

Getting The Most Out Of Native PDF Support - The Next Milestone In Qualitative Computing

by Dr. Susanne Friese
QUARC Qualitative Research and Consulting



Susanne Friese runs Qualitative Research & Consulting, Germany (QUARC), which specializes in methods and techniques of qualitative data analysis, especially in the area of computer-aided qualitative data analysis. In the US, she has worked with the developers of *The Ethnograph* at Qualis Research Associates. She has gained experience in a number of qualitative software packages as a trainer in the CAQDAS networking program at the University of Surrey. She has used ATLAS.ti in her own academic work at the University of Sussex, UK, the University of Hohenheim, Germany, the Copenhagen Business School and, at present, the University of Hanover.

On the Web: www.quarc.de, Email: info@quarc.de

For many years, it's been common for researchers to receive, download, upload, create or send PDF documents. But have you ever tried to attach some codes to them? A handful of QDA software packages allow you to work with "stripped" PDF files: more or less what you get when you copy and paste the text of a PDF document into a Word file. But by sacrificing the PDF's formatting, you end up with a document that's both unsightly and a pain to use.

Thank goodness the solution has arrived. The programmers behind ATLAS.ti v6.0 have succeeded in creating full native PDF support – meaning greater functionality, fewer tedious steps, and higher-quality research.

Dot-matrix To [Dot]com: A Brief History Of PDFs And QDA

Just in case you're new to QDA, let me take you on a journey through time, back to its beginnings in the mid-1980s – a time when researchers huddled around a dot-matrix printer to celebrate the first outputting of a simple retrieval. This meant no more stacks of paper on the floor, no more rifling through boxes of cut-up data segments, no more searching for the data segment that you were certain was marked in red and somewhere at the top of a page. PDFs weren't around yet: back then, we worked with simple, unformatted text files called ASCII. The first Windows programs arrived in the mid-1990s, but they still only supported text files.

The advent of PDFs was both a blessing and a challenge for researchers. When it comes to coding, PDFs work differently from text files, graphical documents, and audio/video clips. In ATLAS.ti, quotations and codes are like a layer on top of the documents; the HU stores a reference about the start and end point in the data where a code is attached. When coding a *rich text file*, this reference is defined by the paragraph number and the character numbers within the paragraph. In *graphical documents*, the coded segment is rectangular: the two coordinates that mark the beginning and end define the rectangle through its upper left and lower right hand corner. In *audio and video documents*, the boundaries of a quotation are defined by time parameters (minutes:seconds:milliseconds) or by frames.

PDFs, on the other hand, consist only of objects and object types. Each addressable object gets a number, and at the very end of the PDF file is a cross-reference table that maps object numbers to locations within the file. PDF renderers start by reading the cross-reference table from the end of the file, and then retrieve the objects randomly by number. For a long time it was impossible to make codes, quotations, memos, or hyperlinks "stick" to data segments in a PDF document.

Fortunately, it didn't take another 20 years to reach the next milestone in qualitative computing: full PDF support. With ATLAS.ti v6.0, the PDF is just one more item on the long list of supported formats – and this one opens up a whole new world of possibilities. (For an overview of ATLAS.ti v6.0 PDF features, [click here](#).)

Native PDF Support: How It Helps Your QDA

And now, let's take a look inside ATLAS.ti's full PDF support – how it works, what you can code and retrieve, and how it can make your life easier.

Keep Your Layout



With ATLAS.ti v6.0, you can keep your original PDFs – layout, graphics, and even tables – so your primary data always remains uncorrupted and complete. This means that you no longer have to be afraid of lost formatting, especially if you work with tables. Your documents can contain virtually any arrangement of columns, text and graphic objects, including embedded graphics. Basically, anything that you can save as PDF file can be coded in v6.0 – the ATLAS.ti website, for example...

The above image shows the ATLAS.ti webpage based on a combination of text and graphic elements saved as PDF file.

The End Of OCR

If you're already using PDF data as source material (e.g. from archives), full PDF support means that you don't have to scan the data using OCR software. As we all know, correcting and editing OCR results takes up considerable time. And less time spent on data preparation means more time for analysis.

Keep Your Line Numbers



When working with such methodological approaches as biographic narrative interviews or conversational analysis, researchers often prefer to work with line-numbered documents. With pre-v6.0 software, this meant working with text-only files saved with a hard return at the end of every line. But now you can simply insert line numbers in your text document, save it as a PDF, and assign it to your HU. When coding, select the text, including the associated line numbers, by drawing a rectangle around them it. When you retrieve these coded segments, the line numbers are displayed below the coded lines.

Keep Your Revisions And Comments



The PDF format allows you to save "track changes" features like revisions and comments in your text documents. What you see on your screen will be preserved in the PDF document. This comes in handy when you need transcript revisions and comments from your respondents, or when you have commented on a document yourself and want to keep your ideas without retyping them as quotation comments.

Code Like The Wind

Classroom Observation

Focus of observation - How is time managed

Time	Observation
7:45	students hand in homework Announcements- last day for MMPC money, Quiz tomorrow (full hour), extra credit for clineX
7:50	Warmup- Dodge grades homework in 7min for completeness
7:58	Pass back homework Go over answers to warm-up; says this is a idea similar to your quiz
8:15	reads answers to homework
8:19	does 2 questions from last nights homework.
8:31	Gives review sheet and students work in groups or on their own, and ask questions
9:14	students pack up early.

- ☒ admin stuff
- ☒ warm-up
- ☒ admin_activity
- ☒ explanation
- ☒ exercise
- ☒ reaction_students

Text: Coding text in PDF files is business as usual – just mark the selection of your choice with the mouse. Selecting across page boundaries is a little tricky, since only one page at a time can be displayed. But you can do it by clicking at the beginning of the quotation-to-be, scrolling to the next page, and clicking the end of the quotation while holding down the "shift" key. Just keep in mind that, when you create output with this quotation, it might include intervening text such as page numbers and footnotes.

Images

Up until now, images embedded within an RTF file could only be coded as one object. Within PDF files, images can now be coded in the same way as stand-alone graphical documents: select a rectangular area of your choice and set it as a quotation or code it at the same time. However, as in the previous version, you still need to view retrieved segments on screen. The output in RTF format contains the reference to the graphical quotation, not the partial image itself.

Tables

When coding text in tables, you can select as much data as you want – the entire text within a cell, or only part of the text within a cell. You can even code text across cells by highlighting a rectangular area.

Output With Ease: Stripping Text

Finally, when you code text in PDF documents and create output, ATLAS.ti strips the text for you – a nice little bonus! That means no need to retype text if you want to cite your sources. You simply code the text that you want to cite and create an output.

Tips For Creating Proper PDF Documents

ATLAS.ti v6.0 currently supports files following the most commonly used PDF 1.4 specifications. When creating PDF documents, make sure you don't use password protection or selective protection of features, e.g. prohibiting text selection. Play it safe by specifying Acrobat 5.0 compatibility (which is PDF 1.4) in the PDF Writer settings.

The image anomaly: Sometimes PDF documents only look like text documents. But if you're scanning documents without OCR (text recognition), all you'll get is a multi-page image file – nothing but one image per page. As a consequence, there is nothing to search and no text to select. If you are the creator or in control of the authoring process, repeat the creation by including text recognition.

The selection anomaly: You will often encounter PDF documents in which blocks of text that appear to be orderly onscreen cannot be selected as expected. This is due to PDF being primarily a *layout format* and *not a text document format*. The author creating this document might have inserted a text box or added a paragraph at the top of the page right before completing the document. Internally, the text is placed at the end – it only appears at the beginning. The selection mechanism cannot deduce the correct logical order of the text, resulting in weird jumps when an author extends a selection.

Bottom Line

If it's on your screen, it can be converted into a PDF. And if it can be a PDF, it can be fully analyzed in ATLAS.ti v6.0. I'm certain that you, the users, will discover countless other ways to take advantage of this new feature.

So now that you've learned the basics, get out there and PDF the world!