

Document Imaging Report

Business Trends on Converting Paper Processes to Electronic Format

4003 Wood Street ● Erie, PA 16509 ● PH (814) 866-2247 ● <http://www.documentimagingreport.com>

September 22, 2017

PDF/raster Doc Imaging Format Published

When I first started writing about the document imaging market, in the late 1990s, I remember questioning **Adobe** representatives at an AIIM show about why they thought anyone would adopt a proprietary format like PDF? Boy, was I off the mark. Nearly 20 years later, everyone has a PDF reader, with many owning multiple versions if you count mobile devices. Digital copiers, and even some document scanners, often produce PDFs as their standard image output. And, oh yes, in 2008, with the blessing of Adobe, PDF became an ISO open standard.

And yet, the go-to file format in many document imaging applications is still TIFF—a specification originally published in the 1980s that hasn't been updated since 1992. Meanwhile, PDF continues to evolve, including the recent release of PDF 2.0 (the first new full version release after several point releases over the years). However, because of its simplicity, TIFF has remained attractive to many document imaging engineers, especially in the scanner market.

In an effort to finally end the grip that TIFF has had on the industry, a couple years ago the **TWAIN Working Group** and the **PDF Association** got together and launched a co-development effort focused on creating a document-imaging specific version of PDF—

PDF/raster [see [DIR 10/30/15](#)]. This month, they announced the joint publication of PDF/raster 1.0. According to the press release, it "is a strict subset of the PDF file format. It was designed for storing, transporting and exchanging multi-page raster-image documents, especially scanned documents."

"Even though we've known for years that PDF is probably the best file format for document imaging, it's mostly been too intimidating to utilize the PDF spec with scanners without setting any boundaries," said Jon Harju, the chairman of the TWAIN Working Group and the CTO of scanner vendor **Visioneer**. "PDF/raster is a highly restricted version of PDF, but it's 100% compatible with any PDF processor and/or reader. Because of its restricted nature, it's possible to write and read PDF/raster files with a very light code footprint."

Like TIFF, PDF/raster supports Group 4 compression, as well as uncompressed raw data. It also supports JPEG compression, which is how most PDF document images coming off scanners today are created. PDF/raster supports color, grayscale and bi-tonal images in a variety of different bit densities. The spec also supports meta data, encryption, and digital signatures—all of which can be created at the time of scanning—and none of which

are supported by TIFF.

"The goal with PDF/raster was to support everything that TIFF does, while also making it a little more user friendly by introducing some features that can be incorporated in a PDF but not in a TIFF," said Harju. "At the same time, we wanted to limit the capabilities so as not to make it too intimidating to try and get a PDF/raster image off a device."

Harju noted that even in its stripped-down state, parsing a PDF/raster file can be more complex than parsing a TIFF. "To address that, the TWAIN Working Group will provide a free lightweight open source PDF/raster reader and writer that will be available from our Web site," he said.

The PDF/raster format is designed to be very versatile. Once a PDF/raster file is created, it can be processed further utilizing traditional imaging technologies. "Even though a PDF/raster is a fully formed file, it still offers access to the raw image data," Harju said. "That means if you wanted to apply something like VRS that requires a defined set of image data, you can accomplish it much more readily than you could working with a fully formatted PDF. In addition, you can apply more advanced PDF processing to PDF/raster files, like JBIG2 compression and OCR, to create more complex PDF files.

There is also a fairly direct path to creating PDF/A (archive) files from PDF/raster.”

PDF/raster was conceived to work in tandem with the TWAIN Direct driverless scanning protocol being developed by the TWAIN Working Group [see [DIR 9/26/14](#)]. TWAIN Direct is designed to move scanning into the future by enabling direct network and cloud connections to document scanners. “PDF/raster was designed specifically to satisfy the criteria of TWAIN Direct,” said Harju. “That means the image data has to be transferred as fully formed files in a universal format that can support the output of modern document imaging devices with minimal transformation.”



**Jon Harju, chairman,
TWAIN Working
Group & CTO,
Visioneer**

Harju added that scanner drivers may have worked okay when the computing world was dominated by Windows and Mac. “But, with the growth in the use of mobile computing and the multi-platform environments that we envision going forward, it makes sense to make scanning a network protocol, so a scanner can be interfaced with just like any other network appliance,” he said.

The adoption of mobile is another reason why TIFF is becoming obsolete. “It can be a challenge to view TIFFs on mobile devices,” said Harju. “At the very least you typically have to download an app. Everyone knows how to use a PDF, and PDF/raster supports everything that TIFF does. Now that the work has been done, it is easy to imagine that PDF/raster could be used outside of TWAIN Direct for any application where TIFF is currently used—bringing with it all the advantages of PDF.”

Moving toward an ISO standard

Now that the PDF/raster standard has been published, the TWAIN Working Group is moving ahead with its plans to officially launch TWAIN Direct. It’s a two-phased approach with the local area network version planned for later this year followed by the cloud version. We should have more for you following the annual **Harvey Spencer Associates** Capture Conference being held Oct. 4-5.

Meanwhile, the PDF Association will begin work on getting PDF/raster accepted as an ISO standard. [Editor’s note: *DIR* attempted to catch up with PDF Association Executive Director Duff Johnson on this topic, as well as the recent release of the PDF 2.0 spec, but he was on his way to surgery when we finally caught up with him. We are all wishing Duff a speedy and full recovery and looking forward to talking with him in the near future.

<http://bit.ly/PDF-raster-presos>; <http://bit.ly/PDF-raster-PR>

Document Imaging Report

Business Trends On Converting Paper Processes To Electronic Format

DIR is the leading executive report on managing documents for e-business.

Areas we cover include:

1. Document Capture
2. Image Processing
3. Forms Processing/OCR/ICR
4. Enterprise Content Management
5. Records Management
6. Document Output
7. Storage

DIR brings you the inside story behind the deals and decisions that affect your business.

Vol. 27, No. 16



Editor: Ralph Gammon
4003 Wood Street
Erie, PA 16509
PH (814) 866-2247
ralphg@documentimagingreport.com

DIR is published 23x per year, on the 1st & 3rd Fridays of the month, by:

RMG Enterprises, Inc.
4003 Wood Street
Erie, PA 16509
PH (814) 218-6017
<http://www.documentimagingreport.com>

Copyright © 2017 by RMG Enterprises, Inc. Federal copyright law prohibits unauthorized reproduction by any means including photocopying or facsimile distribution of this copyrighted newsletter. Such copyright infringement is subject to fines of up to \$25,000. Because subscriptions are our main source of income, newsletter publishers take copyright violations seriously. Some publishers have prosecuted and won enormous settlements for infringement. To encourage you to adhere to this law, we make multiple-copy subscriptions available at a substantially reduced price.

Subscriptions: \$597 (electronic) or \$670 (paper) per year.