



GLOBAL GRAPHICS®
software

When Content Matters

Capturing the richness of page description languages

OctoberPDFest
ONLINE

Who am I?

- David Stevenson
 - Product Manager
 - Global Graphics Software
- This presentation will refer to Mako
 - Mako is a “Swiss Army knife”, a toolbox for creating, editing, converting and rendering PDF and other PDLs



Why it matters?

- We have customers who need to extract information from print streams created in page description languages (PDLs)
- It might be as simple as knowing if a job is monochrome or colour
- Or to populate a content management system (CMS) with PDFs
 - Content that can be indexed and later searched
 - Content that can be tagged with metadata (eg XMP)
 - Add a visible identifier to a page such as a UID or QR code

Why bother?

Extracting useful information from a print stream applies to various scenarios

- Re-use:
 - Searching content for specific terms
 - Reverse engineering of transactional content to populate a database
 - e-discovery, identifying document content for legal purposes
- Regulatory pressure, eg GDPR
 - Ensuring private data is kept private
 - Redaction of sensitive information
- Job costing
 - Accurate billing based of colour vs. mono, simplex vs. duplex etc.
 - Correct identification for client billing
- As part of a print workflow or when processing legacy documents

Use case (Racami)

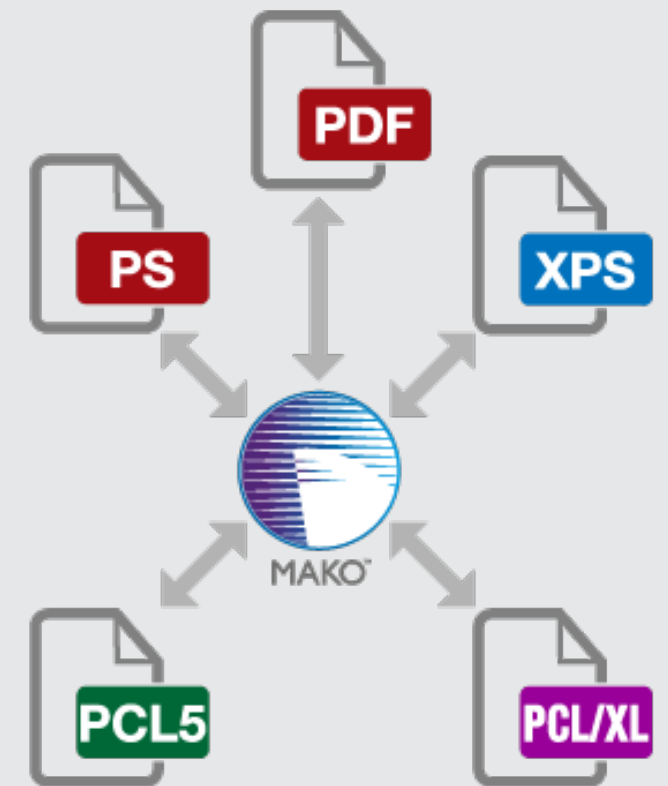


- Racami is the author of the Alchem-e™, used by companies to perform multi-channel communications
- Mako is used to “normalize” content by converting multiple PDLs (PS, PCL) into PDF/X as part of a CCM workflow
 - Allows content to be indexed and later searched
- Mako is also used for combining thousands of transactional PDFs into one for printing



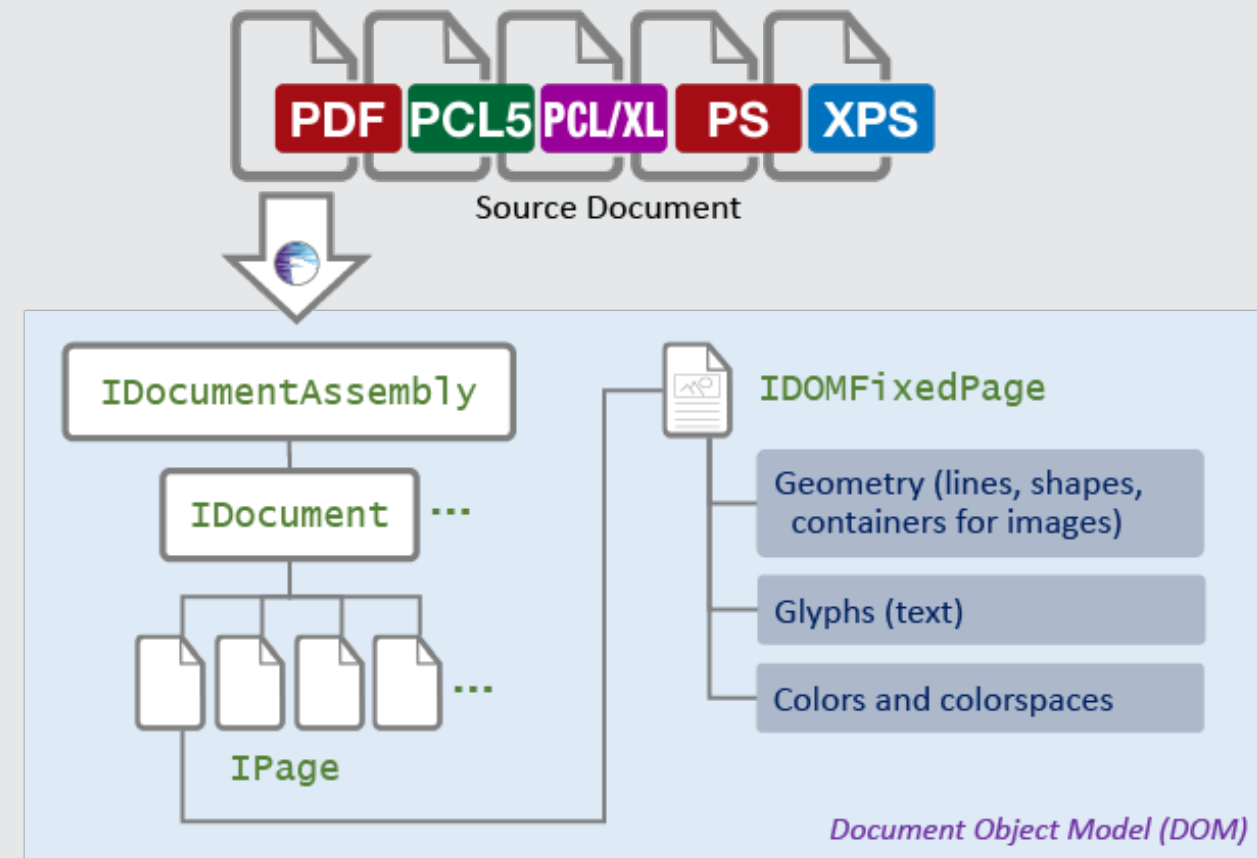
“Our customers can rest easy knowing that when it comes to Postscript, PCL, and PDF processing, they have the combined expertise of Racami and Global Graphics.”

- Anything in PostScript can be represented in PDF
 - So nothing is lost in conversion to PDF
- For other page description languages (PDLs) it's not so easy
 - PDLs such as PCL are concerned only with the representation on the page
 - A different approach is required to acquire useful text and data
- Mako is a software library from Global Graphics that employs this approach to PDL conversion



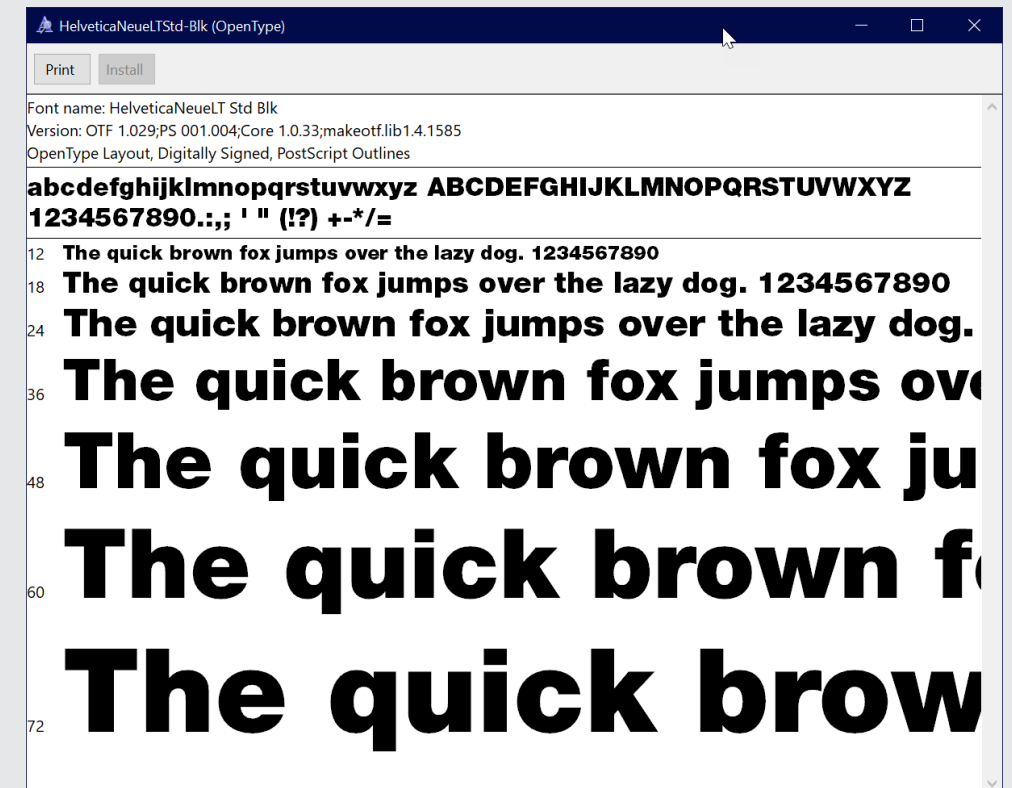
Document Object Model

- Parsing PCL & other formats
 - Treated as a ‘recipe’ for making a page
 - Mako API calls construct a DOM (Document Object Model)
 - DOM represents document & page content in a format-agnostic way
 - DOM can store metadata including print instructions
 - Process PDL to populate

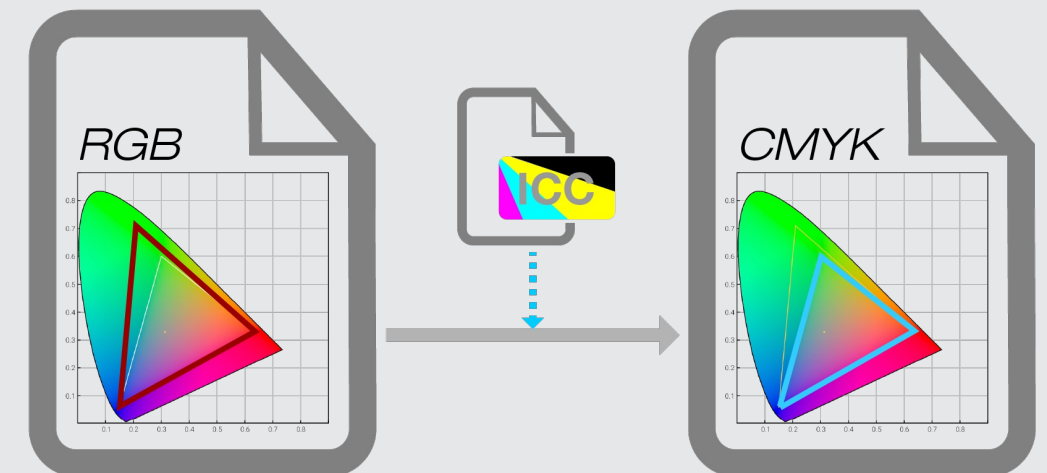


Fonts and text

- All fonts represented as OpenType or Type 3
 - Map bitmap fonts to Type 3
- Identify symbol sets
- Encoding
 - Glyph in the right position?
 - ToUnicode()
- Not always possible for downloaded fonts
 - Could apply OCR to identify glyphs



- PCL5 and PCL/XL have an unsophisticated color model
 - RGB or Gray
 - Convert to a target space in DOM or specify an output intent when saving to PDF
- PCL ROPs (Raster Operations) enable colors to interact
 - Processed during input phase
 - Can be ignored to improve performance if text (or image) extraction is primary goal



- Do you need to build fast, scalable solutions for your print workflow?
- Then you need **Mako**
- Like a Swiss Army knife, Mako has many 'blades', each one matching a job that you need to get done
- Utilizing its unique document object model, Mako's C++ and C# APIs offer control over color, fonts, text, images, vector content, metadata and more, combining precision with performance





Any questions?

david.stevenson@globalgraphics.com