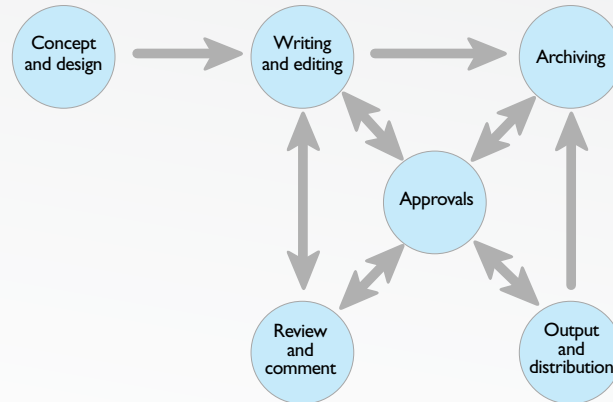


DIFFERENT FORMATS FOR DIFFERENT JOBS



By Martin Bailey
Chief Technology Officer
Global Graphics Software

▷ Document lifecycles can be very complex, but if you strip away the infrastructure used to store or exchange them, most can be represented reasonably accurately as something like this diagram.



Most text-centric documents created in an office environment are written in Microsoft Word, Open Office, or a similar creation tool. Many documents are also circulated around departments and between companies as Word “.doc” or “.docx” or as Open Office “.odf” files. I'd like to explore with you whether that's always the best choice and to look at alternatives. For convenience I'll use the term “authoring formats” to refer to the native file formats used by Word, Open Office and other office document creation tools.

Clearly it makes sense to use an authoring format for those parts of the workflow where the document is being actively edited. If you're passing a document to a colleague and want them to change the text as they see fit, you should send them a suitable file that both of you can work with. But does it also make sense to send an authoring format to somebody in another department or company, or as a submission to an external organization such as a government agency?

To a large extent that depends how you expect them to use the information. If the content is important, but the layout and design are not, perhaps for a press release, then send an authoring format so that the recipient can re-use the text as they want. You'll need to be a bit careful if you're writing in a language that makes extensive use of accented characters or a non-Latin script to ensure that they can access the text properly, but that's easier with modern document applications than it was in the past.

But if you're sending a document where the appearance matters as well as the content, then is an authoring format still appropriate? “Appearance” here includes the look and feel, perhaps for a product brochure, but it also encapsulates an assurance that everyone who reads the document sees the same text on the same lines on the same pages. If you're going to refer to a particular position in the document then you'll want everyone to associate that with the same sentence.

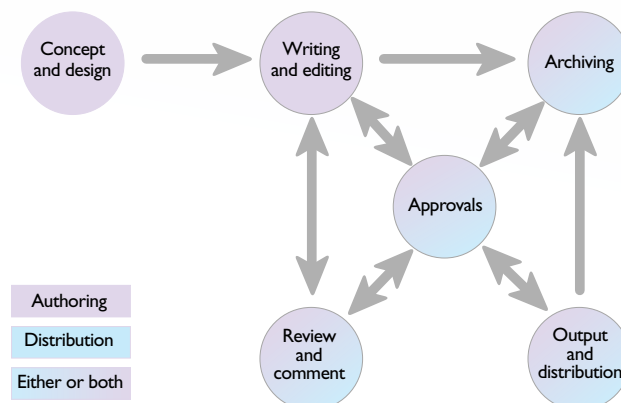
In the past many word processing applications would re-flow documents automatically depending on whether you had all of the right fonts installed on your computer; or on your default printer selection. These days that kind of silent sabotage of your work is far rarer; but there are still similar risks associated with exchanging documents in authoring formats:

- The most common way of opening an authoring format to read or print it is to open it in the same application that it was created in. That normally means that the reader has full editing capabilities, and it's rather too easy for them to accidentally change something while rushing to print the file before their next meeting
- If somebody trying to read the document has a different word processing application, or even a different version of the same one (such as Microsoft Word versions 2003 and 2007) then the text may flow differently in those different programs, usually without any warning that there may be a problem
- If you work internationally some people who receive your document may print it to letter sized paper; while others will print on A4. Some applications will re-flow the text, while others will take the safer route of scaling the whole page without changing line-ends.

At this point it's probably useful to try to construct a description of an ideal format for distributing documents so that both the content and the appearance are correctly reproduced. It should:

- not accidentally change in any way, so you can be sure that "line 5 on page 23" identifies the same text to everyone who reads it
- be difficult (or even impossible) to edit, so that you have some confidence that recipients can't accidentally alter your original document
- display on a computer monitor and print on a printer at high quality
- not be larger than necessary so that it does not take longer than necessary to transmit as an email attachment or download from a web page and does not eat up your disk space too fast
- be possible to create, read and print on a wide variety of printers using inexpensive tools.

Fortunately there are at least two formats that fulfil all of those requirements: the Portable Document Format (PDF, published as International Standard 32000) and the XML Paper Specification (XPS, published as ECMA-388). Both PDF and XPS are therefore good choices as distribution formats. Word and Open Office files do not, but they do work very well as authoring formats.



So now we can re-draw the simple workflow diagram that I showed up above, this time drawing a distinction between those processes that are best done with authoring formats and those done with distribution formats:

I've set out the advantages of using different formats for different parts of the workflow here, but there is a potential cost as well. Do you now need multiple applications on your desktop to ensure that you can deal with them efficiently? Fortunately there are tools to help you there. Make sure that you can work with your files in the way that you need to; visit Global Graphics' gDoc Fusion and take it for a test drive.

December 2009



www.globalgraphics.com

Global Graphics Software Inc.

31 Nagog Park, Suite 315, Acton
MA 01720, USA
Tel: +1-978-849-0011
Fax: +1-978-849-0012

Global Graphics Software Ltd

2nd Floor; Building 2030
Cambourne Business Park
Cambourne, Cambridge
CB23 6DW UK
Tel: +44 (0)1954 283100
Fax: +44 (0)1954 283101

Global Graphics KK

Level 14, Hibiya Central Building
1-2-9 Nishi-Shimbashi, Minato-ku
Tokyo 105-0003, Japan
Tel: +81-3-5532-7340
Fax: +81-3-5532-7373