

Using SAS to Speed up Annotating Case Report Forms in PDF Format

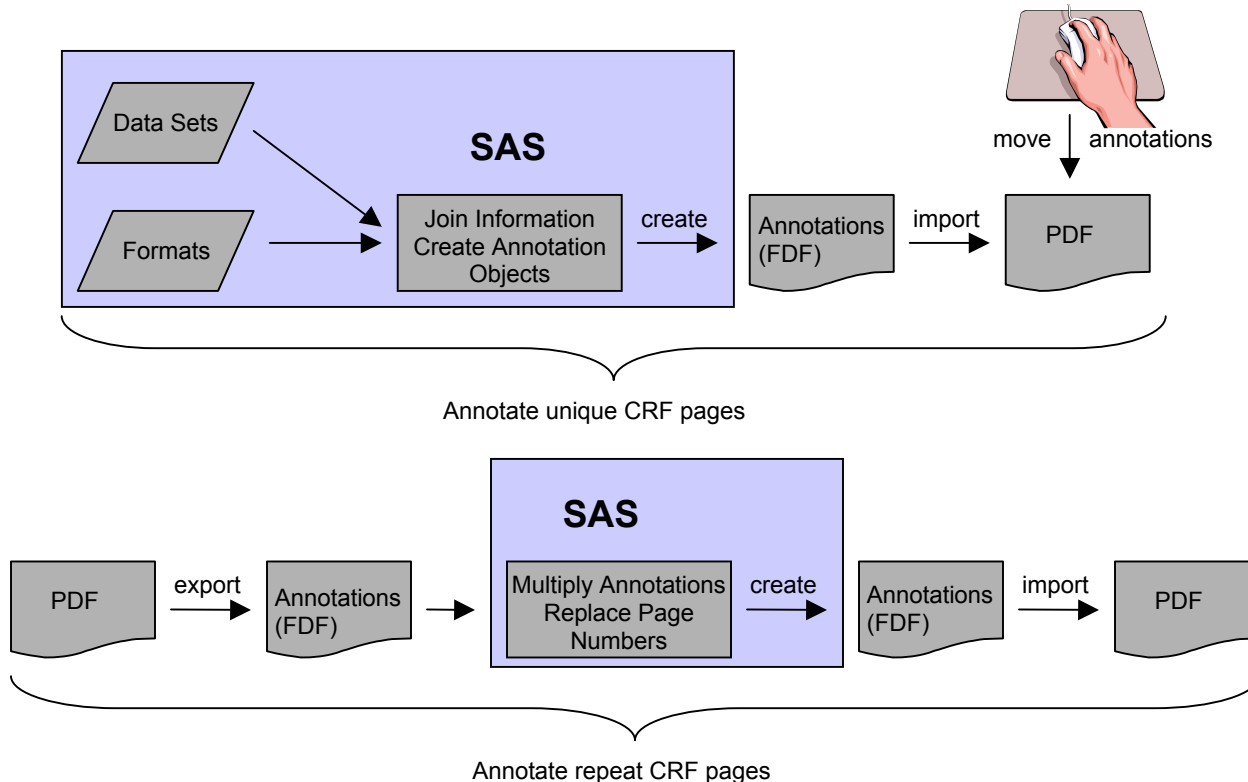
Dirk Spruck, Covidence GmbH, Marburg, Germany
Monika Kawohl, Covidence GmbH, Marburg, Germany

ABSTRACT

One of the first things programmers do before they start working on clinical data is to create an annotated case report form (CRF). Variable names are written on the CRF to provide a link between the fields on the form and the variables in the data sets. Annotating the CRF electronically increases the readability and reusability of the document. For electronic submissions the Food and Drug Administration (FDA) requires a completely annotated CRF in PDF format.

This paper describes the technology used to place annotations in a PDF file and the process of copying annotations between pages. SAS is used to write the variable names from specified data sets and the associated formats as notes in Forms Data Format (FDF). After the CRF is converted to a PDF file the notes are imported into the PDF document and manually placed in the appropriate position on the CRF page. For repeat pages the notes are exported. Using SAS the exported notes are copied and the page numbers are replaced. The notes are then re-imported into the PDF document.

The figure below illustrates the process:



INTRODUCTION

An annotated CRF is a CRF in which the variable names are written next to the spaces provided for the investigator. It serves as a link between the database/data sets and the questions on the CRF. The annotations help both the programmer and the reviewer to understand the data sets and are a vital tool for programming. Annotated CRFs are usually created before programming starts.

There are different ways to annotate CRFs. In the past the variables were hand-written on a printed CRF. However, in the case of electronic submission of data sets, electronically annotated CRF become mandatory. Sharing an electronic copy of an annotated CRF is more convenient than sharing a copy of a handwritten annotated CRF. The FDA requires all documents to be submitted as PDF files.

Sample of electronically annotated CRF (variables BIRTHDI and SEX with format SEXA) in PDF:

DATE INFORMED PATIENT CONSENT WAS SIGNED:		SEXA 1=Male 2=Female	
DATE OF BIRTH:	<input type="text"/> <input type="text"/> <input type="text"/>	BIRTHDI	SEX: <input type="checkbox"/> 1 Male <input type="checkbox"/> 2 Female

The availability of an annotated CRF in electronic format has many advantages and a commonly used file format is the Portable Document Format (PDF). Unfortunately the handling of annotations in PDF documents with Adobe Acrobat is not very convenient. The “copy and paste” function has only limited use and annotations cannot easily be copied between pages, making the annotation of similar pages in Acrobat a tedious manual task.

Fortunately, annotations in PDF can be exported and imported as FDF files. FDF is based on PDF with essentially the same structure, syntax and object types. The structure of exported annotations makes it easy to create and manipulate them in SAS.

Use of data set information, e.g. variable names and formats, minimizes spelling errors and reduces the risk of inadvertently omitting variables. Having formats on the annotated CRF is not mandatory but is certainly an advantageous add-on, which is possible with minimum effort. The electronic copying of annotations ensures that similar pages look alike.

The terminology in Adobe Acrobat 5 is different from that in Adobe Acrobat 4: “annotations” are now called “comments”. Since Adobe’s PDF Reference continues to use the term “annotations” this paper will use the terms “annotations” and “comments” as synonyms.

IMPORTING VARIABLE NAMES AND FORMATS TO PDF

After the CRF is converted to a PDF file, e.g. by scanning the paper CRF, we want to use variable names and formats available in the data sets to create annotations. There is no way to automatically place the annotations in the right position on the CRF but at least all annotations can be placed on the right page. Moving the annotations is a manual process but moving is still easier than creating all annotations from scratch.

The variable names and associated formats can be retrieved using PROC CONTENTS or the SQL dictionary views. Common variables that are in every data set but not printed on the CRF, e.g. database key variables, should be deleted from the data sets. In this way manual deletion after importing them into the PDF file can be avoided.

Sample data set definition:

Data Set	Variable	Format
DEMO	INVSITE	
DEMO	PT	
DEMO	BIRTHDI	
DEMO	SEX	SEXA.
...

Using PROC FORMAT the codes and decodes for the variables can be extracted from the format libraries.

Sample format definition:

Format	Code	Decode
SEXA	1	Male
SEXA	2	Female
...

In this first step of annotating the CRF, only the unique pages are filled with the variable names and formats. Using SAS we can next join the data set and the format information and add the numbers of the pages on which the variables should be placed.

Date Set	Page
DEMO	9
VITAL	14
AE	23
...	...

Combining this information results in the creation of a FDF file. This FDF file can be imported into PDF and contains definitions for annotations representing the variable names and the format definitions.

The figure below shows the menus for importing annotations/comments in Adobe Acrobat:



New annotations after importing into Adobe Acrobat:

INVSITE
PT
INCL
VISDI
INFDI
BIRTHDI
RACE
SEX

INCLUSION / EXCLUSION & PATIENT DEMOGRAPH

Project No.	Protocol No.	Invest. No.	Patient Initials	Patient Alloc No.
	201	US01916	_ _ _ _	_ _ _ _
Date of Visit:				
_ _ _ _				
Patient meets inclusion/exclusion criteria. <input type="checkbox"/> No <input type="checkbox"/> Yes				
Comments: _____				
DATE INFORMED PATIENT CONSENT WAS SIGNED: _ _ _ _				

FDF FILE STRUCTURE

The FDF file consists of a header, a body and a trailer. The header statement is always "%FDF-1.2". The first element of the body is typically the root element of the document. The body consists of objects representing the catalog of the file referencing objects and of object definitions. In our example, the root element is the dictionary catalog listing all annotations that are defined later in the body of the FDF file. The trailer references the root element/object and closes the FDF file. For a complete description of FDF and PDF files, please refer to Adobe's PDF Reference [1].

Here is an example of SAS code creating the header information and the catalog dictionary with three entries for a FDF file:

```
*-----;
*) Initialize FDF file;
*-----;
put '%FDF-1.2';           * FDF header;
put "1 0 obj";           * Root object;
put "<</FDF";
put "<</Annots [ ";
put "2 0 R";               * Reference first annotation - object 2;
put "3 0 R";               * Reference second annotation - object 3;
put "4 0 R";               * Reference third annotation - object 4;
put "]" >>";
put ">>";
put "endobj";             * End of root object;
```

Sample SAS code for creating the trailer information for a FDF file:

```
*-----;
*) Trailer - close FDF file;
*-----;
put "trailer";           * Begin document trailer;
put "<<";
put "/Root 1 0 R";       * Reference to root element;
put ">>";
put '%EOF';              * End of file;
```

DEFINING ANNOTATIONS

For annotating CRFs we use two kinds of annotations: free-text and note annotations. Free-text annotations are used to annotate the CRF with variable names. They are simple text boxes with attributes like color, size and text style.

Note annotations combined with pop-up text annotations are used for annotating the format definitions. They look like notes stuck on the document and can either be open or closed. Besides typical attributes, like color, size and text style, annotations also have titles and their creation date can be displayed.

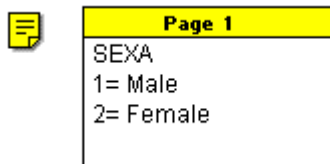
Both types of annotations can be created manually using the Notes Tool in Adobe Acrobat.

Sample SAS code for creating the annotation for variable SEX on page 9:

```
*-----;
*) Annotation of variable name;
*-----;
put "2 0 obj";           * Begin object 2;
put "<<";
put "/Type /Annot";      * Define Annotation;
put "/Subtype /FreeText";
put "/Rect [ 485 412 542 425 ]"; * Size and position;
put "/C [ ]";           * Color background;
put "/DA ([1 0 0.5] r /Helv 10 Tf)"; * Text style;
put "/BS << /W 0 >>"; * Border style;
put "/F 4";             * Flag with display options;
put "/Contents (SEX)"; * Contents;
put "/Page 8";         * Page to put annotation on;
put ">>";
put "endobj";           * End of object;
```

Please note that internally PDF documents start numbering pages with 0. Therefore when annotations are created the page number in the object definition needs to be the displayed page number minus one.

For annotating the format descriptions it is necessary to create two objects. One is the small icon displayed when the note is closed. The other is the open note with the actual text displayed. The two objects have to reference each other:



Sample SAS code for creating the pop-up icon associated with the note displaying the format definition on page 9:

```

*-----;
*) Annotation of pop-up icon;
*-----;
put "3 0 obj";
put "<<";
put "/Type /Annot";
put "/Subtype /Popup";
put "/Rect [ 400 400 495 470 ] ";
put "/Open true";
put "/Parent 4 0 R";
put "/Page 8";
put ">>";
put "endobj";
* Begin object 3;
* Define annotation;
* Size and position;
* Parent of note object 4;
* Page to place annotation on;

```

Sample SAS code for creating the annotation for the format SEXA on page 9:

```

*-----;
*) Annotation of format decode;
*-----;
put "4 0 obj";
put "<<";
put "/Type /Annot";
put "/Subtype /Text";
put "/Rect [ 510 420 530 440 ] ";
put "/Name /Note";
put "/C [ 1 1 0 ] ";
put "/Popup 3 0 R";
put "/T (PAGE 9 ) ";
put "/Contents (SEXA\r1=Male\r2=Female ) ";
put "/Page 8";
put ">>";
put "endobj";
* Begin object 4;
* Define annotation;
* Size and position;
* Display as a note;
* Color;
* Reference to pop-up icon object 3;
* Title;
* Contents - format definition;
* including line breaks;
* Page to place annotation on;

```

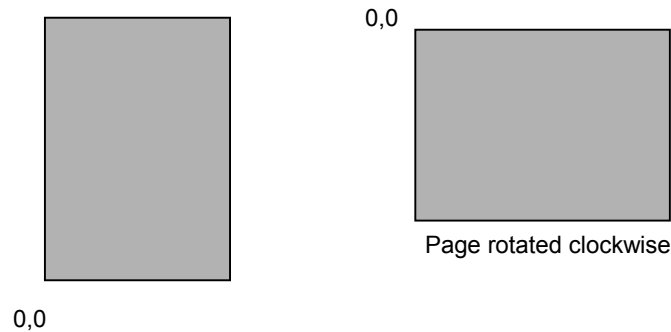
POSITIONING OF ANNOTATIONS

When new annotations are created they must be placed somewhere on the page. We decided to place new annotations on the left side of the page starting at the top and going down. If there are a lot of variables in the data set a second column is started. The pop-up notes containing the format descriptions are placed next to the variable the format belongs to. Once the annotations are on the page they can be manually moved to the correct position on the CRF page using the mouse. Repositioning the annotations manually takes time but it is far less time consuming than creating new annotations.

The position and the size of an annotation is determined by the

```
/Rect [ x0 y0 x1 y1 ]
```

entry in the object definition. The coordinates x0 and y0 define the lower left corner of the object and x1, y1 the upper right corner of the object. The origin of the coordinate system is in the lower left corner of the page. If the page is rotated the origin rotates with the page. For example after the page is rotated 90° clockwise the origin is in the upper left corner.



If a free-text annotation, e.g. variable name, is placed on a rotated page a `/Rotate nnn` entry with the degrees of clockwise rotation has to be added to the object definition. In the above example where the page is rotated to the right the entry would be `/Rotate 90`. The pop-up notes used for format definitions cannot be displayed rotated and the `/Rotate` entry does not need to be specified. However the coordinates need to be adjusted for both object types.

COPYING ANNOTATIONS BETWEEN PAGES

Often, a CRF contains many similar pages, e.g. for different visits. Without additional plug-ins it is not possible in Adobe Acrobat to select and copy multiple annotations from one page to another. Annotations can only be copied one by one and they have to be repositioned, which is very tedious.

Under File/Export/Comments all annotations can be exported to an FDF file. This file is read and converted to a SAS data set. All annotations from the specified page are extracted. These annotations are then written to a new FDF file whereby the old page numbers are replaced by the new ones. The positions and attributes of the annotations are preserved. The numbering of the objects has to reflect the structure of the new FDF file and cannot be retained from the old file. After the FDF file is created it can be imported into the PDF document.

Exporting annotations does not delete them from the PDF document. When annotations are imported they are added to the already existing annotations.

HARD- AND SOFTWARE REQUIREMENTS

The following programs were used under Windows 2000:

SAS 8.2 (server)
Adobe Acrobat 5.0

The project was started using Adobe Acrobat 4. The developed techniques were then extensively used under version 5 without any problems. We have limited experiences with the newest version, but since the basic functionality of FDF and PDF files hasn't changed no problems are expected.

CONCLUSION

Annotating CRFs electronically has many advantages. Besides better readability and higher quality, the speed of annotating can be increased. Once the system is set up, electronic annotation of unique pages might be faster than making handwritten comments depending on the skill of the person concerned. However, if all pages of a CRF need to be annotated, an electronic system that makes it easy to copy annotations between pages will save a tremendous amount of time, effort and cost. The techniques described in this paper can easily be implemented in SAS and Adobe Acrobat. They are a good alternative to Adobe Acrobat plug-ins or other third party tools.

REFERENCES

- [1] Adobe, PDF Reference, Fourth Edition, Version 1.5
<http://partners.adobe.com/asn/tech/pdf/specifications.jsp> (February, 17, 2004)
- [2] SAS Institute, SAS Online Doc Version Eight
<http://v8doc.sas.com> (February 17, 2004)
- [3] FDA, Guidance for Industry – Providing Regulatory Submissions in Electronic Format – General Considerations
<http://www.fda.gov/cder/guidance/2867fnl.pdf> (February 17, 2004)
- [4] Thomas Merz & Olaf Drümmer, Die PostScript- & PDF-Bibel
<http://www.pdflib.com/de/produkte/mehr/bibel/index.html> (February, 17 2004)

ACKNOWLEDGMENTS

We would like to thank David Izard for encouraging us to write this paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Dirk Spruck
Covidence GmbH
Softwarecenter 3
35037 Marburg, Germany
Work Phone: +49 (6421) 94849-37
Email: Dirk.Spruck@covidence.com

Monika Kawohl
Covidence GmbH
Softwarecenter 3
35037 Marburg, Germany
Work Phone: +49 (6421) 94849-20
Email: Monika.Kawohl@covidence.com
Web: www.covidence.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of the respective companies.