

---

A survey of current best practices for technical communication professionals dealing with

# legacy content

---

Heather M. Snow

University of Minnesota – Twin Cities

WRIT 8505 – Spring 2020

snow0040@umn.edu

---

## Abstract

### Purpose

To examine practitioner experiences and literature pertaining to legacy content, and to establish a set of general guidelines for the benefit of inexperienced technical communicators.

### Method

Search for articles, blog posts, and other web-based resources regarding legacy content; analyze search results for general principles advocated by expert practitioners.

### Results

Management of legacy content is a vital aspect of good content strategy. Left unmanaged, legacy content can detract from user experience, bring the organization into disrepute, or even lead to litigation. Content needs regular assessment in the form of a content audit in order to determine whether it still holds the same value and fulfills its purpose within the organization. When content has been assessed as no longer fulfilling its original purpose, it should be subjected to one of three disposal methods: refresh, reuse, or retire.

### Keywords

Legacy content, content strategy, content audit, content reuse, archives

---

## Introduction

While both content creation and management receive considerable attention in technical communications programs and publications, the management of legacy content goes largely unaddressed. The general assumption, that old content will somehow take care of itself, ignores the pervasive and cumulative problems it can cause. In small-scale instances, outdated contents cause user dissatisfaction, as when a user unwittingly accesses a previous version of how-to documentation and cannot meet their needs because the content simply doesn't match their product. This can lower the organization's public reputation and drive users to seek other avenues to meet their needs. At the other end of the consequence spectrum, incorrect or outdated content can put an organization into noncompliance with regulating bodies, and this can lead to major repercussions that affect the organization's very existence.

The aim of this paper is to establish a view of legacy content in the broader context of content strategy and to explore how experienced technical communication practitioners recommend dealing with legacy content. From this information, this paper will propose a set of guidelines for the benefit of inexperienced technical communication practitioners faced with the challenge of legacy content.

## Methods

Legacy content in the context of technical communication has not made its presence much known in academic literature. A survey of the University of Minnesota online library catalog using the exact-match search "legacy content" produced a small number of results ranging from an article on improving the quality of old videos to an editorial notice by the Journal of Neurosurgery about archiving old articles on their publishing website. One brief, highly technical article discussed ways to convert old, unstructured PDF documents into a searchable format. Nothing in the search results, however, would be of any use in advising new and inexperienced members of the technical communication field about best practices for dealing with, for example, an inherited collection of product documentation which has never been audited and which contains a mix of current and outdated content.

To continue my search, I turned to the literature of the practitioners on the web. In their blogs, I discovered several articles that were key to building my understanding of the concepts dealt with in this paper. I executed a Google search using the keywords "legacy content" and "tech comm" to see what results came up. The search results displayed a number of blog posts from technical communication-related blogs with which I was already somewhat familiar and which I knew to be reputable sources of information. For example, close to the top of the results was the DITAWriter blog, written by Keith Schengili-Roberts. I discovered that Schengili-Roberts has a four-part series on converting legacy content to DITA. This series, titled "Thinking in DITA," provides an in-depth example of the process of assessing a legacy document, planning its conversion from unstructured hard copy to DITA topics, and then putting the plan into action. I found this very helpful in putting concrete details to what had been for me a nebulous concept.

Other, less familiar sources appeared on the results list with resources that I found invaluable. Scriptorium Publishing, Inc., has a compact book titled Content Strategy 101 which is posted in

its entirety online at [contentstrategy101.com](http://contentstrategy101.com). After skimming the online version, I decided to purchase the Kindle version to keep in my e-book library because I found the information in it so useful for establishing a basic understanding of content strategy in general and strategies for legacy content in particular.

In addition to Google searches, I intended to reach out to technical communication professionals via LinkedIn and company websites. My initial tool for this outreach was to be a nine-question survey, eliciting basic feedback on tech comm practitioners' most recent experiences with legacy content. My plan was then to follow up on the survey with interviews, conducted asynchronously through email or in person where possible. However, due to a variety of circumstances, including the COVID-19 pandemic, this plan proved to be ineffectual and was discarded. I limited my outreach to industry-specific information on one significant local technical communication field: medical device labeling.

## **Key concepts and findings**

### **What is content strategy?**

With the rapid accumulation of content in a multiplicity of formats, the pressing need to manage all that content has led to the development of content strategy as an area of specialty. It is a broad discipline closely tied to the fields of technical communication, marketing, web design, UX (user experience) design, copywriting, and more. Halvorson and Rach (2012) offer this definition: "Content strategy guides your plans for the creation, delivery, and governance of content" (p. 28). Content strategy deals with what might be considered the macro level of content (i.e., what are the organizational goals that content must support) as well as in the fine details, like auditing existing content and meeting with the various parties responsible for generating new content. A strong content strategy keeps track of all content, from creation to implementation and onward.

Examining the broader field of content strategy helped me to place legacy content into context. That context may be summarized in one statement: content is a valuable asset and as such repays close attention to how well (or poorly) it supports organizational goals. Content that no longer supports the organization becomes a potential liability called legacy content.

### **What is legacy content?**

The umbrella term "legacy content" includes any type of content that is no longer valuable or usable in its current form. It may exist in an outdated format, such as WordPerfect documents, or the information in it may no longer be correct. This encompasses a variety of content, such as broken hypertext links, blogs that have not been updated for several years, or documentation for products no longer necessary to users. One might say that the category of legacy content includes all content that has reached the end of its lifecycle.

## Why is legacy content a problem?

Paul Boag, a user experience consultant in the UK, describes two primary negative impacts that website legacy content can cause when it is not managed. The first is findability, or navigation. Boag warns that “every piece of content you add to a website adds to the user’s cognitive load and increases the likelihood of them being unable to find the answer they need” (“Do you know,” 2020). This in turn increases the likelihood that the dissatisfied user will turn to a competitor and urge others to do the same. In extreme cases, negative user reviews can severely damage a company’s reputation.

The second impact is inaccuracy, as outdated content can give users incorrect or incomplete information, thus opening up the potential for damage to the organization’s reputation. It can also lead to potential loss. In a worst-case situation, incorrect or incomplete content could lead to expensive litigation or noncompliance penalties. Even if worst does not come to worst, the cost of maintaining a bloated collection of web content or storing unconverted hard copies of content can become excessive.

## How does one identify legacy content?

According to all of the practitioners whose literature I consulted, the first step to dealing with legacy content is identification. In order to weed out legacy content, it is usually necessary to perform a content audit. O’Keefe and Pringle (2012) describe a content audit as a way to “determine what information you have and how useful it is” (chapter 5, section 1, para. 1). While I found no consensus on how often an organization should audit its content, the general opinion was in favor of performing audits regularly. In some cases, such as when dealing with large collections of content, the audit may become an ongoing project or “rolling audit— an audit that basically never ends” (Halvorson and Rach, 2012, p. 62). The importance of knowing what content your organization owns and, in many cases, who is in charge of it will determine the decisions you will make regarding legacy content.

Richard Pope offers a helpful example of a content audit in his blog post, “Introducing the Needotron.” When the British Government Digital Service (GDS) decided it was necessary to overhaul the entire gov.uk website, the audit team went through “long lists of pages and search terms from Directgov [the original government website]” (“Needotron,” 2011, para. 5). They worked this down into a list of perceived user needs, gave each of these needs a priority ranking, determined whether the new website should deal with each of these needs, and cluster the needs into website functions. They went so far as to design an app, the Needotron, to track the whole process of deciding what should and should not be included in the new website.

## How does one deal with legacy content?

In the case of the gov.uk website audit, the intention was to start almost from scratch— a solution that O’Keefe and Pringle refer to playfully as “the scorched earth option” (2012, chapter 5, section 1, para. 4). According to the practitioner literature that I consulted, this is the most

extreme method of dealing with legacy content and should only be adopted if the content itself is beyond saving, or if the context in which the content resides is to be radically changed, as in the gov.uk redesign example. More commonly, the fate of legacy content falls under one of three Rs: refresh, reuse, or retire.

### 1. **REFRESH** legacy content

Some legacy content is outdated but can be brought up to date with a few minor alterations, such as a change in dates or other cosmetic details. The automated content marketing firm Socialtoaster recommends that organizations ought to “dedicate (at least) 25% of their content marketing strategy to repurposing and reposing [i.e. refreshing] legacy content” (“5 ways”, section two, para. 3). For example, the marketing department might have a social media post from an event that took place several years ago, which could fit a similar current event after minimal changes. They are careful to note that this only works well with short content, such as email and social media content.

### 2. **REUSE** legacy content

For the purposes of this project, I differentiate between refreshing content and reusing content according to the degree of alteration required. Content may be refreshed with very little effort, whereas reuse calls for more drastic changes. Converting legacy content from one format to another, especially converting it into structured content, falls under the reuse category.

As mentioned in the Methods section of this paper, Schengili-Roberts offers on his blog a four-part series on converting legacy content to DITA. “Thinking in DITA” provides insights into the process of assessing a legacy document, planning its conversion, and then putting that plan into action. Schengili-Roberts’ series focuses specifically on analyzing hard-copy content for importing into a DITA-based content management system (CMS), but the principles he advocates can be applied to practically any content conversion.

The one principle emphasized by every source I found was this: plan as much as possible before starting any reuse project. The more planning that goes into a conversion project beforehand, the fewer snarls the project encounters while in-process. Schengili-Roberts refers to this process as a document audit process: a close examination and analysis of a specific piece of legacy content, while asking such questions as, “What audience is this intended for? What scenario(s) are readers likely to be using this manual?” (2012, “Thinking... Part 1,” para. 10). Kai Weber, in one blog article on the subject, offers this advice: “Even if it takes a week or two, be sure you have 80, 90% of your structure in place before you start tearing up contents to convert them [to topic-based authoring]” (2011, “Top 4 tactics,” para. 6). It is essential for anyone undertaking a legacy content conversion project to understand the content as it currently exists in order to reuse it effectively.

The real benefit offered by content reuse lies in the renewal of an existing document’s value. There are, however, situations in which content cannot or should not be reused: whether because it refers to a product or service so out-of-date that it no longer has an active use, or because it was so badly written that replacing it with completely new content would be more

efficient, or simply because it was specific to one moment and has no value beyond that moment. In cases like these, the third R is the most appropriate method.

### 3. RETIRE legacy content

Under the third method, that of retiring legacy content from circulation, there are two main paths a technical communication professional can choose from: either to dispose of content or to archive it. The nature and qualities of the content will determine which path is more appropriate in any given situation.

Permanent disposal should be reserved for content proven to have no further use in your organization. Sometimes, old content simply does not have enough value to merit the effort and expense of reuse. This may be because the content pertains to a product or service that has become so obsolete that no one uses it any longer. Keith Schengili-Roberts' example for the conversion of legacy content into DITA is able to use the Radio Shack TRS-80 Expansion Interface Operator's Manual precisely because the product has been obsolete for so long that the manual has passed into public domain.

In some cases, the legacy content in question is so badly written that it would be more efficient to dispose of it and start from scratch. This may seem a radical solution at first glance, but every decision about legacy content always comes back to the concept of value. Content reuse does cost an organization, any reasonable strategy to convert legacy content must take into account the fact that not everything is worth keeping. Here we return to O'Keefe and Pringle's "scorched earth policy" mentioned earlier: if analysis of a document results in a content reuse plan that takes more time and effort than the creation of a new document would take, then that content qualifies for permanent disposal.

There are other factors that might lead a technical communicator down the permanent disposal path. For instance, redundancy is a strong indicator that a piece of content is unnecessary. Boag lists it first in his acronym ROT: "redundant, out-of-date, and trivial" (2020, para. 22). In cases where content has resided for years in departmental silos, for example, there is a high probability that some of the content has been duplicated. The politics of sorting out which version should be culled may well be tricky to navigate, but success in this area automatically pares down the load of legacy content bogging the organization down. The same holds true of content that falls under Boag's third heading, the trivial content. He recommends establishing specific criteria for distinguishing the trivial from the indispensable, such as how well a piece of content supports organizational or user goals (Boag, 2020).

Disposing of legacy content is not always advisable or possible, or even desirable, especially in regulated industries. In these cases, archiving legacy content is the only option. For example, I consulted Daphne Walmer, an expert in the field of medical device labeling, about the best practices for dealing with legacy content in that industry. I asked three questions:

- ✦ What happens to medical device labeling after a device stops being manufactured?
- ✦ How long must one keep this legacy labeling content?
- ✦ Do the regulations differ between US regulating bodies and European regulating bodies; if so, how so?

Her response brought out some unexpected details. First, in the medical device industry, there is a vast difference between a device no longer being manufactured and no longer being used. She gave the example of pacemakers, which might be in use in a human body long after that particular model of pacemaker is replaced by a newer model. There are also temporary pacemakers that are reusable, thereby lengthening the device's time of use. For that reason, the legacy content connected with these devices cannot simply be discarded when the devices themselves fall out of production. Walmer said that, "in practice, it was easiest to keep the manuals pretty much forever" (personal communication, March 30, 2020).

Not surprisingly, regulations about maintenance of legacy content differ between the US and the EU. In terms of legacy content and regulatory compliance, Walmer described current practice at the time of her retirement in 2013 as follows: "the U.S. only required that the most current manual be available on the website. For the European Union and other places that require [the] CE mark, the requirement is that all released versions of manuals be available online" (D. Walmer, personal communication, March 30, 2020). It is important to obtain clear guidance specific to your industry's regulations when determining if legacy content must be archived (and if so, for how long).

Legacy content in some cases still has value as reference material for research. This can be considered a different way that content can support organizational goals, thus giving it a different value. Much of the source material I was able to use for my research came from blog archives and slide sets left over from conference presentations. Although these are no longer considered active content, they can still be made available for reference because of the immense archival capacity of the web. Boag notes that "completely removing content from the web... can often be an unwise decision" and recommends instead "isolating the page from the rest of the site... [removing] it from site navigation" so that it is accessible only via external web searches (2020, "Do you know"). Dealing with the content in this way reduces the occurrence of broken hyperlinks, in the event that someone has linked to the content in the past, and makes the content available for reference purposes without cluttering up its original website with archival materials.

## Conclusions

Legacy content compels us to examine the value of content. A good content strategy must take into account the fact that all content eventually loses its value and must be dealt with, not simply left to become so much clutter in the organizational system. A case in point: a certain company recently released to employees their quarterly list of required training courses. One employee went into the training course library, pulled up a course that matched a name on that list, and spent an hour and forty-five minutes going through the course, only to find out that she had just taken last year's course, under the same course name, which did not fulfill compliance requirements for this year. That was an hour and forty-five minutes of company time wasted, not to mention the heightened frustration on the employee's part, simply because no one dealt with the legacy content in their training course library. And this happens regularly in many other contexts. Although a single incident may be minor, these incidents accumulate, and the costs

are real. Legacy content, left unmanaged, can cause issues ranging from user dissatisfaction to litigation. Legacy content needs to be dealt with.

The benefits are real as well. Dealing well with legacy content can provide a cost-effective alternative to creating new content from scratch. In some cases, it can provide an archive of rich reference material upon which users can draw for years to come. At the very least, a good strategy for legacy content helps create a more orderly, more streamlined collection of content. This benefits users and employees alike, which in turn benefits the whole organization.

To deal effectively with legacy content, one needs an awareness of existing content in general. There is widespread agreement among the sources I consulted that a content audit should always be the first step when beginning to deal with legacy content. Regular content audits are a way to ensure that a valuable resource retains its value over time. This helps not only with identifying legacy content, but also gaining an understanding of what exists generally in an organization's content collection. If it is not possible to audit annually, at least every two or three years should be the goal, so as to make sure that content is either updated regularly or dealt with according to the three methods recommended in this paper: refresh, reuse, or retire.

To review, the three Rs apply to different kinds of legacy content. Practitioners need to assess the content on a case-by-case basis to determine whether it would benefit from a quick refresh (cosmetic changes with rapid turnaround), a thorough reuse (major reconstruction with a slow turnaround) or a permanent retirement (secure destruction or storage in some form of archive).

Based on the information I studied in this project, I would recommend further research in the area of industry-specific legacy content. This project was only intended to be a survey, teasing out general principles that could apply to any industry, but it made me aware that certain industries have very specific requirements and needs. This is particularly true of regulated industries such as may be found in medical, financial, and legal organizations. New technical communication practitioners entering into these industries need to familiarize themselves with the specific requirements and expectations, but with the paucity of attention given to legacy content, they may not know where to find this information.

Moreover, I would also recommend that programs of study in technical communication should include at least an overview of content strategy. This area of specialization offers a big-picture view of content, at the collection level, where some of the most sweeping decisions about content are made. With the speed of content creation always increasing, students of technical communication would benefit from an introduction to the skills and knowledge underlying these decisions. Such an introduction would enable them to craft better content and would prepare them for dealing with what is an undervalued but significant area of technical communication: content management at the end of the content lifecycle.

---

## References

- 5 Ways To Make Legacy Content Part of Your Content Marketing Strategy - SocialToaster: Automating Content Marketing. (2019, March 26). Retrieved March 19, 2020, from <https://www.socialtoaster.com/legacy-content-marketing-strategy/>
- Boag, P. (2017, July 19). Content Management: How to effectively overcome legacy content. Retrieved February 29, 2020, from <https://boagworld.com/content-strategy/content-management-legacy/>
- Boag, P. (2020, January 23). Do You Know How to Deal With Legacy Content on Your Website? Retrieved February 29, 2020, from <https://boagworld.com/content-strategy/legacy-content/>
- Bowerman, K., and Kara, L. (2016) Legacy content: applying your new content strategy to old information [PowerPoint slides]. Retrieved from <https://www.slideshare.net/salesforceeng/legacy-content-applying-your-new-content-strategy-to-old-information>
- Coles, L. (2019, September 20). How to Make Legacy Content Work for You: Blog. Retrieved March 19, 2020, from <https://bookmarkcontent.com/blog/marketing/how-to-make-legacy-content-work-for-you/>
- Edic, M. (2016, June 2). The Documentation Problem: Is Structured Content An Order of Magnitude Improvement? Retrieved April 1, 2020, from <https://techwhirl.com/documentation-problem-structured-content-order-magnitude-improvement/>
- Halvorson, K., & Rach, M. (2012). Content strategy for the web, second edition. Berkeley, CA: New Riders.
- Johnson, T. (2016, February 21). Answering questions in James Gill's upcoming book, How to Get Started as a Technical Writer. Retrieved April 1, 2020, from <https://idratherbewriting.com/2016/02/21/technical-writing-careers-a-day-in-the-life/>
- O'Keefe, S. S., & Pringle, A. S. (2012). Content Strategy 101. [Kindle edition] Durham, NC: Scriptorium Publishing Services, Inc.
- O'Keefe, S. S., Pringle, A. S. (2009). Technical Writing 101: A Real-World Guide to Planning and Writing Technical Content (Third Edition). (n.p.): Scriptorium Publishing Services, Incorporated.
- Pope, R. (2011, September 19). Introducing the Needotron: working out the shape of the product. Retrieved February 29, 2020, from <https://gds.blog.gov.uk/2011/09/19/introducing-the-needotron-working-out-the-shape-of-the-product/>
- Schengili-Roberts, K. (2015, February 11). Thinking and Writing in DITA: Converting Legacy Content to DITA, Part 1. Retrieved February 29, 2020, from <https://www.ditawriter.com/thinking-and-writing-in-dita-converting-legacy-content-to-dita-part-1/>
- Schengili-Roberts, K. (2015, July 27). Thinking and Writing in DITA: Converting Legacy Content to DITA, Part 2. Retrieved February 29, 2020, from <https://www.ditawriter.com/thinking-and-writing-in-dita-converting-legacy-content-to-dita-part-2/>
- Wachter-Boettcher, S. [O'Reilly] (2013, March 29) How to get unstuck: content strategy for the future [Video file]. Retrieved from <https://youtu.be/scTpNWhwQew>
- Weber, K. (2011, August 13). Top 4 tactics to structure legacy documentation. Retrieved February 29, 2020, from <https://kaiweber.wordpress.com/2011/08/15/top-tactics-to-structure-legacy-documentation/>