

Using PDFs to manage and share content. The balance between quality and size.

Whether you are trying to leverage the intellectual property that is valuable to your organization, improve decision making, increase sales opportunities, or simply manage a vast amount of digital documents, the Portable Document Format (PDF) is often the standard of choice. PDF delivers portable and consistent content, to both represent your company and make it easy for a user to access your information. While PDF is the most consistently used document type for long-format materials, it can sometimes provide challenges due to the wide variety and inconsistencies in both authoring and consumption tools available.

Consider:

In 2014, 3,408 B2B companies were surveyed and the results published in the B2B Content Marketing Trends-North America report from the Content Marketing Institute, MarketingProfs, 2014. According to the report, respondents produced over 42,000 pieces of content per year to support content marketing efforts. Of this content, 37% (or 15,692 assets) represented long form content such as eBooks, white papers, and case studies, where the PDF works best to meet this need.

According to the article “PDF in 2016: Broader, Deeper, Richer”, from the PDF Association, Phil Ydens, Adobe Systems VP of Engineering, had some interesting estimates to share during the 2015 PDF Technical Conference:

- In October, 2015, there were approximately 1.6 billion PDF documents on the web. About 80% included the string “2015”, implying that content is fairly current.
- 60% of non-image attachments in Microsoft Outlook Exchange Enterprise emails are PDF documents.
- A conservative estimate is that 2.5 trillion PDF documents are created each year.
- PDF is the most popular format in cloud storage repositories; in many cases PDF accounts for greater than 50% of the files stored.
- 17% of files in Box are PDF documents.
- 18 billion PDF documents are in DropBox.
- 73 million new PDF documents are saved each day in Google Drive and Gmail.

PDF is the industry standard, and a leader when it comes to sharing information. The PDF file format's goal is to present document content reliably regardless of the software, hardware, or operating system in use, and is popular for use with long-form content. It is great for generating print-ready content, and will open and display well on any workstation or laptop after being downloaded from a server, making it an ideal choice for delivering content. However, once PDF became an ISO standard in 2008, and as more vendors entered the market to create and use this royalty-free document format, the variety of PDF creation and consumption tools on the market added a level of inconsistency to the process, and in many ways, reduced the reliability and consistency of PDF files.

The popularity of the format and nearly global acceptance combined with the advent of open standards for PDF creation means that hundreds, maybe even thousands, of software products can now generate a PDF document. With other added complexities, like variations in rendering, loose adherence to the PDF standard, and the variety of PDF authoring and consumption tools available, PDF documents, now more than ever, can lack basic consistency. This provides a challenge to content providers for appearance and reliability. As gloomy as this message seems, it is possible to minimize these and other issues related to PDF content and provide flawless representation of your brand along with the ease of use that users expect from PDFs.

Take content marketing, for example. According to the Content Marketing Institute's 2016 B2B/B2C Content Marketing Trends Report, 70% of a buyer's journey is complete before a buyer reaches out to sales. As a result, marketers are now looking to deliver information that makes prospective buyers more intelligent about their products. Because of this, businesses are focusing on delivering valuable information, in a consistent format, across all platforms, and need to make sure that information is easy to download, easy to open, the right size for emailing, as well as impeccably representing their brand.

Information is most effective when it is written, designed and distributed in a manner that is tailored to a specific reader. That's why it is important to consider not only what you're saying, but also the format and the method of distribution so that the reader of the content has the best experience possible.

Imagine you're planning your content marketing strategy. You want to distribute a white paper in PDF format online, and intend for your customers and prospects to open and view it on their mobile devices. Will it look the way you want it to look? Will the images, colors, and fonts you so carefully selected continue to appear as expected to accurately represent your brand? Will it resize appropriately?

And what if you want to share your PDF documents as email attachments, or offer them for download from your web site? Will they be too large for the user's email server? What if your four-page marketing document featuring color photographs, diagrams and charts, name and address fields, a few check boxes, and some unique fonts, tips the scale at 65 MB? Just how long will it take your prospect to give up during the download process? According to MarketingSherpa, "Conversion Rates Correlated to Page Load Times (2014)," slow pages produce lower conversion rates. In fact, a decline in conversion rate begins after waiting only 4 seconds. That's how important it is to ensure you've optimized your documents.

PDF document authoring tools adhere to the PDF specification in varying degrees, and they are rarely held accountable for the quality, usability or accessibility of the documents produced. Everyone with a computer can create and display a PDF. Content managers rarely have control over the way a PDF file gets created. This can result in a document that is too large, often not optimized for fast web viewing. It can be missing fonts, it can be just an image, or even a broken document that won't meet the promise of a PDF, i.e, to present document content reliably. PDF document consumption tools will account for and often repair documents on the fly, to the best of their ability, but with varied success. Other tools might just give up and leave the file unchanged. No matter what, it is imperative for document consumption tools to have a high quality process in place to correct and optimize PDF documents both for internal processes and external user distribution.

The PDF standard is 1100 pages long, so it comes as no surprise that it takes a lot of expertise to craft a good PDF file. It takes even more know-how and expertise to fix a poorly crafted one. The more advanced your set of PDF tools are, the less you need to worry about creating good PDFs.

Understanding Fonts in PDFs

When working with fonts, PDF documents are designed to be self-sufficient. The best practice is to store the font sets used for the text in a PDF document within the PDF document itself. In other words, font files are saved as part of the larger PDF file. That way, when a user opens the PDF file, that user's computer system already has all of the fonts it needs. The machine or device used to open and display a PDF document doesn't have to go looking for a set of characters stored on the local workstation or smartphone, and maybe guess at a reasonable replacement if the right font file is not available. This standard practice is called font embedding and is an important part of the PDF format. It guarantees that the text in a PDF document will always look the same no matter how or where it is displayed.

Embedding fonts is a great way to ensure consistency, but it comes at a cost. Font files can be quite large, especially if the font includes Asian languages with vast character sets, like Mandarin or Japanese. You can make a PDF document smaller and easier to share by removing these font sets if they are not needed. You need to make sure that the fonts used in the PDF document are common, so that when a customer opens the file, the device the user is working with can find the fonts locally. Ideally, you want to embed only the font characters used in the document, creating smaller file sizes while ensuring the proper characters are in place to represent the file as the author intended.

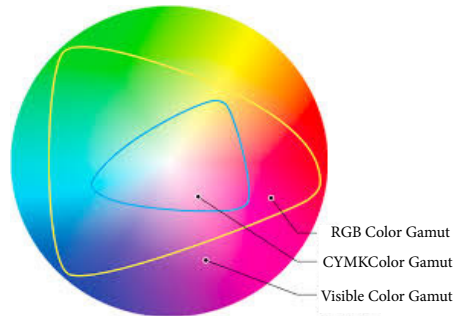
Color that Represents your Branding

Starting in the 1960s and 70s, computer and printing technology grew steadily more sophisticated. A wide range of computing devices were introduced by a wide variety of manufacturers, including workstations, monitors, scanners, copy machines, printers, digital cameras, presses, many different kinds of printers, and eventually, laptops and mobile devices. How this hardware presents colors on screens or in print varies from one device to another, because a device is designed to independently verify the colors it uses.

To create a uniform standard to govern how computer technology worked with colors, the International Color Consortium (ICC) developed a color specification in 1993 that was designed to work across all hardware and software platforms and for all operating systems. A variety of color profiles have been introduced, in the form of .icc profiles, which are a mathematical description of a device's color space, and allow multiple devices to select the appropriate colors to use when rendering content. A color management system (CMS) compares the color space in which a color was created to the color space in which the same color will be output, then makes the necessary adjustments to represent the color as consistently as possible among different devices.

A color profile can be stored on a hardware device to define what a monitor can display or a printer can print. Some color models have a fixed color space because they relate directly to the way humans perceive color. Other color models, (RGB or CMYK) can have many different color spaces. Because these models vary with each associated color space or device, they are often described as being device-dependent. It's important to remember that no device is capable of reproducing the full range of colors viewable to the human eye. Each device operates within a specific color space that produces an individual range, or gamut, of colors. The diagram below illustrates a standard color gamut.

A color gamut



Reproducing color across different mediums accurately is a difficult problem to solve. Color-matching problems result from various devices and software using different color spaces. Because of these varying color spaces, colors can shift in appearance as you transfer documents between different devices. Color variations can result from differences in image sources, the way software applications define color, print media, and other variations, such as manufacturing differences in monitors or monitor age. This shift in colors represents a real problem when documents are intended to represent your brand. Rendering the content for an individual device does not ensure that the content will be read just on that specific device. Users can very easily read a PDF on multiple device platforms such as an iPad, desktop computer, laptop, or a phone. As a result, users creating content must utilize tools that have color management capabilities so that your brand is accurately represented across platforms. Accurate color representation is particularly important when it comes to presenting items that define a business' identity like company logos, for example.

Balancing the Image Quality with Portability

When a PDF document includes images, such as diagrams or photographs, these images are embedded in the PDF document as graphics files, such as JPG, TIF, and BMP files. If the original images are created with high resolution, they will look good but may be quite large, and as a result, the PDF document in which they are stored will also be quite large. To make a PDF document smaller, you can reduce the size of images in a PDF document by down-sampling them.

The down-sampling process reduces the resolution of an image, while protecting the integrity of the original colors displayed. When you down-sample, information is deleted from the image. Effectively, the software reduces the width and height of images, in pixels. However, not all down-sampling results in acceptable image and document quality. There is a real art to creating files that are optimized for sharing while still representing the document and your brand accurately. Compression and down-sampling with the right tools can significantly reduce the size of a PDF with little to no loss of detail and precision. However, you will want to experiment with various PDF options to find an appropriate balance between file size and image quality.

Transparencies

A transparency refers to objects on a page, such as images or text, which are transparent or ‘show through’ in some way. Examples may include lightening part of an image so that text shows through, or showing parts of an object that would otherwise be hidden under a layer. Transparencies are generally used to add complexity and richness to documents. While providing great benefits, transparency is a complex technology to process. In fact, the Adobe technical documentation on transparencies is over 100 pages long. Nonetheless, today’s document creators are still regularly creating documents that contain multiple transparent elements.

Processing transparencies can be tricky, and the outcome can vary based on the tool used. This can result in an inconsistent visual message across your users. Generally, if you want to distribute a PDF document with transparencies to a broad audience, it makes sense to first flatten the transparencies. This process works just the way it sounds—the layered transparencies are squeezed together and saved as a single permanent image that is embedded in the PDF document. Flattening transparencies creates a stable standard version of the PDF document; the resulting file will look the same, regardless of the hardware or software used to view it later. This process also tends to make the PDF document larger, and it is a good idea to optimize the resulting images.

It is important to note that other software products used to view or work with flattening PDF documents are often designed to automatically flatten any transparencies found in a PDF. All the more reason to flatten transparencies in PDF documents yourself, as this allows you to control the settings used versus relying on an automated process that may not deliver the desired outcome.

PDF and the Intended Reading Space

One of the big issues to consider is that PDFs are still somewhat difficult to view on mobile devices. Each operating system handles them differently, forcing users to download various apps to view the PDFs, making them hard to consume. Readers have to resort to scrolling or pinching-and-expanding to read the document.

These PDF obstacles make it less likely that mobile device users will read your content. As mobile content consumption has surpassed desktop consumption, it's important to prepare documents for the intended viewing method and to optimize accordingly to address the platforms used to consume your content.

Managing PDF Documents using the Adobe Library and PDF Optimizer

When working with PDFs, you have to manage your documents with two things in mind: quality and file size.

You need to find a way to reduce the size of your file as to make it easier to use, while making sure that you don't lose anything that might jeopardize your brand or the integrity of the original document. You want to protect your reputation for quality; that is not going to happen with a white paper that might contain fuzzy drawings, images, and mismatched fonts, no matter how quickly the reader can open the document on mobile devices.

That's why Datalogics offers PDF Optimizer which is built upon the Adobe PDF Library. PDF Optimizer, available as an API or as a command-line utility, works with the Adobe PDF Library to simplify the process of altering existing PDF files to increase compatibility with user's viewing environment, while making the files smaller. Essentially, it automatically reduces the size of a PDF document without compromising the quality of that document. With it, you can embed tools in your process that:

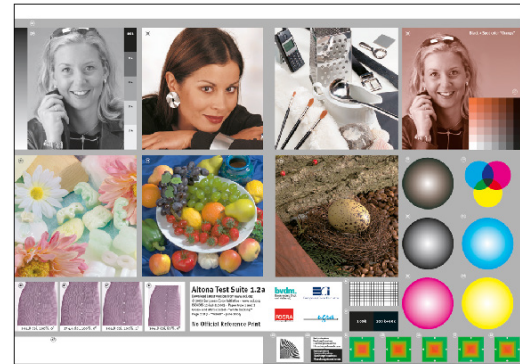
- Create a PDF file optimized for the targeted viewing method
- Reduce image resolution while maintaining image fidelity
- Convert color in a way that is ideal for the target medium
- Flatten transparencies to improve document portability
- Embed only the necessary subset of fonts to view the document
- Apply customized settings with the API toolkit or execute process automation using the command-line utility
- Implement settings without the need for prior PDF experience

And with the Adobe Library and PDF Optimizer, you can be confident that your readers won't notice any changes in how a PDF document appears after you have reduced its file size. The colors remain true, the images still look as good as ever when displayed, and the fonts won't change. This applies across Windows and Mac browsers, Apple and Android mobile devices, Kindles or Nooks, or any of the variety of devices and platforms your customers are likely to use. With the Adobe PDF Library and PDF Optimizer, you can repurpose your existing PDF documents and the PDF workflow you already have in place to create PDF documents that are smaller and easier to distribute, while maintaining the integrity of the original document.

Altona Test Suite



Original file size 11.6Mb



Optimized file size 2.44Mb

You may have considered switching from PDF to HTML or EPUB files for presenting a document to users with mobile devices. But, what if you already have in place a process that generates the PDF documents you need, and that has worked just fine for many years? Installing new software, retraining your staff, and building new batch processes to replace your PDF content with EPUB or HTML content could take a lot of time and money, which may not make economic sense for your organization. You have the PDF files you need. Simply use the PDF Optimizer from Datalogics to make these portable files more portable. If your goal is to create PDF documents that otherwise look good and don't change, and that are small enough to distribute online, PDF Optimizer will help address this need.

Adobe Acrobat provides a related function. If you open a PDF in Acrobat and select 'File' and 'Save As', 'Other', and then choose 'Optimized PDF', you can create a compressed version of the file. The feature in Acrobat offers detailed parameters that you can work with to adjust how the software product manipulates your PDF document, and the final result can be dramatic. Using Acrobat for optimizing is only practical if you are prepared to open and manually optimize each PDF document one file at a time, and may not be practical for larger batch processes.

What if you have hundreds, or maybe thousands, of PDF documents you need to work with? What if you are creating hundreds of new PDF documents every day that need to be optimized? The advantage of working with the Adobe Library and PDF Optimizer is that you can use it as a foundation for building your own code to optimize PDF documents, and run that code or execute the command-line utility from a server. You can embed the optimizer code in another application or run it as a batch process after business hours, automatically processing dozens or hundreds of PDF documents overnight. With PDF Optimizer, you can easily add a simple step to your existing PDF document workflow to optimize each PDF document automatically as it is created, and with the settings that make the most sense for your needs.

How PDF Optimizer Works

PDF Optimizer looks at every part of the PDF document for elements that can be squeezed, or that don't need to be there. First, the utility downsamples, and/or compresses images found in the document. Then it identifies a variety of items that could be considered optional. This would include some images and font files stored within the PDF document itself, as well as other items such as comments and other changes, metadata, color profiles, bookmarks, thumbnails, labels, and dictionaries.

If the primary goal is to reduce the size of a PDF document so that a user can easily open and read it online, there are many items within a PDF document that can be removed. For example, some PDF documents store multiple copies of images to use as alternates for printing or displaying on certain kinds of devices. Usually the complete fonts that a PDF document needs are stored within the document itself, so that the system used to display a PDF does not need to look for copies of the fonts on the device where the PDF is opened. You can also probably forgo bookmarks and some non-essential navigation tools. You can set up PDF Optimizer to identify these sorts of objects within a PDF document and remove them, or you can use the default set of settings provided.

Managing Images

You can set a target resolution for the images in a document, and the PDF Optimizer tool will reduce the resolution for each image found in that document to match this target. Or, if a single image appears multiple times in a PDF document, PDF Optimizer can select one example of that image and use it as a benchmark. All of the other examples of that same image in the document would be reduced to match the resolution of the image you selected.

PDF Optimizer can also reduce the size of images in a document by compressing them using a format such as Flate or JPEG. The tool can even un-compress an image and then recompress it using a different compression method, if the new compression method is more effective at reducing file sizes. Finally, the tool looks for and discards duplicate images embedded in a PDF document that are not needed. Sometimes a PDF document holds alternate images, or multiple copies of a single image, so that one can be presented to a printer, and another to a monitor.

What customers are saying about Datalogics:

"Datalogics has provided superb help with very short turnaround time that enables us to resolve problems in our products accurately and quickly."

"I love using the Adobe PDF Library - we were using a mish-mash of other tools and this API makes our product so much better."

"I really enjoy working with the team and appreciate the extra effort that goes into the quick answers to engineering questions."

Managing Fonts

If you are using a standard font in a PDF document that is likely to be found on any machine, you can use PDF Optimizer to remove font files saved within the document. The resulting PDF file can be dramatically smaller, and the appearance will not change when the file is opened. If an uncommon font is used, the tool can automatically create a font subset containing only the necessary characters to open the document, resulting in increased compatibility and reduced file sizes.

Removing Objects in the PDF Document

You can set up PDF Optimizer to remove items from a PDF document to enhance its portability. If your goal is to make a PDF document as compact as possible, so that you can distribute it easily online, and you know that the document will likely not be printed, you can set the tool to eliminate objects automatically. This includes:

- Bookmarks
- Page thumbnail images and labels
- Attached files
- Several kinds of internal dictionaries that aid navigating through the document or that are used for editing and managing the PDF file
- Comments and metadata

Making Adobe PDF Library and PDF Optimizer Work for You

Datalogics provides a product that allows you to process dozens or hundreds of PDF documents at a time, significantly reducing the size of those documents without compromising their quality. With the Adobe PDF Library and PDF Optimizer, you can take existing PDF documents and make them much easier to share. To download a free evaluation of PDF Optimizer go to <http://www.datalogics.com>.

About Datalogics

Datalogics, Inc. is a PDF, EdTech and eBook technology company, dedicated to providing the highest quality software. Based in Chicago, IL, we support hundreds of customers worldwide who are using our technology in diverse applications. With over 50 years of industry experience, we provide unmatched support, enabling customers to bring their products to market faster. Datalogics is on the PDF Association board, and is a member of W3C Transitional Publishing Industry (TPI) and the Radium Foundation.



Datalogics, Inc. | 101 N. Wacker Drive | Suite 1800 | Chicago, IL 60606 | USA
datalogics.com | sales@datalogics.com | 312.853.8200

Datalogics and the Datalogics logo are registered trademarks of Datalogics Incorporated. Adobe and the Adobe logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States and/or other countries. All other trademarks are the property of their respective owners.
©2017 Datalogics, Inc. NBEK17