



PP23

# Extracting CRF Annotations to Support Define.xml Generation and Review

REATA

Mike Tan ([mike.tan@reatapharma.com](mailto:mike.tan@reatapharma.com)), Richard Addy ([richard.addy@reatapharma.com](mailto:richard.addy@reatapharma.com)), Deidre Kreifels

Reata Pharmaceuticals

## ABSTRACT

- The annotated Case Report Form (aCRF.pdf) is a crucial component in a data submission package. The variable and value annotations on the aCRF are linked to PDF page references in the define.xml, allowing reviewers to quickly associate submission data with the collection forms. For that to work, aCRF annotations and define.xml origins must be consistent. Aligning CRF page references in the define.xml with the aCRF annotation, and keeping them aligned, is an arduous task if handled manually. In this poster, we will share a programmatic approach that extracts CRF annotations (including pdf page numbers) into a SAS dataset. Previous literatures suggested exporting CRF annotations in FDF format. However, we found that exporting as XFDF format is more user-friendly because of its XML-like nature. We will share a CRF annotation schema that allows for efficient extraction by SAS programs. Furthermore, we will share a program design that uses global macro variables to turn on and off extraction features based on the need of a specific study. Finally, we will show how these metadata extracted from annotated CRF can be incorporated in programming specifications, define.xml generation and used to support define.xml review.

## ANNOTATION SCHEMA

The annotation schema is a set of rule to follow when annotating aCRF so a SAS program can correctly extract the necessary information from each comment.

Optional cleaning:

- Remove text inside parenthesis:

Date of First Dose Administration (dd MON yyyy) **EXSTDTC (date part)**  
Time of First Dose Administration (HH:nn) **EXSTDTC (time part)**

- Remove "If ... then" text:

If Yes then VSPOS = SITTING

Each annotation is categorized as one of these groups:

- Group1: NOT SUBMITTED: check for brackets.

Patient Initials **NOT SUBMITTED**

- Group2: Notes: If annotation starts with "Note:..."

Note to Reviewer:  
Each occurrence of a data collection module is annotated in full the first time it is used. Additional occurrences are cross-referenced.

- Group3: xxDTC: As is, use where applicable.

Visit Date: (dd MON yyyy) **xxDTC**

- Group4: See PDF Page XX **See PDF Page 18**

- Group5: Value Level: check for "where" keyword. If found, variable to left stores result and values to right stores code.

2.e Number of Fractions Received xxx **OVAL where SUPPLYA.QNAM=NUMFRAC**

| Height | VSORRES where VSTESTCD = HEIGHT | Centimeters |
|--------|---------------------------------|-------------|
| Weight | VSORRES where VSTESTCD = WEIGHT | Pounds      |

- Group6: Assignment: if "=" found:

Domain Label: verify against dataset list.

FA = Findings About

Variable Level **AGEU = YEARS**

- Group7: Variable Level: if text length <= 8.

Age **AGE**

- Group8: Not Recognized

## COLOR SCHEMA

The color schema is used to group different annotations to their corresponding domain on the same pages.

- First Domain: light yellow

Hue: 40 Red: 255  
Sat: 240 Green: 255  
Lum: 200 Blue: 170

- Third Domain: light green

Hue: 80 Red: 191  
Sat: 240 Green: 255  
Lum: 210 Blue: 191

- Second Domain: light pink

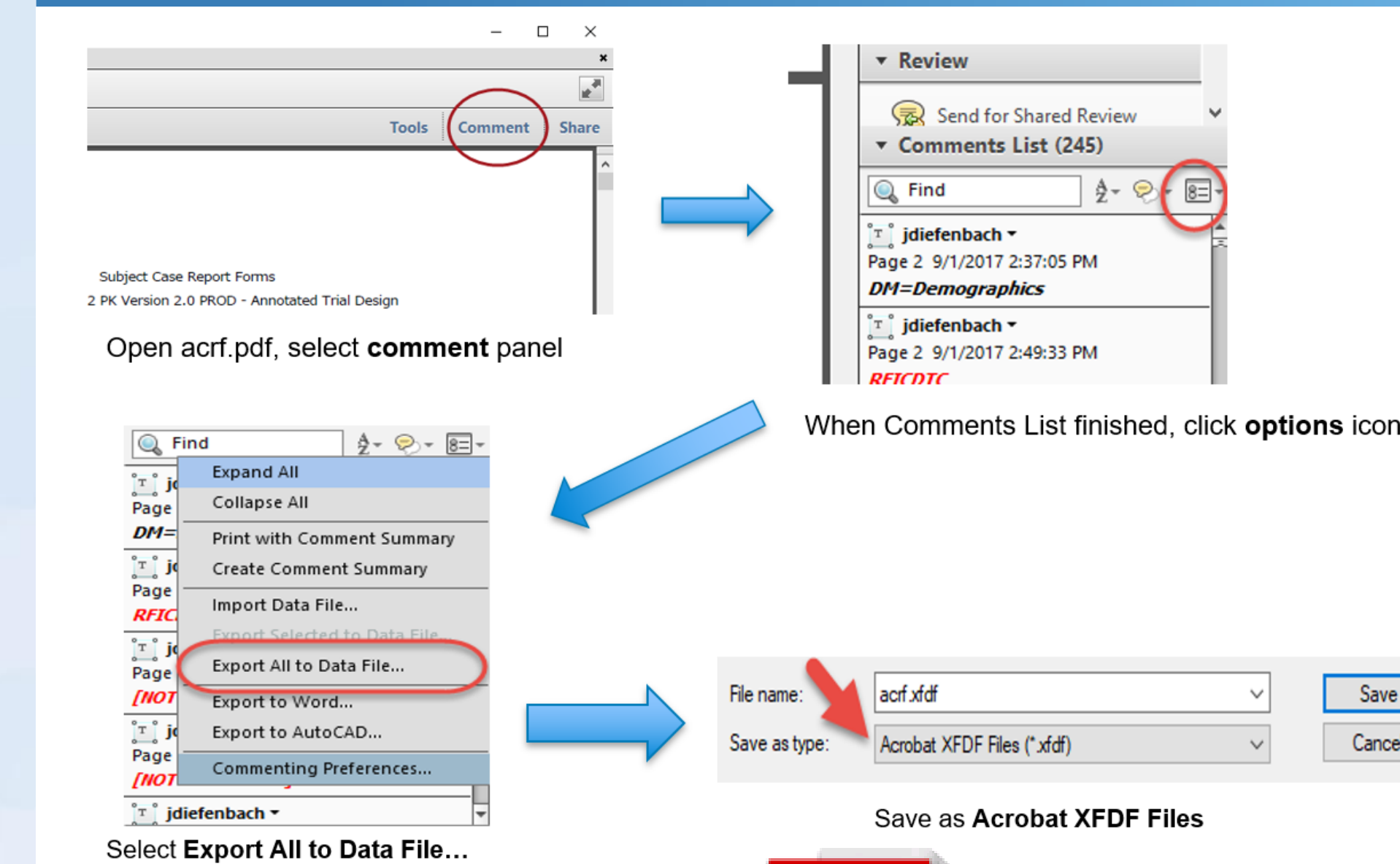
Hue: 10 Red: 255  
Sat: 240 Green: 191  
Lum: 200 Blue: 170

- Fourth Domain: light blue

Hue: 120 Red: 191  
Sat: 240 Green: 255  
Lum: 210 Blue: 255

- See reference site5 for corresponding Hex color values that will be exported to .xfdf file.

## EXPORTING TO XFDF



## XFDF VS. FDF

1 <?xml version="1.0" encoding="UTF-8"?>  
2 <xfdf xmlns="http://ns.adobe.com/xfdf/" xml:space="preserve">  
3 <annots>  
4 <freetext color="#FFFFAA" creationdate="D:20100216084342-05'00'" flags="print" dace="D:20190409085659-05'00'" name="9b1d31fe-ea3d-4c94-9e3d-4940bd267533" page="0" rect="51.657500,740.740000,167.903000,757.775000" subj>  
5 <content-richtext>  
6 <body xmlns="http://www.w3.org/1999/xhtml" xmlns:xfa="http://www.xfa.org/schema/xfa-data/1.0/" xfa:APIVer>  
7 xfa:spec="2.0.2" style="font-size:10.0pt;text-align:left;color:#0000FF;font-weight:normal;font-style:normal">  
8 <p dir="ltr">  
9 DM = Demographics</p>  
10 </body>  
11 </content-richtext>  
12 </defaultappearance>  
13 </defaultstyle>  
14 </freetext>  
15 </annots>  
16 </xfdf>

- Screenshots of xfdf format (above) and fdf format (below) rendered in Notepad++. The xfdf format is xml-like, which made it easier to read.

- These files stored all information about each annotation, including text, font size, color, background color, author, x-y coordinates, created date and time.

1 %PDF-1.2  
2 %  
3 1 0 obj  
4 <</FDF<</Annots[2 0 R 3 0 R 4 0 R 5 0 R 6 0 R 7 0 R 8 0 R 9 0 R 10 0 R 11 0 R 12 0 R 13 0 R  
5 endobj  
6 2 0 obj  
7 <</C[1.0 1.0 0.666656]/Contents(DM = Demographics)/CreationDate(D:20100216084342-05'00')/DA  
8 al;font-family:Times;font-stretch:normal"><p dir="ltr">DM = Demographics</p></body>>/Rect[51  
9 endobj

## XML MAPPER

- Sample mapper to extract page number and background color of each annotation. Note: the page number starts at 0 and background colors were converted to Hex color values.

1 <?xml version="1.0" encoding="UTF-8"?>  
2 <SASMAP name="xfdfmap" version="1.2">  
3 <TABLE name="Annots">  
4 <TABLE-DESCRIPTION>CRFAnnotationExtraction</TABLE-DESCRIPTION>  
5 <TABLE-PATH syntax="XPath">/xfdf/annots/freetext/contents-richtext/body/p</TABLE-PATH>  
6  
7 <COLUMN name="rawPage" retain="YES">  
8 <PATH syntax="XPath">/xfdf/annots/freetext/@page</PATH>  
9 <TYPE>numeric</TYPE>  
10 <DATATYPE>integer</DATATYPE>  
11 <LABEL>Raw page number beginning at zero</LABEL>  
12 </COLUMN>  
13  
14 <COLUMN name="boxColor" retain="YES">  
15 <PATH syntax="XPath">/xfdf/annots/freetext/@color</PATH>  
16 <TYPE>character</TYPE>  
17 <DATATYPE>string</DATATYPE>  
18 <LENGTH>64</LENGTH>  
19 <LABEL>Color of annotation for grouping domains</LABEL>  
20 </COLUMN>  
21 </TABLE>  
22 </SASMAP>

## REFERENCES

- Guideline for submission ready aCRF: <https://www.phusewiki.org/docs/Frankfurt%20Connect%202018/SA/POI/Papers/PP12-pp02-19541.pdf>
- QC of SDTM and aCRF using SAS: <https://www.phusewiki.org/docs/Conference%202016%20AD%20Papers/AD03.pdf>
- Have SAS Annotate your Blank CRF for you! Plus dynamically add color and style to your annotations: <https://www.pharmasug.org/proceedings/2015/AD/PharmaSUG-2015-AD05.pdf>
- Using SAS XML Mapper to Generate and Update an XMLMap: <https://documentation.sas.com/?docsetId=engxml&docsetTarget=n0mxy7afwqron1pioilw0vzqx1.htm&docsetVersion=9.4&locale=en>
- RGB to Hex color conversion: <https://www.rapidtables.com/convert/color/rgb-to-hex.html>

## ANNOTATED CASE REPORT FORM

DA= Drug Accountability EX = Exposure DM = Demographics

Indicate the dosage dispensed at the immediately prior visit. **DAORRES where DATESTCD = DADSPRE**

Will study drug be dispensed to the patient? **If No then set DASTAT = NOT DONE where DATESTCD = DAALL**

Date dispensed (dd MON yyyy): **DADTC**

Enter the dosage being dispensed to the patient at this visit. **DAORRES where DATESTCD = DADSPCLUR EXDOSE EXDOSU = mg**

The IWRS will display the Kit #(s) to be dispensed in this field. (Derived) **DAORRES where DATESTCD = DADSPHWR**

Please retrieve the Kit from storage and manually key in the Kit # of the box being handed to the patient. **DAORRES where DATESTCD = DADSPKIT EXREFID, multiple kits per dosing interval will be concatenated**

Date of First Dose Administration (dd MON yyyy) **EXSTDTC (date part) RFXSTDTC (date part)**

Time of First Dose Administration (HH:nn) **EXSTDTC (time part) RFXSTDTC (time part)**

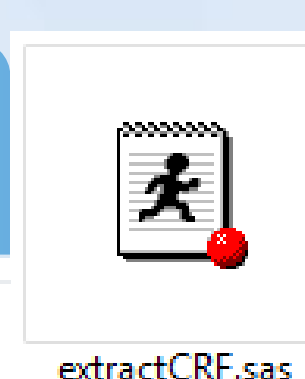
## IMPORTING XFDF INTO SAS

```
1 * mapping files for extracting XML file formats ;  
2 filename xdfmap "C:\Users\mike.tan\Desktop\xfdf.map";  
3  
4 * ACRF annotation extracted XFDF file ;  
5 filename crf_xfdf "C:\Users\mike.tan\Desktop\acrf.xfdf";  
6  
7 * create libname and load XML files into work directory ;  
8 libname crf_xfdf xml xmlmap=xdfmap access=READONLY; * read CRF annotation ;  
9  
10 proc copy in=crf_xfdf out=work;  
11 run;  
12
```

- The remaining extraction methods can be inferred using the annotation schema and the color schema.



CRFMeta.sas7bdat



extractCRF.sas

