. There is some customization required to get this feature to work, but it is one that is well worth the effort.



Note Remember that structured metadata is metadata that is generated by the system, such as author and modified date.

Farm administrators can do even more to assist the user in finding content. They can do things like *visual best bets*. This is where a particular document is highlighted at the top of a search based on a keyword. The relationship between the keyword and the best bet has to be established manually. This helps users know what is the most important document for that search term.

The trick with the best bets option is keeping it current and relevant. This usually requires an administrator specializing exclusively in search and who has the relevant operational domain knowledge or knows how to obtain that knowledge within the organization.

Another common feature is linking content to individuals by including people search in the search results. By default, this appears on the right side of the search results and is effective only if you have active MySite pages for your users.

Viewing

After a user has browsed to or searched for a document, they need to have access to view and edit the content.

In SharePoint, there are two options for viewing content: either via the client application or via the browser. Surprisingly, the answer as to which to use is highly dependent on the licenses your organization owns and the makeup of your users' desktops. For example, if your users have Microsoft Office installed on their PCs, viewing via client application might be very convenient. If you have users accessing content on mobile devices, you need to spend time on a great browser viewing experience.

It also depends on your SharePoint licenses. To get the most robust experience with in-browser viewing and editing, the organization should invest in licenses to the Microsoft Web App server. This is a separate SKU and server on the farm that manages the real-time editing of documents in a browser. This tool allows both mobile users and PC users to have 80 percent of what is available in any client application for office documents.

To help ease your decisions, we recommend being consistent. Supporting various types of clients can be difficult. For example, if your organization is dominated by local network PCs with Microsoft Office installed, it's recommended to make the "open in client" application the default use case. Additionally, this will help if you have other file formats that are not a part of the Microsoft Suite.

If you can purchase licenses to add Microsoft Web Apps Server, we recommend that you standardize on editing and viewing in the browser, to the extent where you remove Microsoft Office from the users' desktops or laptop PCs. For unique circumstances for some users, you might allow the client integration with the local version of Office.



Note Separate the readers from the creators. In many organizations, those who create the content and those who read the content are not the same. It is possible to consider different modes of operation for each. For example, readers should be able to work with all in-browser viewing of content, while creators might need Office client applications.

Finally, there are third-party viewing tools available that cross a spectrum of file formats. When your users are doing a lot of read-only access to documents, these tools might be useful.

Preservation

Preservation of content is an aspect of delivery, because typically preservation is a process of document conversion and movement to another location.

Preservation focuses on content that has historical importance to your organization. The types of content can be diverse and vary depending on the type of organization, enterprise governance, and any regulatory considerations you might have to contend with. The breadth and depth of preservation is very noticeable in the government arena, where many documents have historical importance and, in some cases, need to be kept forever.

For example, the documents associated with city planning could have a life cycle that requires them to be kept forever, to retain the historical significance of when, where, and how land use decisions were made. It's key to understand where SharePoint ECM begins and should end in terms of active content and archiving.

In the case of historical documents, the importance of managing a content life cycle does not generally have an impact for a SharePoint ECM solution, because many of the historical documents should be stored either on permanent media or on microform.



Note We recommend that you focus your SharePoint ECM solution on active content. This is content that is being created, reviewed, and shared and that has value to the organization at large for completing a job function or transaction. We do not recommend that you use SharePoint as an archive or long-term preservation solution.

Even if preserved content is still a part of SharePoint, it's recommended it be moved to a completely different SharePoint farm or, at minimum, to a site collection. The reason for this is that SharePoint is built for high activity and interaction from users, whereas preservation is about content that is interacted with very rarely or only for research purposes.

If you are keeping this content in the active SharePoint environment, there is a large opportunity cost that is taking valuable resources away from other high-value content. It is taking up space from your primary storage that should be dedicated to active content. Also, it could be impacting the performance of the farm, but most importantly, it's accessible by users who might or might not know the difference between preserved content and active content.

We recommend the following steps when determining how content moves from the active portion of its life cycle to the preservation stage:

1. Determine where the preserved content will live. It is assumed that as a part of the process of preservation the content will move. If you are saying to yourself that preserved content will simply be placed in a separate library in the same location as active content, this alone is not considered preservation.
2. Determine the long-term preservable format that the content requires.
3. Determine the automated or manual process that will move and convert the content.
4. Determine a custodian for preserved content. This will be an individual or external organization.
5. Determine the format of metadata for preserved content, and include in that metadata the last location where the content was saved in SharePoint. This is important to establish the chain of custody.

If you choose SharePoint as the final location for preserved content, consider methods of remote blob storage (RBS) for the preserved content. RBS is a database configuration that allows certain content to live outside of the SharePoint database while maintaining its accessibility through SharePoint. This allows you to keep the content in the SharePoint user interface without the load on the database. Using RBS is not recommended for active content but only for site collections where you have moved or are referencing preserved content.

It is a great practice for organizations to start either preserving or offloading inactive content, or disposing of it according to a predetermined retention schedule. This ensures the efficiency of the

system and prevents users from being overloaded with content that is no longer relevant. This should be done in accordance with the following activities performed by an experienced Records Manager:

1. Create a records inventory to determine the breadth and depth of all your organizations records, both physical and electronic.
2. Create a retention schedule that clearly outlines the definitions of content, based on several parameters that are relevant to your organization. At a minimum, this usually includes document type, records series, retention period, active date, and inactive date.

Legal Holds can also be established for items identified during litigation discovery processes. These are known as *Interrogatories* and *Requests for Production (ROGs)*. In this case, content objects can be placed on hold with a timer. If they are touched or further requests are made before expiration of the timer, it can be reset accordingly.

3. Create a records retention policy that clearly outlines the procedures and governance of records for your organization. This usually includes the review and blessing of your legal department and executive branch.

After these activities have been performed, you can use the information in the retention schedule to configure SharePoint to support the preservation of content at the appropriate stage of its life cycle.

You have now completed the reading necessary to understand all the series of stages that content traverses during its life cycle. We began with defining Enterprise Content Management in our introduction. We then walked you through each of the major steps required for a SharePoint ECM solution. To recap, we have covered the following areas:

- **Getting content in, or Capture** Upload, MS Office, scanning, native documents, forms, and streams
- **Configuring site(s) collections and document libraries, or Store** IA, versioning, formats, and transformation
- **Moving content from person to person, or Process** Business process management, workflow, business intelligence, and eDiscovery
- **Finding and collaborating on content, or Manage & Deliver** Navigation, editing, viewing, searching, and preserving

Next steps

We will now outline two different strategies for using your newly acquired ECM knowledge. First we will look at a small or departmental deployment of SharePoint ECM, and then we will make the necessary adjustments for a much larger scale project. We will be providing step-by-step instructions and giving you some nice outlines that you can use for documenting your IA.

Index

Symbols

#SPHelp, 260

A

ABA Rule 1.1 of Professional Conduct, 224
accessibility
 to content, 83
 with Office 365, 247
access levels, 18–19
accountability in records management, 191
accuracy of documentation, 165
active content, vs. historical documents, 89
Active Directory, 73
Add & Manage Sources dialog box, 233–234
ad hoc capture, 11
administration functions, access to, 76
administrator permissions, 106
Administrators, for site collection, 117
administrators recycle bin, 77
advisory governance, 181
AIIIM, 262
analysis paralysis, team size and, 137
application usage policies, example, 188
approval workflows, 23
archive approach to site disposition, 213
archive content, 26
archive tools, 248
ARMA, 262
as-is state, 61
AT (after termination) date, 112
audit documents, location for, 218
audit tracking, 225
authentication, Single Sign-On for, 40
automation, for content load into SharePoint, 9
availability, in records management, 192

B

bandwidth, for Office 365, 247–248
bargained governance, 181
bargained hard line setting, for site disposition, 213
behavior, encouraging user adoption, 152–160
best bets, visual, 87
best practices, documenting, 164
BigData, 23–24, 249
black and white images, vs. color, 45–46
Boolean logic, 21, 86
born-digital method of capture, 9
branding, 156–158
browser
 creating documents in, 10, 42
 for viewing content, 21, 87
browsing
 and navigation, 83–86
 vs. search, 83
budget, for change management, 156
bulk load content, 8
bulk operations, user interface limitations on, 9
business analyst, 4
business drivers
 documenting, 163
 for records management, 193–195
Business Intelligence (BI), 23–24, 68, 249
Business Process Management (BPM), 23, 249–250
Business Productivity Online Standard (BPOS), 244
business unit, and ECM, 134

C

Capture stage in ECM stack, 5, 7–12
 best practices, 36–47
 content streams, 11–13
 document scanning, 11, 43–47

case in eDiscovery

- electronic form capture, 10–11, 43
- file upload, 8–9, 36–39
- native SharePoint documents, 10, 41–43
- Office documents, 9–10, 39–41
- case in eDiscovery, 230
 - home page for new, 232
 - multiple eDiscovery sets per, 236
 - site home page, 240
- case study, managed metadata, 96–109
- central administration isolated, Office 365 and, 245
- Central Administration dashboard, 92
 - Manage Services Applications, 103
 - Manage Web Applications page, 130
- centralized governance, 181
- chain of custody, 57, 89, 225
- change
 - end users fear of, 82
 - resistance to, 70
 - support for adoption, 145
 - to taxonomy, 177
- change control, 20
- change management, 144
- change manager, 156
- change request documents, 142
- check-in | check-out document management, 40, 79–80
- child term, in taxonomy, 98
- city, Information Architecture for small, 169–171
- CL (close, or completion), 112
- client-side applications
 - removing from desktops, 43
 - for viewing content, 87
- cloud
 - quantity of data, 227
 - security for, 246–247
 - service providers, 246
 - vs. on-premise, 243
- CloudShare, 259, 260
- Coho Winery case study
 - characteristics, 92
 - ECM solution, 149
- collaboration, 3
- color images, vs. black and white, 45–46
- colors, branding with, 157
- columns in views, guidelines for, 86
- comma-separated value (CSV) files. *See* .csv (comma separated values) format
- Common law duty, 224
- communication by ECM team, 141
- communication gap, 5
- community of users, 154–155
- compatibility, 41
- completion, percentage of, 158
- compliance
 - in records management, 192
 - with regulations, 29
- Compliance Details dialog box, 202–203
- compressed files, 15, 26
- conditional formatting, 65
- conduct statement, 189
- confidential library, 168
- confidential records, 228
- configuration blueprint for documentation, 164
- Configure Send To Connections link, 208
- Connection Permissions for ECM MMS dialog, 105
- consistency, 82–83
 - in keywords, 159
 - in navigation, 84
 - in site collection, 168
- content
 - authoring, 42
 - delivery, 81–88
 - duplication of, 31
 - eDiscovery for isolating, 223
 - exporting for eDiscovery, 237–240
 - preservation, 88–90
 - structured vs. unstructured, 30
- content actualization, 24
- content audits, 151, 193, 202
 - report options, 204
 - responsibility for, 218
- content custodian, 193
- content databases, 12, 26
 - determining number of, 51
 - estimating initial number, 48
 - maximum content, 91
 - multiple, 168–169
 - size recommendations, 47
 - for web applications, 52
- Content Deployment paths, 207
- content enrichment, third-party tools, 250
- content governance, 180–189
 - documenting, 185
 - progression for, 181
- content life cycle
 - end, and content deletion, 31
 - managing, 88
- content management, 69–81
 - repository, 73–77

- responsibility for, 71–73
- security, 73
- site administration and recycling, 76–77
- team site, 75–77
- Content Management Interoperability Services (CMIS), 252
- content managers, 72, 134, 225
- content organizers, 144, 205–207
 - automating notifications with, 63
 - workflows vs., 207
- content posting and editing policy, example, 187–188
- Content Posting Cycle, 187
- content routing, 63
- content streams, 11–13
- content type hub, 173
 - order of implementing, 111
 - retention schedule and, 110
 - service connection linked to, 106
- content types, 32, 109–117
 - automatic application of, 250
 - creating, 114–117
 - defining, 37
 - documenting, 172–176
 - documenting steps to create, 164
 - governance for, 187
 - for metadata, 55
 - on-premise SharePoint vs. Office 365, 245
 - publishing, 56, 116
 - settings, 121
- Content Type Settings page, 115
- conversion, 17
- copiers, digital, 45
- costs
 - of deploying ECM, 4
 - of eDiscovery, reducing, 222
 - of governance, 184
- Create New Managed Metadata Service dialog, 104
- cross-functional projects, 75
- crowd sourcing, 152
- .csv (comma separated values) format, 97–102
 - for taxonomy, 177
 - functional taxonomy outline, 100–101
 - period taxonomy outline, 99
 - region taxonomy outline, 99
- custom content types
 - adding to library, 121
 - documenting, 172, 175
- custom solutions

- determining need for, 57
- potential impact, 80
- custom workflows, complications from, 67
- CYE (current year end), 112

D

- dashboard, 24
 - for reports on user adoption, 153
- database. *See also* content databases
 - naming convention for, 93
 - SQL, 47
- data mining, 24
- deactivating records management, 200–203
- decisions, information for, 29–30
- declaring document as record, 201–204, 218
- deleting information, 214
- Deliver stage in ECM stack, 5, 20–22, 81–88
 - editing and viewing, 21–22
 - publishing, 22
 - search, 21
- departmental/distributed capture, 11
- department managers, in ECM project, 135
- deploying ECM, costs of, 4
- deployment
 - assumptions, 91–96
 - large-scale, 129–131
 - small-scale, 125–129
- desktops
 - client authoring tools, 42
 - removing client-side applications from, 43
- diagram of information architecture, 167
- digital copiers, 45
- disaster recovery, 249
- discoverable content, 228
- discovery, 193. *See also* eDiscovery
 - procedures validation in advance, 224
- disposition library, 123–125
- Disposition Library Send To Connections page, 124
- disposition of records, 192, 202–205
- disposition workflow, 63–64
- distributed document scanning, 45
- documentation, 163–166
 - configuration blueprint, 164
 - of Content Organizer, 207
 - of content types, 172–176
 - Information Architecture, 166–180
 - as never-ending, 165–166
 - of content governance, 185

Document base content type

- peer review of, 165
- purposes, 163
- quick tips for, 165
- roles and responsibilities in, 72
- of taxonomy, 177–180
- Document base content type, 56
- Document Center template, 38, 40, 95
- Document content type, 112
- Document ID, 57–58, 191
- Document ID Service, 120
- document imaging, 253
- document libraries, 32, 168
 - adding, 119
- document life cycle, 88. *See* ECM Stack
- document management, 79–81
- document readers, vs. creators, 88
- documents
 - declaring records as, 218
 - dragging between locations, 207
 - links to, 41
 - user actions on, 77–79
- document scanning, 11, 43–47
- dots-per-inch (dpi), for scan setting, 46
- “downstream” effects of document life cycle
 - stages, 6
- draft item security, 81
- dragging documents
 - between locations, 207
 - into SharePoint, 38–39
- Drop Off library, 123–125
 - creating rules for, 205–206
 - URL for, 124
- duplicate records, content audit and, 151
- duplication of content, 31

E

- early case assessment, 222
- ECM. *See* Enterprise Content Management (ECM)
- ECM Committee, 72
- ECM environment, recommended libraries, 54
- ECM hub
 - accessibility, 93
 - site collection, 168
- ECM_MMS service
 - administrator permissions for, 106
 - user permissions for, 105
- ECM solution
 - core feature sets, 2

- order of operation for deploying, 96
- success vs. failure, 161
- ECM stack, 5–7
 - Capture stage, 7–12
 - Deliver stage, 20–22
 - Manage stage, 17–20
 - Preserve stage, 25–26
 - Process stage, 22–25
 - Store stage, 12–17
- ECM team
 - communication, 141
 - culture, 140–141
 - importance, 133
 - org chart, 140
 - participants, 135
 - practitioner as implementer, 146–147
 - pre-mortem, 145–146
 - selection, 137–139
 - time and conflict, 135–136
 - transition, 152
- ECM team roles and responsibilities, 139–147
 - quality control, 144–145
 - subject matter expert, 142–143
 - technical team, 143–144
- eDiscovery, 25, 29, 210, 221, 263
 - cost reduction, 222
 - exporting content, 237–240
 - holds and, 221–227
 - implementing in SharePoint, 228–241
 - In-Place Hold option, 235
 - isolating content, 223
 - litigation support, 223–225
 - notification of content owners and consumers, 240–241
 - Office 365, 227
 - process, 232–233
 - reasons for using, 222
 - risks in not understanding, 227
 - site collection for, 229–230
 - stages, 225–227
- eDiscovery Center, 229
 - home page, 231
- eDiscovery Download Manager, 238–239
- eDiscovery set creating, 233
- eDiscovery template, 230
 - custom lists in, 232
- editing, 21–22
- e-forms, 10
- 80/20 rule, 67

- Electronic Data Reference Model (EDRM) load file format, 238
 - electronic discovery. *See* eDiscovery
 - electronic form capture, 10–11, 43
 - email-based libraries, 158
 - email-enabled lists, 158
 - email web application, 171
 - employee files, libraries for, 168
 - empowered governance, 181
 - empowered owner approach to site disposition, 213
 - end-user recycle bin, 77
 - Enterprise CAL with the Performance Point features, 249
 - Enterprise Content Management (ECM), 1–5, 262
 - catalyst for, 133
 - enforcing plan, 160
 - reasons for SharePoint failure, 4
 - reasons to use, 27–29
 - SharePoint counter features, 158–159
 - target of, 31–32
 - with vs. without records management, 215
 - Enterprise Edition SKU of SharePoint, 124
 - Excel spreadsheet, 24
 - for audit data, 204
 - for taxonomy structure, 97
 - Exchange, 32
 - exclude logical operators, 236
 - executive summary
 - in governance plan, 181, 186
 - in documentation, 165
 - expectations, 32
 - experts, 155
 - expired content, 25
 - determining action for, 63
 - Explorer view, 9, 158
 - exporting content for eDiscovery, 237–240
 - exports library, history of all exports, 239
 - external applications, 255
 - External BLOB Storage (EBS), 49
 - external capture applications, 44
 - extranets, 22
- F**
- farms
 - multitenant, 243
 - performance, and user experience of Send To locations, 210
 - Favicons, branding with, 157
 - feature sprawl, 146
 - File Explorer view, uploading files from, 38
 - file formats
 - best practices for, 16
 - policy example, 188
 - for scanning, 46
 - file naming conventions, 37
 - file upload, 8–9
 - best practices, 36–39
 - filter
 - for eDiscovery query results, 235
 - for search results, 86
 - findability, 27
 - flat hierarchy for sites, 53
 - flexible content structure, 30
 - flexible owner approach to site disposition, 213
 - flowcharting in Visio, 66
 - folders
 - as metadata, 14
 - structure, 152
 - structure as taxonomy, 59
 - folksonomy, 50, 58–61, 180
 - implementing, 60
 - user generation, 97
 - formatting, conditional, 65
 - form design, 11
 - FRCP Rules, 224
 - full-text search, 86
 - functional taxonomy, 60, 96
 - construction process, 101–102
 - outline .csv file, 100–101
 - principles for building, 100
- G**
- gamification, 158
 - governance, 17, 32, 164. *See also* content governance
 - bargained, 181
 - documents, 187–189
 - features and application relationships, 185
 - goals of plan, 183
 - importance of, 70
 - in legal discovery, 227
 - management of, 217
 - plan implementation, 183
 - plan outline, 186–187
 - third-party tools, 251, 255
 - graphics, branding with, 157

groups

- groups, 74, 77, 108
 - adding users, 118–119
 - in views, 86
 - security level for, 118

H

- habits, and user adoption, 158
- hard copies, converting to digital, 45
- hardware drivers, 44
- hash tags, 260
- headings, in quick launch, 85
- hidden columns in views, 86
- historical documents, 88
- holds, 193, 210, 221–227
 - content access during, 228
 - library for, 168
 - notification before setting, 241
- hosting vendors, 244
- hub for content types, 56
- hybrid approach for record center, 216–217

I

- iFilter, 16
- image capture, 43
- Image content types, 112
- image quality, from production scanning, 45–46
- implementation, ECM team members and, 138
- Import Terms, for group, 108
- index, 86
- index crawler, 47
- InfoPath, 10, 43
- information
 - for decisions, 29–30
 - gathering by team members, 138
 - management policies, 202–205
 - options for organizing unstructured, 2
 - organization of, 51
 - rapid growth in, 48
 - value of, 30
- Information Architect, 72
 - vs. business analyst, 4
 - qualities, 5
- Information Architecture (IA), 12–15, 49–51
 - application of, 2
 - beginning of document, 171
 - documentation, 166–180
 - importance of, 70

- shared, 117–125
 - SharePoint components, 50
- inheritance, 75
- in-place content holds, vs. moving contents, 228–229
- in-place records declaration, 198–201
- in-place records management, records center
 - vs., 215–217
- integrity
 - of document catalog, 20
 - in records management, 192
- intellectual property rights, 228
- intelligent character recognition (ICR) software, 253
- internal support, lack of, 146
- Internet, security threat of, 246
- Interrogatories, 90
- intranets, 22
- isolated central administration, Office 365 and, 245
- item-level security, 75
- IT Governance Institute, 180
- IT role in ECM project, 134

K

- Key Performance Indicators, 24
- keywords, 21, 159, 250
- knowledge, 29
- knowledge workers, access for those bridging departments, 51

L

- labels, for workflow steps, 67
- landing page, 92
- language of stakeholders, 4
- large-scale deployment, 91
 - creating, 129–131
- LCID (language code), 177
 - in taxonomy, 98
- leadership, 133
 - by example, 146
- least common denominator, user as, 149–150
- legacy ECM applications, 41
- Legal Holds, 90
- legal standards, 224
- legal team, 225
- liabilities, old records as, 194
- libraries, 39, 53–55
 - adding custom content types to, 121

- documenting architecture, 166–172
 - document storage in, 36
 - in-place records management for, 200–203
 - modifying Quick Launch navigation for, 120
 - number of, 120
 - size recommendations, 54
 - views for, 83, 85–86
 - Library Record Declaration Settings dialog box, 200–203
 - Library ribbon, 121
 - licenses for Microsoft Web App server, 88
 - “Like” metadata element, 60
 - line-of-business (LOB) applications, 2, 91–92
 - integration with, 251–252
 - links to documents, 41
 - lists in SharePoint, 39
 - litigation, 32, 149
 - eDiscovery for, 25
 - eDiscovery support of, 223–225
 - reactive scenario in, 29
 - locating document, effort required, 27
 - location
 - for audit data, 204
 - for audit documents, 218
 - for content types, documenting, 172
 - dragging documents between, 207
 - for eDiscovery set, 233
 - for hold and move discovery contents, 229
 - logical storage, 12. *See also* Information Architecture (IA)
 - Lync, 32
- ## M
- mail server, discoverable content in, 228
 - major versions, 16, 80
 - Managed Business Objective (MBO), 72
 - managed metadata, 176, 188
 - managed metadata service (MMS), 107, 176
 - initial configuration challenge, 107
 - “management by the red asterisk”, 18–19
 - management portals, 51
 - Manage stage in ECM stack, 5, 17–20
 - change control, 20
 - policy, 19
 - records management, 17–18
 - security and access, 18–19
 - mandatory metadata, 18
 - manifest XML file, 238
 - matter, 221
 - held content associated with, 230
 - MCM (Microsoft Certified Master), 155
 - media files, 48
 - media library, 168
 - meetings of ECM team, 137–139
 - metadata, 6, 50, 55–57
 - content streams and, 11
 - content types and, 37, 112–113
 - for file upload, 38
 - folders as, 14
 - managed, 96–109
 - mandatory, 18
 - on-premise SharePoint vs. Office 365, 245
 - quality control for, 145
 - for records, 18
 - for scanned document, 44
 - standards from, 26
 - storage, 12
 - structured, 87
 - use of, 15
 - user actions and, 158
 - workflow use of, 23
 - metadata model, 49
 - Metadata Navigation Settings dialog, 122
 - methodology, ECM as, 5
 - metrics, on user adoption, 153
 - Microsoft Certified Master (MCM), 155
 - Microsoft Excel 2013 Web App, 42
 - Microsoft Office
 - for capture, 39–41
 - and SharePoint, 22
 - Web Apps, 42
 - Microsoft Partner community, 3
 - Microsoft SQL databases, 12. *See also* content databases
 - Microsoft Web App server, licenses for, 88
 - minor versions, 16, 80
 - MMS (managed metadata service), 107, 176
 - initial configuration challenge, 107
 - mobile authoring, tablets and smart phones for, 42
 - Most Valuable Professionals (MVPs), 155
 - multifunction peripherals, 45
 - multiple files, uploading at once, 38
 - multitenant farms, 243
 - MySites, 61, 159–160, 254

N

- names
 - for ECM project, competition for, 153
 - for eDiscovery set, 233
 - of holds, 221
 - naming conventions, 165
 - documenting, 164
 - for database, 93
 - for files, 37
 - for site names, 157
 - for workflows, 67
 - in governance document, 187
 - native SharePoint documents, 10, 41–43
 - native SharePoint solutions, 254–255
 - natural language processing (NLP), 250
 - navigation, 76
 - browsing and, 83–86
 - New Site Content Type dialog box, 114
 - New Subsite option, 127
 - non-native SharePoint solutions, 254–255
 - nonrecord documents, 194
- O**
- OASIS, 252
 - Office 365, 243–248
 - bandwidth and accessibility, 247–248
 - eDiscovery and, 227
 - file size limit, 245
 - Office documents, 9–10
 - bar code on, 58
 - creating in SharePoint, 40
 - Info portion, 40–41
 - SharePoint viewers for, 22
 - Office Web Applications (OWA), 10
 - one-to-one relationship, for security and access levels, 19
 - on-premise solution, SharePoint as, 243
 - optical character recognition (OCR), 44, 46, 253
 - organization
 - changes, 146
 - inconsistency in following rules, 225
 - preparing for user adoption, 150–152
 - Organizational Policies and Procedures, 17
 - organization of information, 51
 - Outlook, as viewer, 21

P

- paper documents, converting to digital, 45
- parent term, in taxonomy, 98
- PDF/A file format, 26, 46
- PDF file format, 15
 - for scanned documents, 46
- PDF iFilter, 47
- peer review of documentation, 165
- percentage of completion, 158
- period-of-time–based taxonomies, 96
- Period taxonomy outline .csv file, 99
- peripherals, multifunction, 45
- physical records, policy for, 219
- physical storage, 12, 47–49
- planning, 6, 35, 50, 163
 - documentation in, 165
 - for records management, 192–193
- policy, 19
- Port 80 web application, 92
- portals, 22
- PowerPivot add-in for Excel, 24
- predictive coding, 250
- presentation layer, 50
- preservation of content, 88–90
- Preserve stage in ECM stack, 25–26
- preview of eDiscovery query results, 235
- priorities, of other projects, 146
- proactive driver of ECM, 28
- procedures, 19
- processes
 - organization requirements for, 62
 - re-engineering, 31
- processing fee per gigabyte, 223
- Process stage in ECM stack, 5, 22–25, 61–68
 - best practices, 61
 - conditional formatting, 65
 - content routing, 63
 - disposition workflow, 63–64
 - three-state workflow, 64–65
 - workflow, 66–68
- production capture, 11
- production farm, development and testing separate from, 53
- production scanning, 45–47
- productivity, content access and, 151
- profile information, 158
- progress report, on implemented changes, 152
- Project Management Book of Knowledge (PMBOK), 141–142

project management, by ECM team, 141–142
 projects, cross-functional, 75
 protection, in records management, 192–219
 publishing, 22, 81
 purge date, for recycle bin, 76
 putability, 27

Q

quality control, 144–145
 query, 21
 creating for content export, 237–238
 for eDiscovery search, 234
 quick launch pane, 84

R

“Rating” metadata element, 60
 ratings, 159
 RBS. *See* Remote BLOB (Binary Large Objects) Storage (RBS)
 reactive driver of ECM, 29
 records, 18, 191
 confidential, 228
 content preserved as, 226
 declaring document as, 218
 disposition of, 202–205
 duplication, content audit and, 151
 records center, 18, 32, 198–219
 vs. in-place records management, 215–217
 records inventory, 48–49, 90, 193
 records life cycle, 192–219
 records management, 17–18, 188, 263
 business drivers for, 193–195
 deactivating, 200–203
 ECM with or without, 215
 factors impacting, 196
 planning for, 192–193
 principles and life cycle, 191–193
 retention schedule, 195–198
 SharePoint features, 198–215
 SharePoint processes, 217–219
 third-party tools, 252
 records management plan, 17
 records manager, 72, 143, 194, 217, 225
 access to content, 218
 activities of, 90
 Records Policy, 110
 records retention policy, 90

recycle bin in SharePoint, 76
 red asterisk, management by, 18–19
 refiners, 21, 86–87
 reformatting content for preservation, 26
 regional taxonomy, 60, 96
 region taxonomy outline .csv file, 99
 relationships of content types, 173
 Remote BLOB (Binary Large Objects) Storage (RBS), 26, 48–49, 89
 on-premise SharePoint vs. Office 365, 245
 third party tools, 250–251
 repositories, 49
 determining factor for creating, 51
 for Information Architecture, 13
 security for, 73–77
 Requests for Production (ROGs), 90
 resolution, for scan setting, 46
 resources library, 168
 resources, manager objection to committing, 135
 Retention File Plan, 110
 retention, in records management, 192–219
 retention periods, and content types, 111
 retention schedule, 18, 193, 195–198
 and content type hub, 110
 example, 197
 terminology, 111–112
 return on investment (ROI), 31
 for business analyst, 4
 risk
 from organization documentation, 194
 in not understanding eDiscovery, 227
 roles, tying security to, 51
 root site collection, creating for web applications, 95
 root site, in site collection, 52
 root web applications, 167

S

scanner drivers, 44
 scanning documents, 11, 43–47
 scope creep, 144
 scope of project, 142
 screen shot, in documentation, 164
 search, 21, 83, 86
 Search site collection, 168
 security, 18–19, 51, 73, 151
 draft item, 81
 ECM principles for, 74
 for exported content, 239

security groups

- for form capture, 11
- in cloud, 246–247
- need for simplicity, 78
- repository-level, 73–77
- retention schedule criteria, 196
- for site collections, 118
- third-party tools, 251
- tying roles to, 51
- security groups, 77
- security trimming, 19
- selection of ECM team, 137–139
- Send To Connection Settings page, 209
- Send To location, 207
 - set up, 208–209
- service-level agreement (SLA), 247
- Seventh Circuit Principles Relating to the Discovery
 - of Electronically Stored Information,
Principle 1.02, 224
- shared drives, organization, 1–34, 71
- shared Information Architecture (IA), 117–125
- shared network drives, 48
- SharePoint events, 261
- SharePoint, 1–34
 - as Save location for Office documents, 41
 - automation for content load, 9
 - documenting reasons for implementing, 163
 - eDiscovery implementation, 228–241
 - Enterprise Edition SKU, 124
 - evolution, 2
 - Information Architecture components, 50
 - as IT responsibility, 71
 - mapping processes to, 62
 - native documents, 10, 41–43
 - navigating, 76
 - as Office 365 component, 244
 - opinions on, 134
 - organization implementation, 69
 - planning for, 35–36
 - as platform, 262
 - reasons for ECM failure in, 4
 - records management features, 198–215
 - records management processes in, 217–219
 - uploading document to, 37–39
 - user interface limitation on bulk operations, 9
 - viewers for Office documents, 22
- SharePoint 2007 MOSS ECM functionality in, 3
- SharePoint 2010
 - functionality, 3
 - limitations in eDiscovery, 229
 - user interfaces, 119
- SharePoint 2013, 3
 - on-premise vs. Office 365, 245
 - user interfaces, 119
- SharePoint Administrator, 143
- SharePoint community, 154, 259–262
- SharePoint Designer, for visual representation, 66
- SharePoint ECM solution, 32
- SharePoint experts, 261
- SharePoint index crawler, 47
- SharePoint partner community, 32
- SharePoint Saturdays, 155, 261
- SharePoint timer job, for audit log details, 205
- SharePoint User Groups (SPUGs), 155, 261
- single-server farm configuration on CloudShare, 260
- single sign-on, 40
- site, 53
 - management, 188
 - top link bar for navigating, 84
 - viewing contents, 86
- Site Administration, Site Closure And Deletion, 214
- site administrator, 76
- site architecture, documenting, 166–172
- site collections, 32, 52–53, 74, 168
 - best practices, 123
 - creating, 117
 - creating eDiscovery set per, 236
 - Document library, 122
 - for eDiscovery, 229–230
 - landing page, 120
 - root, creating for web applications, 95
 - site permissions for, 118
- Site Content Types – Edit Policy page, 116
- site disposition, 210
 - decision process for settings, 212–213
- site permissions, for site collection, 118
- site policies, 210–215, 218
- site provisioning, 188
- Site Settings – Site Features page, 123
- small-scale deployment, 91, 125–129
 - site structure, 126
- smart phones, for mobile authoring, 42
- social media, 155
 - third-party tools, 253–254
- sprawl, 214
- SQL databases, 47
- SQL server, 249
- stakeholders, 133
 - language of, 4
 - and taxonomy constraints, 102
- stamping document in SharePoint, 191–219

standardization methods for capturing, naming and storing content, 2

storage

- calculating growth, 48
- usage policies example on quotas, 188

Store stage in ECM stack, 5, 12–17, 47–61. *See also* metadata

- document ID, 57–58
- Information Architecture, 13–15
- libraries, 53–55
- physical storage, 47–49
- taxonomy and folksonomy, 58–61
- transformation, 16–17
- versioning, 16

strategy, 35

structured content

- lists for, 53
- vs. unstructured, 30

structured metadata, 55, 87

subject matter expert, 142–143

subsites

- for content marked for hold, 229
- creating, 126–129

super users, 72, 142–143, 154

synonyms in taxonomies, 60

system design flaws, responsibility for, 144

systems integrators, 256–257

T

tables, for IA documentation, 169

tablets, for mobile authoring, 42

taxonomy, 50, 58–61

- creating, 96–109
- design, 97
- documenting, 177–180
- guidelines, 59
- keywords and, 159
- outline for .csv file, 98
- rules, 178
- steps for creating terms, 178–179

team site, 3, 75

Team Site template, 38, 95

technical team, 143–144

technology organization response to, 69

templates, 53

term sets, size of, 108

terms stores, creating, 103–107

Term Store Management Tool, 108

text files, 15

themes, 157

third-party tools, 248–257

- for backup and recovery, 248–249
- business intelligence, 249
- business process management, 249–250
- content enrichment, 250
- document imaging, 253
- evaluating, 159
- general considerations, 254–256
- questions for vendors, 251
- records management, 252
- Remote BLOB Storage (RBS), 250–251
- social media, 253–254

three-state workflow, 64–65

TIFF Group 4, 46

time-based taxonomy, 60

time to value, 137

to-be state, 61

tools, 259–262. *See also* third-party tools

- CloudShare, 260
- SharePoint community, 260–262

top link bar, 84

tracking documents, 120

training, 149, 156

Training Manager, 143

transformation, 16–17

transparency, in records management, 192–219

tree view, 85

Twitter, 155

Twitter hash tag, 260

U

unstructured content

- libraries for, 53
- vs. structured, 30

unstructured information, options for organizing, 2

unstructured metadata, 55

updates, consistency of, 82

upgrades, third-party tools and, 256

usability issues, 144

usage policies, 193

- for content types, 173

usage quality control, 145

user access, usage policies example on, 188

user adoption, 241

- actions to gain, 150

user-driven Business Intelligence

- encouraging behavior, 152–160
- enforcing plan, 160
- issues preventing, 158–160
- preparing organization for, 150–152
- user-driven Business Intelligence, 24
- user interface
 - in SharePoint 2013, 3
 - SharePoint 2010 vs. 2013, 119
- user-level security, 74
- user permissions for control of ECM_MMS service, 105
- users
 - actions on documents, 77–79
 - and library creation, 54
 - community, 154–155
 - document readers vs. creators, 88
 - experts, 155
 - as least common denominator, 149–150
 - profile information, 158
 - super users, 154
 - training, 156
- US (until superseded), 112

V

- versioning, 16, 80–81
- viewing, 21–22, 87–88
 - site contents, 86
- views, 83, 85–86
- Visio, for flowcharting, 66
- visual best bets, 87

W

- waterfall workflow, 64
- web applications, 52–53, 74
 - creating, 93–94
 - creating root site collection for, 95
 - for large-scale site structure, 130
 - as root site collection, 125
- Web Applications Companion (WAC), 42
- web application sites, 167
- web interface. *See also* browser
 - limitations on bulk operations, 9
- Web parts, branding with, 157
- web searches, 155
- Windows Image Acquisition (WIA), 44
- Windows Workflow Foundation, 66
- wisdom, 29

- Word documents, 15
 - scanned documents as, 46
- workflow, 23, 66–68
 - best practices, 67
 - disposition, 63–64
 - three-state, 64–65
 - vs. Content Organizer, 207
- Workspaces in SharePoint, 9, 158
- WYSIWYG administration of workflows, 68

X

- XML file, 238

About the authors



CHRISTOPHER RILEY To say someone is predisposed to be in the field of Enterprise Content Management (ECM) is a little silly. There is no ECM gene, but there are certain characteristics that lead one to be interested in the space. You have a strong sense of order and organization. The statement “information is power” awes you, and you probably graze the boundaries of obsessive-compulsive disorder, or in my case full-on OCD.

I started my career life in content management, specifically in the area of *image capture*. As you will soon learn, capture sits in the early portion of the document life cycle, and at first glance seems so basic. But it is not. The sole job of image capture is to transform paper into value for a content management system, marrying physical documents with electronic. When you start to get into the details of how capture is done, its complexity becomes overwhelming fast. Do you capture all information? Some? What is the right format? What if you need to repurpose it? What do you do with the original? What happens when the types of content you capture are varied? Capture is just one piece of ECM, and as I’ve learned, all pieces follow the same pattern; they seem so rudimentary, but quickly you find that to execute on them well takes thought.

Think of all the unclaimed content or information out there. Think of the information that has haphazardly been pushed aside, a form of capture, and the implications that made on its findability in the future. We know how to generate information but not yet how to use it.

“When information is cheap, attention becomes expensive.”

— James Gleick, *The Information: A History, a Theory, a Flood*

If you were not born with the ECM gene, don’t worry; this book is still for you. The benefits of ECM, I can safely say, are the same for all. Although it might be a taste of medicine for some, the results of proper ECM are felt everywhere.

Whether you insist on hoarding information and organizing it like I do or whether you operate within the confines of normal organizations, with their hodgepodge of personified drive letters and multiple unplanned repositories, the benefits are the same. A well-designed and implemented ECM solution is a chugging engine that gets you on the path to better content utilization.

We all intend to transcend ECM and the simple aspects of managing content and move into a world where the content we spend so much time creating delivers real

value. And by real value, I'm referring to all those technology terms that really excite us: BigData, Cloud, Unified Information Access, and Semantic Web. These more entertaining technologies make one major presumption: that we have succeeded in capturing and storing the content in a way these evolving technologies can consume. So far we haven't.

Today I believe that the technology world has finally mastered the capture of information, but we still are very poor at storing it, and we are only just now flirting with the grand potentials that come when we try to actualize it. Everyone needs the processes of ECM to be successful with their content actualization. This book takes the "why?" and "how?" of ECM and smashes them together into a comprehensive tool. A tool allowing you to form SharePoint into the best content storage and delivery machine it can be.

Writing this book in itself was an ECM exercise. And like all ECM projects, it was hit with detractors and challenges from every angle. While writing this book, I had a job change, fought the economy with the rest of the U.S., and finally said hello to the possibility of being a father. I dedicate this book to everyone who felt the blunt end of my stress to get it done, to Shad for making the efforts see the light of day, and to my beautiful wife and future child.



SHADRACH WHITE Enterprise Content Management is an industry-adopted term that is often referred or rather mislabeled by many. The reasons for this are due to the evolving landscape of information management. Whether it's early card catalogue techniques, microform, document imaging, or records management, the goal is always the same. Take vast and varied amounts of information and organize it so that anyone can find what they are looking for. The truth is, managing and organizing information in any form so that it is easy to find after you store it is one of the biggest challenges that all organizations grapple with, regardless of the technology used. My co-author Chris Riley is a rare example of someone who analyzes and then organizes every bit and byte of information. He refers to this as the ECM gene, and I am lucky enough to have a few of those chromosomes myself. This is one of the reasons he can accomplish so many complex tasks across a wide variety of subject matter and operational boundaries. In this book, our goal is to share our experience, best practices, and our obsessive desire to organize information so that it becomes easy to find and share.

My first role in the ECM field was working as a systems analyst and network engineer in my home state of Alaska. I worked in all facets of the business, from sales to implementation, training, and rollout. After successfully selling, implementing, and supporting dozens of document imaging and optical storage solutions across Alaska, I moved to Seattle. Working closely with upper management, I implemented some of the earliest and most successful document imaging solutions in the Northwest at companies like Eddie Bauer, Boeing, and Costco. It was during this time that I became frustrated with the lack of proper

expectations being set with customers prior to selling them an ECM solution. I recognized the need to create best practices for architecting a document management solution and building a project delivery model. Prior to that, the approach was to sell the software and figure out the details later. Unfortunately, this is still a common practice today, and the results are either shelf-ware or poorly architected solutions. After reading this book, you will have tools and examples that can be used to prevent these outcomes and help you deliver an ECM solution that balances your business objectives, technical requirements, and budget.

The importance of developing a plan and following best practices can be challenging, and in the rush to just get the software deployed and the project completed, many important factors are too often overlooked. To write this book, I have relied on my experience in over 300 content management projects where I variously participated as a lead technical architect, project director, and later as an executive sponsor. Because not all projects are winners, my goal for you as a reader is to share the knowledge that came from both success and failure.

In nearly every project, you will encounter challenges from one of three primary areas:

- **Resources** Time, infrastructure, money
- **Technology** Functionality, integration, devices
- **People** Users, management, project team

I have been asked many times about what the main difference is between a highly successful project and others. Without hesitation, it's about the people. Without a great team, solid executive sponsorship, and engaged users who together take ownership, you can have all the resources and technology in the world and fail spectacularly.