

TONY DAVIES COLUMN

Adobe PDF use in analytical information storage and archiving

A.N. Davies

External Professor, University of Glamorgan, UK, Director, ALIS Ltd and ALIS GmbH—Analytical Laboratory Informatics Solutions

Background

*There are more things in Heaven and
Earth Horatio,
Than are dreamt of in your philosophy.¹*

Yes, the children are now studying Hamlet at school and this seemed an appropriate quote to start of this edition.

When we set up ALIS GmbH one of the first major “discoveries” was probably the most embarrassing for me. Having worked on analytical data standards for so long, I seem to have successfully generated a blind spot for the developments which have taken place in the structure, standardisation and functionality of the Portable Document Format (PDF).² Maybe it’s due to a subconscious aversion to what I had for a long time seen as a simplistic “get out” solution for those too lazy to convert data into a long-term, stable, vendor-neutral format. How often have we heard the “well... we just print to PDF” as an excuse for not having in place a properly thought through analytical data storage and archiving policy taking no account of the future use to which that data may well be put within an organisation.

Anyway, it has been pretty difficult for me to admit that my knowledge of the available functionality lay somewhere back in the early 1990s (see Figure 1) but I hope in this column to make some amends!

We will look at the PDF file format and standardisation by international standards bodies. On the functional side we will look at:

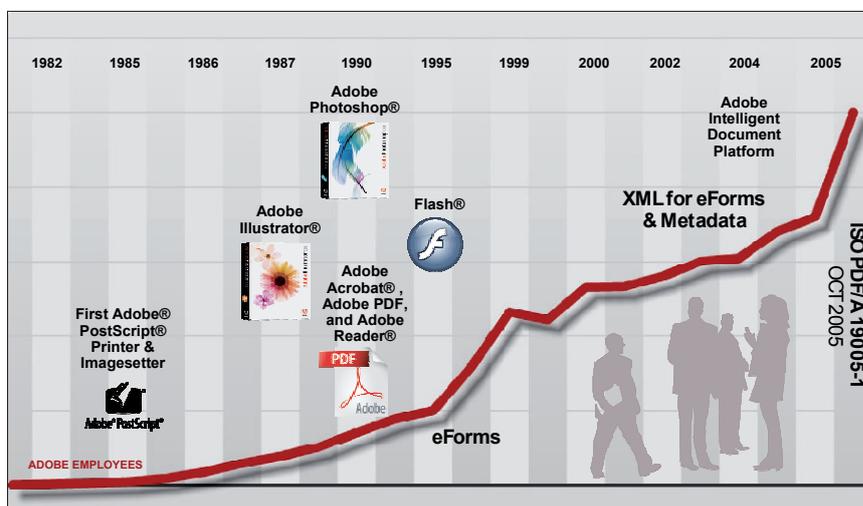


Figure 1. Adobe historical company development leading to ISO standardisation of PDF/A for Document Archiving.³

- How the PDF document can work in the Analytical Laboratory environment not only as a stable document format but also communicating with other computer systems such as a LIMS or ERP software suite such as SAP, essentially making the PDF form into a database data entry and reporting tool.
- Integrating spectroscopic and other analytical data into a PDF document.
- The use of so-called rights enabled PDF documents within the free Adobe Reader software package providing additional functionality such as editing and saving content.
- In addition I will briefly outline how a PDF document can work in an analytical laboratory environment pulling and

pushing data like a database front-end system.

- To finish off I will very briefly touch on document security and digital signatures.

History

The “PDF format” is actually a series of formats which have built on each other with increasing (and sometimes specifically limiting) functionality (see text box).

The PDF format was originally intended to provide a platform and generating software for independent documentation representation. So whatever computer operating system is your personal favourite from Apple Macs, Microsoft Windows or UNIX-based etc., you can receive, view and print documents from

TONY DAVIES COLUMN

Formal agreed ISO PDF Standards

PDF/X for the printing and graphic arts. The guidelines for developing the PDF/X standards were published as *Graphic technology—Prepress digital data exchange—Guidelines and principles for the development of PDF/X standards* ISO 15929:2002 and a succession of PDF/X standards were agreed under the general ISO 15930 numbering with increasing levels of functionality and allowed graphical types and handling options.

- ISO 15930–1:2001–Part 1: *Complete exchange using CMYK data (PDF/X-1 and PDF/X-1a)*
- ISO 15930–3:2002–Part 3: *Complete exchange suitable for colour-managed workflows (PDF/X-3)*
- ISO 15930–4:2003–Part 4: *Complete exchange of CMYK and spot colour printing data using PDF 1.4 (PDF/X-1a)*
- ISO 15930–5:2003–Part 5: *Partial exchange of printing data using PDF 1.4 (PDF/X-2)*
- ISO 15930–6:2003–*Complete exchange of printing data suitable for colour-managed workflows using PDF 1.4 (PDF/X-3)*

PDF/A for archiving in corporate, government, library etc. environments as ISO 19005–1:2005 published on October 1, 2005. *Document Management—Electronic document file format for long term preservation—Part 1: Use of PDF 1.4 (PDF/A-1)*. A key element to this reproducibility is the requirement for PDF/A documents to be 100% self-contained.³

- Audio and video content are forbidden.
- Javascript and executable file launches are prohibited.
- All fonts must be embedded and also must be legally embeddable for unlimited, universal rendering. This also applies to the so-called PostScript standard fonts such as Times or Helvetica.
- Colourspaces specified in a device-independent manner.
- Encryption is disallowed.
- Use of standards-based metadata is mandated.

Additional PDF standardisation under development targeting specific application areas:

PDF/E currently under development for exchange of engineering drawings

PDF/UA also under development for universally accessible PDF files.

PDF/H (PDF for Health Care Industry). Under development for the storage and transport of patient care data such as:

- Diagnoses
- Lab reports
- Pictures
- X-Ray and Computer Tomography data
- ECG and EEG Information etc.

colleagues who prefer to work in one of the other environments without worrying whether the document you are viewing is actually showing the same image as the original author intended.

Although the PDF format is proprietary, it is also open and well documented allowing other software vendors to work with PDF formatted documents without resort to licensing the format from Adobe.

A number of versions of the PDF format have been and continue to be adopted as ISO standards (see text box). Adobe published the first PDF standard in 1993 and since then the standard has been enhanced with additional functionality with almost every major release of the Adobe Acrobat software. The organisation coordinating these ISO efforts is called AIIM and was orig-

inally founded in 1943 as the National Microfilm Association. It later became the Association for Information and Image Management (AIIM) before mutating in a similar manner to ASTM by dropping its original title in favour of the initials and is now known as "AIIM—The Enterprise Content Management Association".⁴ AIIM is accredited by ANSI (American National Standards Institute) as a standards development organisation and holds the Secretariat for the ISO (International Organization for Standardization) committee focused on Information Management Compliance issues, TC171.⁵

OK, so PDF is now an internationally recognised documentation standard, but how can that help us in the analytical field? Well let's look at the internal workings of the PDF file.

PDF structure

What surprised me most when learning about how far PDF formats had come while I wasn't watching, was to learn about the intelligence underlying the visible document which we are used to seeing. I could write for several hundred pages on the structure of the PDF document but as a picture paints a thousand words you will be spared! Figure 2 shows a schematic representation of how a modern PDF file is built.

The presentation layer is essentially what we see either on the computer screen or when printed out. This may be only a fraction of the information actually stored within the PDF file down at the XML level as the display is controlled in part by specific business rules.

In an analytical laboratory this may well mean that a single PDF form contains data entry fields for all the possible information to be captured for all the different types of analyses carried out in the company. A very simple PDF presents itself to the user when a new work order is first opened... say just enough information for the user to identify themselves and their group. This information can be used to control the presentation layer restricting the available fields to those relevant to the actual user. A user registering a sample for analysis in the NMR laboratory would, for example, only see the

TONY DAVIES COLUMN

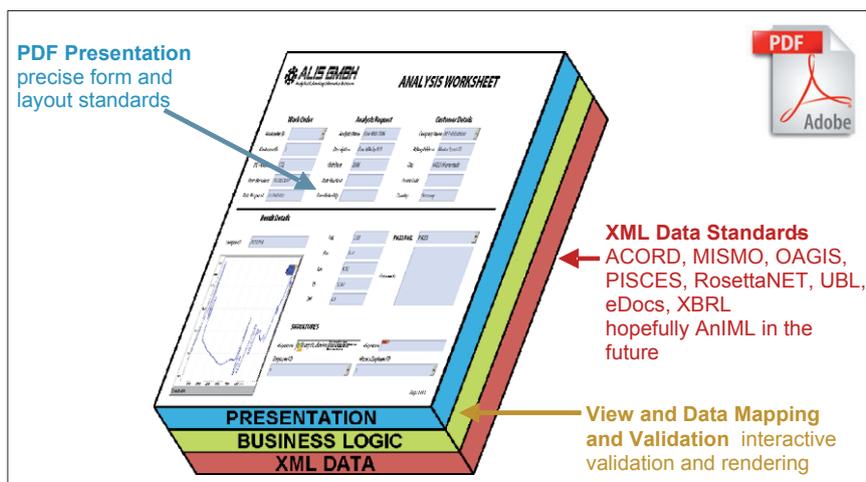


Figure 2. PDF document structure can be explained in a three tier model of a Presentation Layer (what you see and print), Business Logic Layer which contains the intelligence and possible interconnectivity information making the PDF Document into a flexible configurable expandable object and finally the XML layer.³

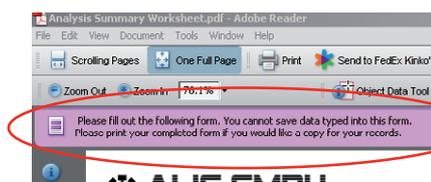


Figure 3. A standard eForm in Adobe reader can be filled out but the only way to store the data entered is to print it.

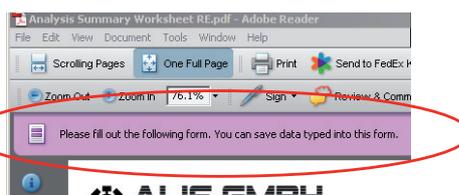


Figure 4. A rights-enabled PDF form can be completed in Adobe reader and also saved to disk with the data entered. This is indicated by a different message to the user within Reader than in Figure 3 showing that a rights-enabled form is being viewed.

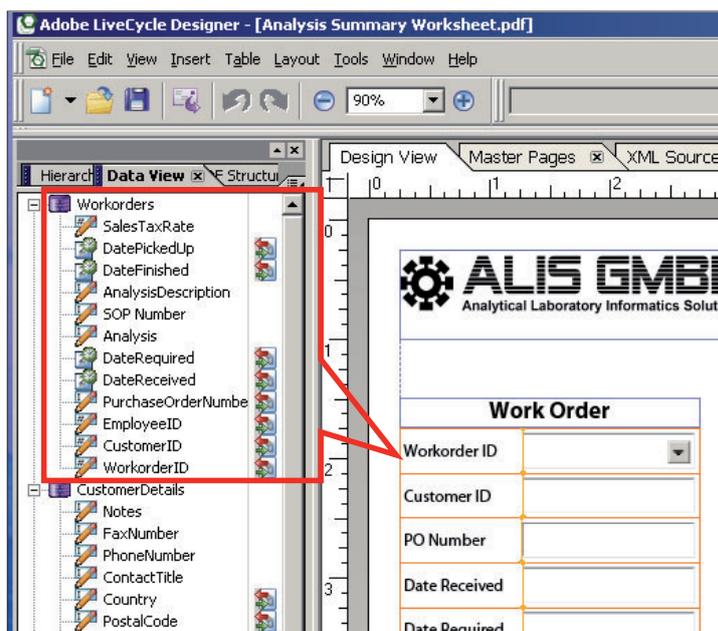


Figure 5. During PDF form development the individual fields can be automatically linked to external IT systems—in this example to an ERP system for administering contract analytical laboratory work.

fields appropriate for the available NMR experiments.

Four levels of functionality increase

The simplest deployment of the available technology would be the classic eForms solutions where the PDF Reader displays boxes to be filled in without any additional rights enabled. The data to be entered has the advantage of being structured. However, with no inherent capability of saving the file with the data or connectivity to another IT system the only way to store and communicate the information further is to print out a hard copy of the file which would then need re-entering to any receiving system (Figure 3).

The next level up is to design the form so that the boxes can validate the data being entered. This brings benefits in that the data in the printout is correctly structured—where a sample code in your organisation's specific format should have been entered you can ensure that it has the correct form, eliminating some simple typing errors.

The larger returns on investment come with rights-enabled forms where the form can now be completed, entries validated and maybe digitally signed and then saved. Rights-enabling a form is done centrally and no additional software needs to be deployed to the client PCs where the Adobe Reader is already in place.

Finally, the top level of functionality which can be deployed to make your life easier is where the individual boxes are linked to other IT systems. For example, where information about the client who has requested the analyses is available in another database system, don't waste time re-entering this data into the form for every sample they submit but connect to your client database and select the client using a drop-down list box which then triggers the automatic population of the other client fields. This can work in both directions—for a new client—enter the details in the respective fields and upload to the database for re-use later on! (See Figure 5.)

And the spectra?

You can work with the spectroscopic data in the same manner. In the form shown

TONY DAVIES COLUMN

ALIS GMBH
Analytical Laboratory Informatics Solutions

ANALYSIS WORKSHEET

Work Order		Analysis Request		Customer Details	
Workorder ID		Analysis Name		Company Name	
Customer ID		Description		Billing Address	
PO Number		Unit Price		City	
Date Received		Date Finished		Postal Code	
Date Required		Date Picked Up		Country	

Result Details	
Sample ID	3132764
Fat	
Pro	
Lac	
TS	
Snf	
PASS/FAIL	
Comments	

SIGNATURES	
eSignature	
Employee ID	
eSignature	
Witness Employee ID	

Spectrum

Page # of ##

Figure 6. In this reporting form an image field (circled) has been added and can be populated with a graphical snapshot taken from the standard spectroscopy software package. With the eForm rights enabled clicking in this field will open a dialog box where the appropriate spectra image can be uploaded.

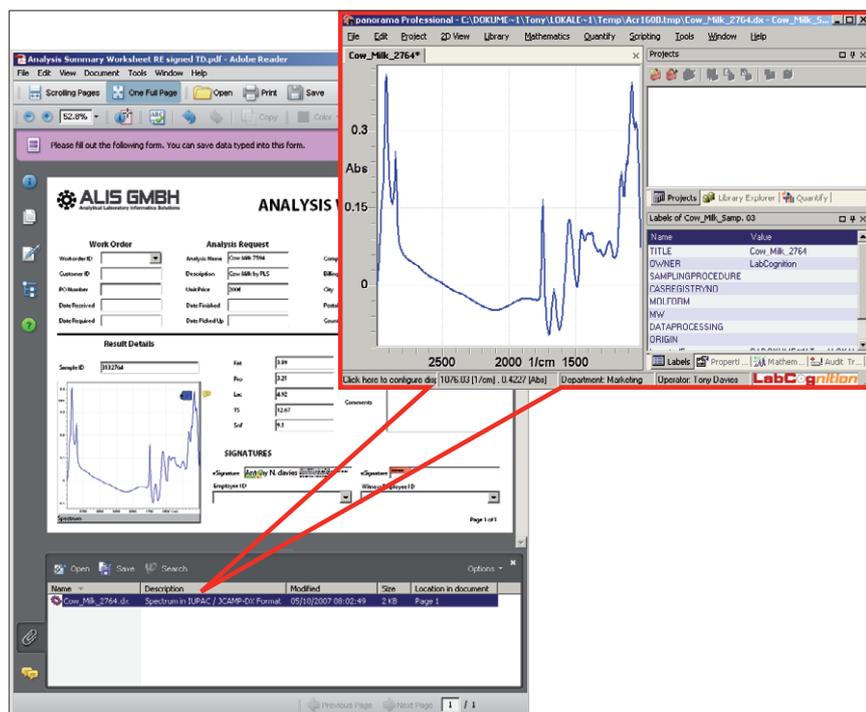


Figure 7. Within Reader the rights-enabled form has now had an image of the spectrum added to the image box and the actually spectrum added as an attachment before the document was digitally signed. Clicking on the attachment launches whichever spectroscopy package is associated with the data format of the attachment—in this case IUPAC/JCAMP-DX is associated with the LabCognition Panorama software.

in Figure 6, an area has been reserved for an image to show the exact view the spectroscopist had on the data as the decision was made to pass the sample.

The actual measurement data in any format can also be saved with the rights-enabled PDF document as an attachment. In our example in Figure 7 the file has been attached as a comment to the uploaded image in IUPAC/JCAMP-DX format to ensure the ability to call up the data for re-analysis at any time in the future. By clicking on the attachment the local client spectroscopy software—in this example the LabCognition Panorama software—is automatically started and the attached file loaded.

Conclusion

"The better part of valor is discretion, in the which better part I have saved my life data".⁶

Don't say it isn't possible to improve on Shakespeare!

...And unfortunately this is only the client-sided half of the story. The Adobe LiveCycle suite of products adds substantially more functionality especially around workflow and document security which is of interest for all regulated areas of analytical work including ensuring your labs are all working to the same SOP for example... but that is a story for another day.

References

1. William Shakespeare (1564–1616), *Hamlet*, Act 1 Scene 5, 168.
2. http://www.adobe.com/devnet/pdf/pdf_reference.html
3. PDF/A, see <http://www.aiim.org/standards>
4. AIIM—The Enterprise Content Management Association has offices in the USA and Europe. see <http://www.aiim.org/article-aiim.asp?ID=20964>
5. http://www.iso.org/iso/standards_development/technical_committees/list_of_iso_technical_committees/iso_technical_committee.htm?commid=53650
6. William Shakespeare (1564–1616), *King Henry IV part I*, Act 5, Scene 4, 119.