Whitepaper

Using PDF to associate processing steps and content data

Author	**Jason Lisi**
	Documentation Officer, Ghent Workgroup
	jlisi@ryerson.ca
Date	May 6, 2015
Status	Final

info@gwg.org

www.gwg.org

Ghent Workgroup

Table of Contents

1 Introduction

1.1 About the Ghent Workgroup

The Ghent Workgroup (GWG) is a worldwide assembly of graphic arts stakeholders (user associations, vendors, consultants, educational institutions, service providers, and end users) founded in 2001. It was formed in response to increased needs for standardization of the different processes in graphic arts workflows, especially in an increasingly globalized service provider landscape. The rules of the group have been carefully conceived to ensure that the group remains practically oriented, and the priority is focused on the needs of the end users.

The GWG focuses on developing best practice guidelines and specifications for graphic arts workflows. While initially focusing mainly on quality control and preflight for PDF workflows in commercial print, that focus has broadened to also include metadata specifications, workflow test suites and increased support for market segments such as packaging.

All material created by the GWG is disseminated free of charge through the website of the GWG (www.gwg.org) and through the vendors and user associations partaking in the work of the group.

The mission statement of the Ghent Workgroup states that the group will "establish and disseminate process specifications for best practices in graphic arts workflows". In practice this means that the group:

- Develops and maintains process specifications and associated documentation for best practices in graphic arts workflows.
- Develops and maintains *reference implementations* to ensure the specifications it develops are usable in the real world.
- Actively promotes adoption of its work in both the graphic arts user and vendor communities.
- Streamlines and coordinates the decision process between its members.

While the group started its work developing guidelines for PDF quality control, it has expanded its scope. The group is now involved in magazine, office, and packaging specific specifications, the development of job ticket metadata specifications for delivering PDF files for advertisements, preflighting PDF files, and in developing test suites to ensure workflows and applications are configured and used correctly.

Much of the work of the group is done through teleconferences and e-mail discussions. Three times a year, the members come together for a three-day face-to-face meeting. To streamline the work and decision process, subcommittees have been organized around specific topics do the actual work. To learn more about the different subcommittees, or to find out how you can contribute to this effort, visit the Ghent Workgroup website (www.gwg.org).

1.2 About This Document

The purpose of this white paper is to provide an overview of specifications and a proposed standard to associate processing steps and content data within a PDF. This document also explores the application of this process and files that have been created to demonstrate the association of processing steps and content data within a PDF.

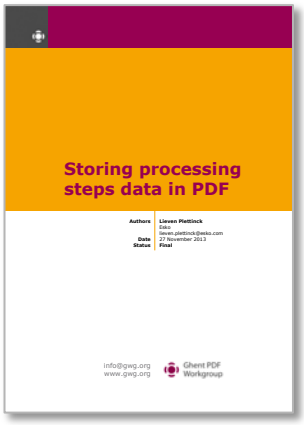
1.3 More Information

The GWG website contains a lot of useful information regarding the GWG, its specifications and deliverables, and caters to vendors, associations and end-users alike. It also features a user forum where you can interact with your peers and ask questions.

2 Background

It is common practice in the packaging and other segments of the printing industry to incorporate objects and metadata in a PDF that will not be used for printing the final product. Examples of such information would be cut lines, legends, Braille, dimensions, and so on. These data correspond to additional processing steps, i.e., steps that occur outside the normal parameters of printing.

Currently there are no standardized methods used to store this "extra" data within a PDF; instead, multiple methods are used depending on the application used to generate the PDF and the preferences of the user. For example, in one PDF cut lines may be stored in a layer called "CAD" using a contour with a stroke in spot color "Cutline", while in another PDF the cut lines are stored in the same layer as the graphics and are represented as a contour with a stroke in spot color "cut". Such variability often leads to problems of interoperability between companies and systems. Suppose a converter (the packaging equivalent of a printer) has configured their RIPs to ignore (i.e. not output) spot colour inks defined as "cut" in incoming PDF files. This would prevent an extra plate being imaged for the cut data, which of course does not print. Now what happens if the same converter receives a PDF file where the cut data is defined as the spot colour "cutline"? If the variability is not caught early, the converter risks wasting plate because the cut data will be output as if it were just another colour for the job. This is not only costly, but also wastes valuable time.



The Ghent Workgroup has developed specifications that describe a systematic method for storing data (both objects and metadata) corresponding to additional processing steps within a PDF. The latest specification, entitled Storing processing steps data in PDF, provides detailed information on how content data can be associated with processing steps in a standardized way.

The Storing processing steps data in PDF Specification can be found on the GWG website (www.gwg.org).

2.1 Associating Processing Steps with Content Data

In its simplest terms, using PDF to associate processing steps and content data means using metadata within a PDF file to define and categorize different types of processing steps within the file in a way that is standardized. Optional Content groups (OCGs) are used to group objects that belong to a specific processing step. The use of OCGs is ideal, because it allows for the labelling of objects within a PDF file in a way that makes it easy to include or exclude them at output.

The concept of layers in Acrobat (i.e. Optional Content Groups) is not new to PDF, nor is it unique to the specifications published by the Ghent Workgroup. In a similar fashion, metadata is used extensively in PDF files for a variety of reasons. What makes the GWG specifications unique is the standardized association of predefined metadata with Optional Content Groups so that the optional data can be categorically defined and understood in a standardized way.

2.2 Chronology of the Work Done By the Ghent Workgroup

The concept of associating PDF processing steps and content data has developed over time through the work of the Ghent Workgroup (GWG) Packaging Subcommittee. The original concept, dubbed "Storing non printing contour data in PDF" dates back to 2010, and eventually led to the development of the Non Print Contours Specification which was approved by the GWG in June 2011. The specification can be downloaded from the GWG website (www.gwg.org).

In fall of 2013, the GWG approved the extension of the Non Print Contours Specification to Processing Steps. A new specification entitled Storing processing step data in PDF was approved by the GWG in the fall of 2013.

With the modification of Non Print Contours to Processing steps, the GWG made a decision to further expand the specification by proposing it as an ISO a standard. In late October 2013 the Processing Steps project was accepted as a working item at the TC130 meeting in Berlin in October as ISO / PWI 19593. Since this time the TC130/WG2/TF2 has been working on the further development of the ISO standard.

2.3 Important Terms and Definitions Explained

To best understand the concepts and explanations found within this whitepaper, it will be useful to become familiar with some of the terminology that will be used throughout the document. This section defines and explains significant terms and concepts.

2.3.1 Contours

Contours are PDF paths or text objects found within the PDF file. An example of a contour would be a cut line. A series of contours can form processing steps objects (see definition below).

2.3.2 OCCD

OCCD is an acronym for Optional Content Configuration Dictionary. An OCCD "represent[s] different presentations of a document's optional content groups for use by PDF processing applications or features " (Adobe, 2006, p. 375). OCCDs are not the same as nested OCGs.

2.3.3 OCG

OCG is an acronym for Optional Content Group. An OCG is "a dictionary representing a collection of graphics that can be made visible or invisible dynamically by users of viewer applications" (Adobe, 2006, p. 364). OCGs appear as "layers" in PDF reader programs that support OCGs. A collection of OCGs can be collected into an OCCD to manage the way they are presented.

2.3.4 OCMD

OCMD is an acronym for Optional Content Membership Dictionary. An OCMD is generally used to express visibility policies that are more complex than those that are as simple as having an entire OCG state as on or off (Adobe, 2006)

2.3.5 Processing Step

A processing step is defined as a step in the production of printed products that is not the actual printing of the graphic elements. For example, die cutting a package or producing Braille objects would be considered processing steps.

2.3.6 Processing Step Optional Content Group

A processing step optional content group is an optional content group representing a processing step. A processing step is characterized by the combination of the GTS_ProcStepsGroup and GTS_ProcStepsType key in the GTS_Metadata dictionary that is part of the optional content group dictionary.

2.3.7 Processing Step PDF Object

A PDF object associated with a processing step. For example, a series of contours can form together to form dieline information within the PDF file.

2.4 Standardizing the Association of Processing Steps and Content Data

There are several benefits that can be attained from standardizing the way content data and metadata are used within a PDF. Many of these benefits can be grouped into three main categories.

2.4.1 Interoperability

The standardization of processing step/content data association within a PDF drastically improves the interoperability between different user sites, companies, and workflow systems from multiple vendors. Improving the ability for different technologies and companies/units to work together is critical for today's fast-paced, competitive, and global print economy.

2.4.2 Error Reduction

A significant benefit of standardization is predictability and consistency. Standardization of associating content data and processing steps will create a singular methodology that eliminates the guesswork of interpreting multiple, non-standardized methodologies. This will reduce errors because the outcomes of a standardized workflow are predictable and repeatable.

2.4.3 Automation

With the standardization of the association of processing steps and content data, systems and processes can be automated to interpret and process files. Such automation can reduce the number of human touch points within a workflow, increasing productivity and lowering costs.

Standardizing the association of processing steps and content data within a PDF can result in tangible cost and time savings for companies that deal with PDF files that contain non-printing elements by creating standard metadata associations for these objects that can then be used to customize various parts of workflow automation. One example of this would be creating preflight profiles that can use the metadata structure to direct the proper generation of errors and warnings.

3 An Overview of Processing Step Data

There are several different types of processing step data (PDF objects) that can be associated with a final print-ready PDF file that are not intended to print on final output. In other cases, the object may be intended to print but must be outside of the image area of the job. This section discusses different types of processing steps data that can be found in a PDF, and provides an overview of why and how this processing steps data is included in a PDF file.

3.1 Types of Processing Steps Data

What follows is a non-exhaustive list of processing step data that can be found within a PDF.

3.1.1 Braille

The addition of Braille to printed packages is becoming more and more common, especially in the pharmaceutical industry. Braille text is applied to packaging after printing. Usually this is done by a die cutter (for a complete sheet of printed boxes), a folder/gluer (for single boxes) or inkjet device.

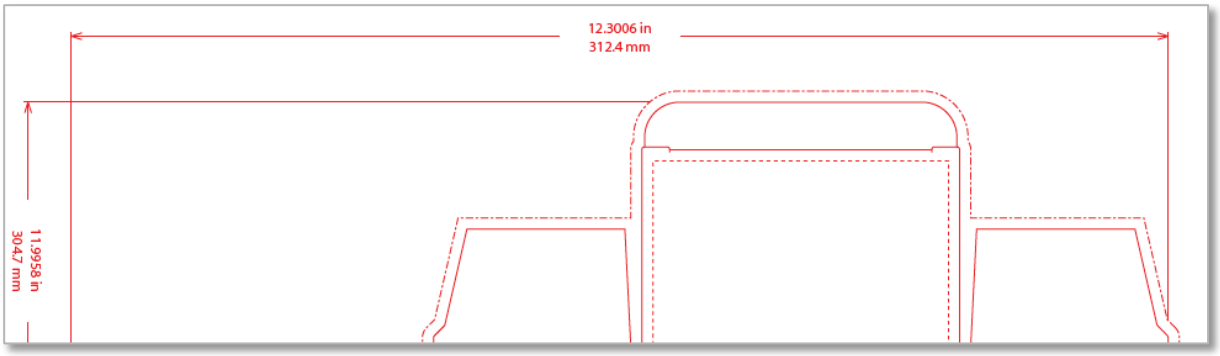
There are strict regulations that must be followed when incorporating Braille onto certain types of packaging. PharmaBraille, a company that develops and markets Braille fonts specifically for the pharmaceutical industry, offers these recommendations (adapted from PharmaBraille, 2014):

- Braille text should be set in its own colour, and that colour should not be used for anything other than Braille;
- Braille text should be constructed on its own layer (or page), separate from the artwork;
- A translation of the Braille text should be placed outside of the dieline for quality assurance purposes.

Braille text is considered processing step data because the braille text in the file will not print via traditional printing, and will instead be applied as an extra step. In addition, the translation of the Braille text is required for preflighting purposes during the proofing stage, but must be outside of the image area and usually does not print on press.

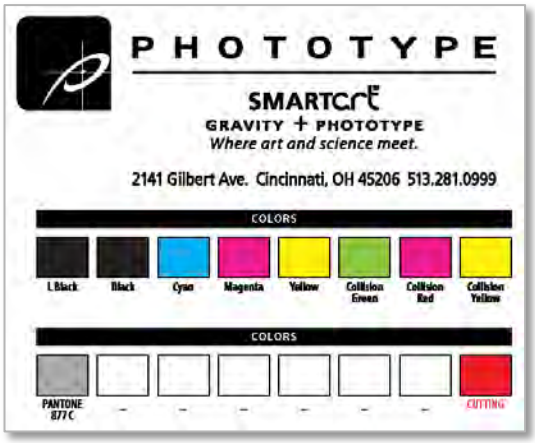
3.1.2 Dimensions

Dimensions are a collection of indications (arrows, numbers and units) of sizes of items in the design. They are used for quality assurance purposes and do not print.



Just like the Braille translation, dimensions must be placed outside of the image area, and are often included during the proofing stage for verification purposes. Dimensions do not print on the final product.

3.1.3 Legend



A legend is defined as an informational panel containing job information that is positioned outside the dieline.

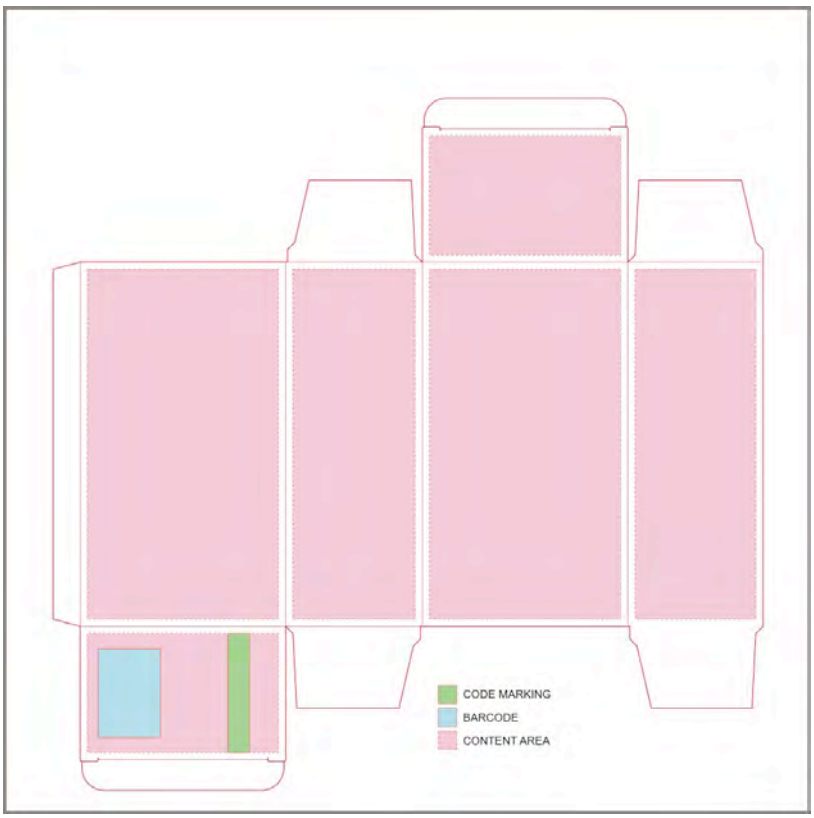
A legend contains reference material that is relevant to the job, such as design identification, customer information, and technical specifications such as colours and project (job name) etc.

A legend will usually be included on both the proof and may or may not be included on press.

3.1.4 Positions

Positions are PDF processing step data that identifies intended or allowed positions for objects such as text, barcodes, holograms, and so on. The image below shows an example of three different types of Positions: code marking, content area, and barcode.

Positions are marking used for informational purposes only and do not print on the final product.

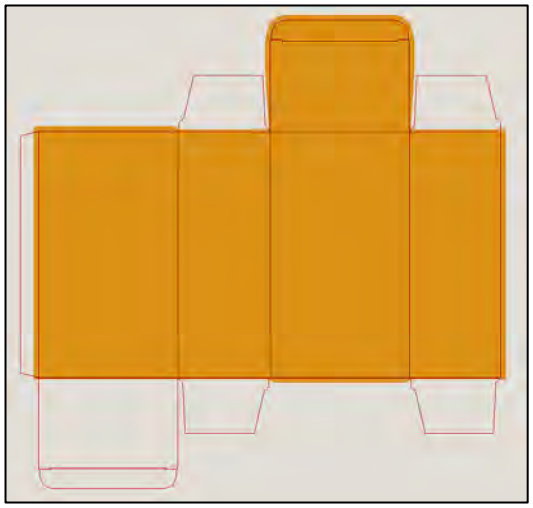


3.1.5 Structural Data (CAD)

Structural Data (dielines) refers to a collection of contours in a PDF document that describe how the label or package will be finished (e.g. cut, folded, glued, embossed, etc.) to produce the end product from the printed substrate. These contours correspond to operations performed on finishing devices like die cutters and folder-gluers. In some cases finishing can result in a simple, flat end product, such as a diecut label. Other times, as demonstrated in the image below, the end product can be a complex three-dimensional shape. Structural data is usually printed on a proof to verify the proper finishing of the product, and is often used to generate the physical dieline used to create the end shape.

3.1.6 Varnish

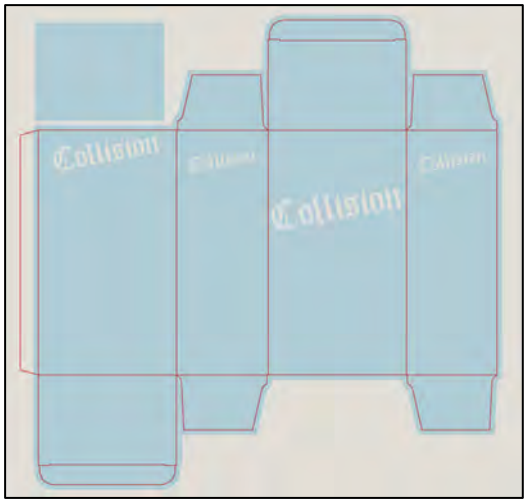
Varnish processing step data is a collection of PDF objects indicating application of varnish after printing.



This image shows the outlines of a folding carton. The orange color indicates areas where varnish will be applied after printing of the actual graphics. Varnish processing step data is used for visualization but does not print as shown. Since this is typically a finishing treatment that is printed, or sprayed on, and is more often clear in nature. There is not a current method to display this information in a non-destructive method.

3.1.7 White

This defined processing step data category represents a collection of PDF objects indicating application of white backing ink on transparent or metallic substrates.



This image shows the outlines of a folding carton printed on a metallic substrate. The greyish background represents the metallic surface. The cyan color indicates areas where opaque white ink will be applied before printing of the actual graphics.

Consequently, although the "Collision" logo views as white in the images in section 3.15 of this document, it is actually the bare metallic substrate without the opaque white ink behind it.

This is one instance where the processing step data would actually print as a separation on press.

3.2 Why Embed Processing Step Data in the PDF?

Some may argue that the surest way to ensure that non-print associated with the design of a project does not print is to exclude it from the file entirely. While this approach would certainly work, in most cases it is impractical and would lead to other potential problems. Having processing step data embedded in a PDF makes it easy to validate structural components of design through conventional proofing methods. Overlaying processing step data with design elements in the same file enables precision alignment of design elements and makes preflighting all aspects of the job easier.

A PDF can serve as a container for all production data, printing and non-printing, of a package, label or any other print file. When done correctly, embedding processing step data into a PDF has many advantages.

3.2.1 Graphic Placement

One of the main advantages of having processing step data in the PDF is to give the graphic designer (creative designer or production artist) a context for correct placement of graphic elements. By providing structural data, dimensions, and positions within a PDF file, artwork can be easily (and precisely) placed into the correct position, with the correct orientation, rotation, and sizing, while allowing for compensation for such things as bleeds and ink-free areas.

3.2.2 Quality Control and Approvals

Processing step data can play an important role in quality assurance and job approvals. For example, structural data embedded in PDFs allows for the verification that the correct die shape was used, and that graphic elements were correctly placed relative to the structural contours. Braille, and in some cases translation text, is included in the PDF using an artificial spot color to make it visible for non-visually impaired QC people. Legends and dimensions are added to give extra context and information for the quality control and approval processes.

Applications like Adobe Acrobat foster quality assurance with the ability show or hide structural contours in the Layers Panel. In addition, processing step data can usually be printed on hard copy proofs made for quality control and approval purposes.

3.2.3 Layout and Tool Manufacturing

The structural data embedded in the PDF of the single artwork file can be used by automated layout applications to fill a sheet or web by repeating the artwork multiple times. This can be a layout of graphics on a sheet or web to correspond with an analog die (e.g. folding carton and labels), or it can also be an optimized nesting algorithm to minimize material waste for digital finishing (e.g. cutting table or laser cut device).

From the imposed layout of the graphics on the sheet or web, the position of the braille elements is determined and that information is used to manufacture the braille dies.

Similarly, the position and shape of varnished or coated elements can be derived through processing step data to make a varnish printing plate or to cut a coating blanket. Similar logic holds true for foils.

3.2.4 Digital Finishing

The structural data embedded in a PDF can be used via automation to drive digital finishing machines. Similarly, Braille data can be extracted from the PDF to drive inkjet braille machines.

4 How to Store Processing Step Data in a PDF

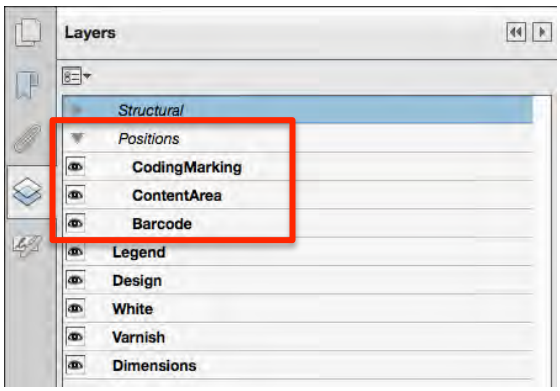
The GWG Storing processing step data in PDF Specification provides detailed information on the technical requirements for storing processing step data in PDF. This document summarizes the key points of the specification document and provides an overview of how processing step data is stored in PDFs. Readers wishing to explore the in-depth technical details are encouraged to read the specification found on the GWG website (www.gwg.org).

4.1.1 The Role of OCGs in Storing Processing Step Data

Just like layers are used in design software to group different elements together and affect their visibility, processing steps data is stored in optional content groups in such a way that they can either be included or excluded from output simply by altering the visibility (turning off or on) of Acrobat layers. Processing step objects of the same kind/functionality are stored in the same OCG (a functional OCG). This grouping in functional OCGs of related objects simplifies the process of layer selection for human users and automated processes by allowing all data of the same processing step type to be turned off or on together as well as individually.

Using OCGs to organize processing step data is only part of the procedure. In addition, the processing step data must be properly identified and categorized so that their purpose and characteristics are easily recognizable by both humans and machines performing automated processes. This is accomplished by using a standardized set of metadata tags that are assigned to the processing steps data. In other words, first the processing step data is physically organized into "layers", and then it is tagged with the appropriate standardized metadata. It is this two-step approach that results in a highly functional use of processing step data.

The following is an example of how processing step data would look like as layers (OCG – Optional Content Groups – see section 4.1.3) in Acrobat, and shows the standardized metadata used to categorize and group the processing step data.



Name="**Positions**" GTS_ProcStepsGroup= "Positions"

Name="**CodeMarking**" GTS_ProcStepsGroup= "Positions" GTS_ProcStepsType= "CodeMarking"

Name="**ContentArea**" GTS_ProcStepsGroup= "Positions" GTS_ProcStepsType= "ContentArea"

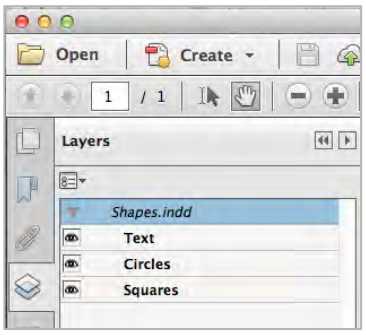
Name="**Barcode**" GTS_ProcStepsGroup= "Positions" GTS_ProcStepsType= "Barcode"

4.1.2 Coloring and Positioning of Processing Steps Objects

In addition to using OCGs to define and categorize processing step data, the positioning and colour of processing step objects must also be considered. In particular, it is important that the processing step data does not interact with the design objects in such a way as to modify their physical attributes – i.e. knocking out the design objects below, participating in transparency blending and so on. For some processing steps objects, this means they have to be placed outside the bleed area so they do not impact the design (e.g. Legends, Braille translations, etc.). Processing step data that must overlap with design elements (e.g. Braille text, dielines, etc.) need to be coloured with a unique spot colour (i.e. not a spot colour used in the design elements), and set to overprint. This will ensure that this data does not affect the rendering of printing objects or inadvertently be output on one of the printing separations.

4.1.3 OCGs vs. OCCDs vs. Layers

Understanding the difference between layers, optional content groups, and optional content configuration dictionaries can be a challenge, especially when most users of PDF only see "PDF Layers". Nonetheless, a basic understanding of these terms is useful when dealing with processing step data.



When viewing "layers" in a PDF reader, what is really being shown is Optional Content Groups (OCGs). Each layer represents an OCG. The image to the left shows a PDF with 3 OCG's: Text, Circles, and Squares.

Where things get a bit confusing is with the concept of optional content configuration dictionaries. Rosenthol (2014) states that an OCCD "represents a preset configuration of the state for one or more [optional content] groups. A PDF (that has OCGs) may contain several OCCDs, but must include at least one" (p. 155).

In other words, the OCCD deals with the presentation of OCGs for use in PDF viewers, including the default state or configuration of these OCGs (aka Acrobat Layers) when the document is first opened.

5 Processing Step Data in Practice: The "Collision Colours" Sample Files

In this section, we will look at some sample files created by the GWG that conform to the Storing Processing Steps Data in PDF Specification. The purpose of these files is to provide working sample of conforming PDFs that can be used for education, testing, and verification. The sample files can be downloaded from the GWG website (gwg.org).

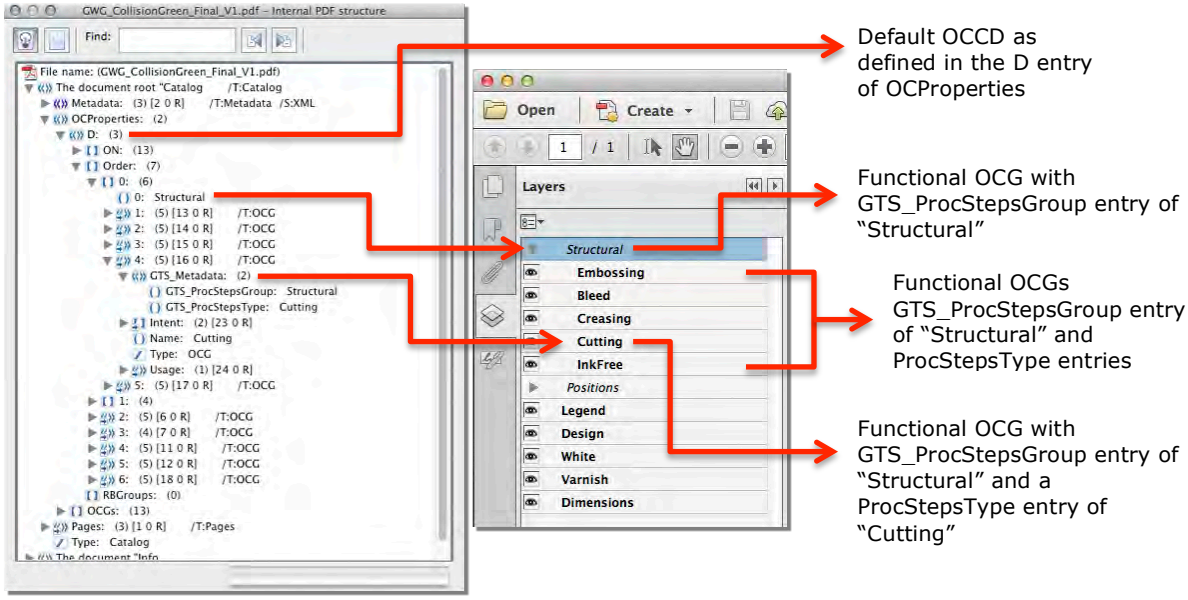
5.1 The Intent of the "Collision Colours" Sample Files

The intent of the "Collision Colours" sample files is to demonstrate conformance to the GWG Storing Processing Steps Data in PDF specification. The "Collision Colours" sample PDF files contain processing steps data within processing steps OCGs with standardized metadata tagging. Each one of the three files has the following characteristics pertinent to the specification:

- A Design OCG containing data for printing
- Multiple Functional OCGs for processing step data
- Independent OCGs that are presented as a group in the PDF reader
- A default OCCD
- Multiple types of Processing Step Data

Users are encouraged to process these files within their workflows to determine the extent to which different workflows can identify and utilize the association of processing steps and content data.

The diagram below shows the association of the OCCD and OCGs within the "Collision Colours" sample files.

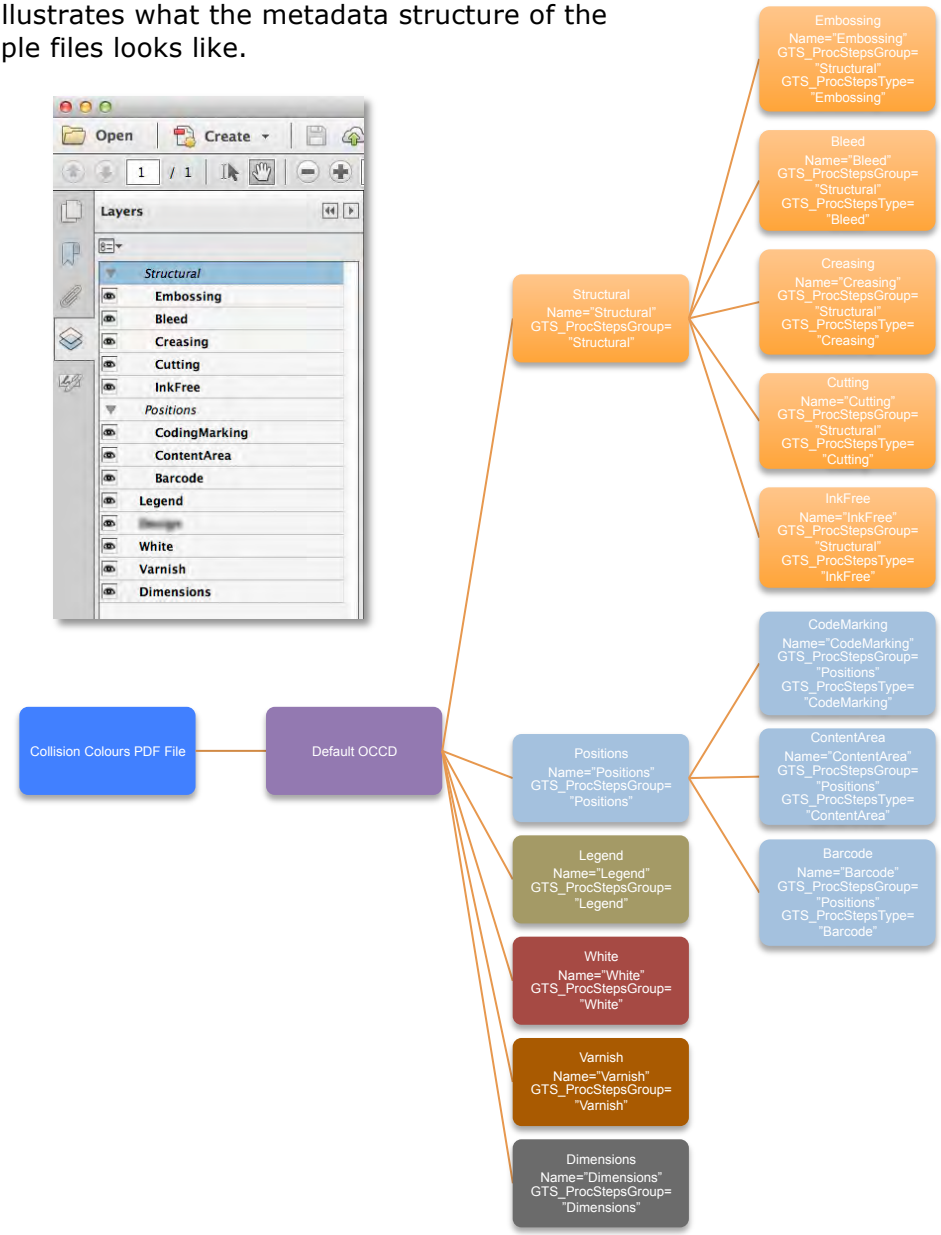


5.2 Content of the "Collision Colours" Sample Files

Each of the three "Collision Colours" sample files contains the following content associated with processing step data:

- Structural:
  - Embossing
  - Bleed
  - Creasing
  - Cutting
  - InkFree
- Positions:
  - Code Marking
  - Content Area
  - Barcode
- Legend
- White
- Varnish
- Dimensions

The chart on the right illustrates what the metadata structure of the "Collision Colours" sample files looks like.



5.2.1 Consuming the "Collision Colours" Sample Files

The "Collision Colours" sample files were created to provide workflow vendors with files that conform to the Storing processing steps data in PDF GWG specification so that they can configure their systems to take advantage of the association between processing steps and content data. Using these files, workflow vendors should be able to configure their systems to use the processing steps to automatically:

- Generate printing plates for design content without generating plates for processing steps;
- Make a proof with the processing data printed on the proof;
- Extract data for such things as white backing ink and varnishes; and,
- Perform other necessary functions relating to process steps.

In addition, these files can be used by preflight software vendors to derive appropriate preflight profiles that can be used to validate PDF files against the specification.

6 Conclusion and Further Information

6.1 Conclusion

The GWG has developed a specification that standardizes the association of processing steps with content data to improve workflows where content is included in a PDF that does not pertain to the printing of the design elements. This specification presents obvious benefits to the packaging printing market; however, other markets segments can also benefit from this standardization.

Presently, the TC130 is developing an ISO Standard (ISO 19593 - Use of PDF to associate processing steps and content data) to expand the reach and functionality of the original GWG approved specification.

To assist industry with understanding and conforming to the standardized association of processing steps and content data, the GWG has developed three sample files that can be used for testing and development within various workflows.

The standardization developed by the GWG makes it easy to preflight and validate conformance of files through industry standard preflight tools.

6.2 Further Information

Further information about the Ghent Workgroup and the Packaging Sub-committee can be found at www.gwg.org.

For more information about the International Standards Organization (ISO), please visit http://www.iso.org.

Background information in the ISO/TC 130 Graphic Technology committee can be found at
http://www.iso.org/iso/home/standards_development/list_of_iso_technical_committees/iso_technical_committee.htm?commid=52214

7 References

Adobe Systems Incorporated (2006) *PDF Reference (6th Ed.): Adobe Portable Document Format Version 1.7.* USA, Adobe Systems Incorporated

ISO International Standard Organization (In Progress) ISO *19593 Graphic technology — Use of PDF to associate processing steps and content data.* Switzerland, International Organization for Standardization

PharmaBraille (2014) *Procedures for Pharmaceutical Braille Artwork.* Retrieved from http://www.pharmabraille.co.uk/pharmaceutical-braille/pharmaceutical-braille-procedures.htm

Rosenthol, L. (2014) *Developing with PDF: Dive Into the Portable Document Format.* Sebastopol, CA: O'Reilly Media Inc.

Plettinck, L. (2013) *Storing processing steps data in PDF.* Unpublished Manuscript