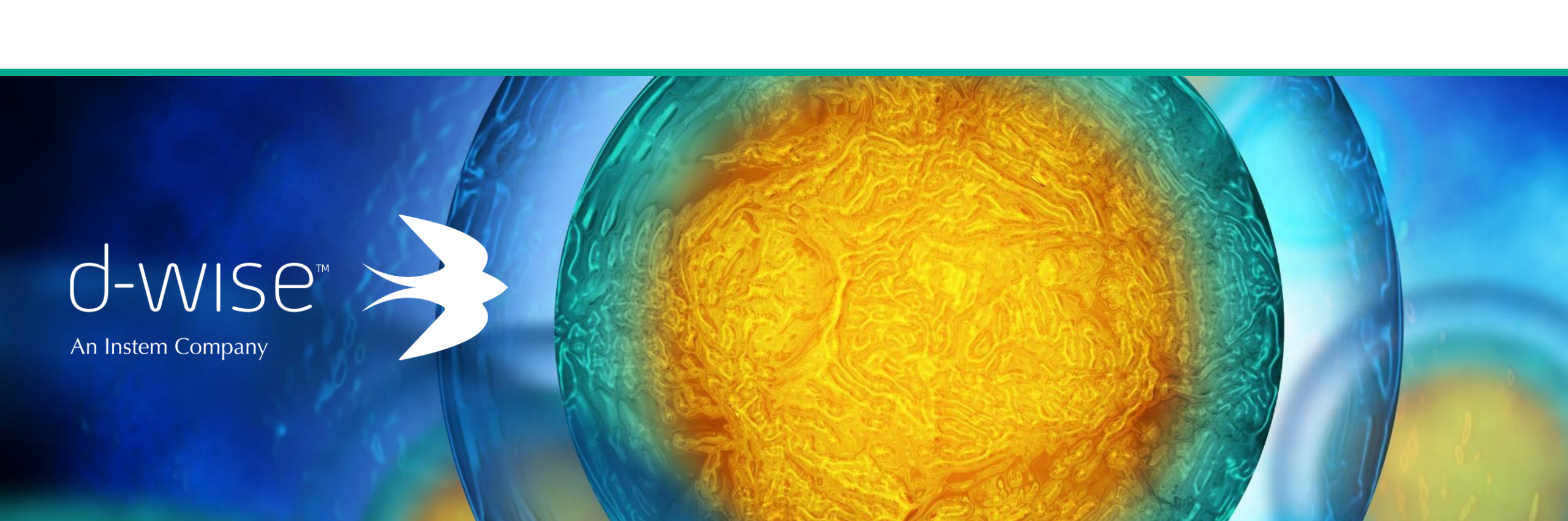




d-wise™

An Instem Company



Processing RTF Documents with SAS

Cara Murphy

Associate consultant

cara.murphy@d-wise.com



Agenda

- **Introduction** – Who am I and why am I here
- **Background** – What was the problem
- **Target** – How could we solve our problem
- **Solution** – How did we do it
 - How does it work
 - What did I learn
- **Summary** – What does it all mean

Introduction

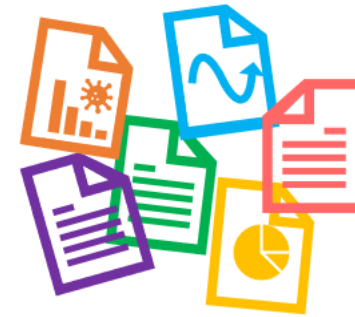
- Who am I?
 - Graduated in June 2020 with a degree in Computing Science
 - Joined d-wise  in September 2020 as a Graduate Consultant
 - First SAS project in October 2020
- Why am I here?
 - My new experience with SAS
 - How I applied it to this project





Background

- The Client produced many **Tables**, **Figures** and **Listings**
- Results all separate documents
- Reports had to be joined together for submission
- This was done manually
- Time consuming
- Not standardised



Target

- The client wanted to automate the process
- To input a list of RTF files containing Tables, Figures and Listings
- And concatenate them
- Saving time and making submissions more simple



Target

- We discussed our options and decided that a **SAS macro** would be the best solution
- It would read in the inputs from a driver file
- Parse the RTF
- Update the format of the document
- Output a single document for submission



Solution - Inputs

task_name	Table_Number	Output_Folder	Output_Filename	Order	Tlf_Section
ITL00016	13-006	C:\Users\cmurp	itl00016.rtf	1	fir
ITL00016	13-006	C:\Users\cmurp	itp00003.rtf	2	fir
ITL00016	13-006	C:\Users\cmurp	ill00002.rtf	3	fir
ITL00016	13-006	C:\Users\cmurp	ill00005.rtf	1	bir
ITL00016	13-006	C:\Users\cmurp	ill00006.rtf	2	bir
ITL00016	13-006	C:\Users\cmurp	itl00010.rtf	1	tir
ITL00016	13-006	C:\Users\cmurp	itl00011.rtf	2	tir

- 1 driver file containing information about the input files
 - Name
 - Table number to be prefixed to title
 - Location
 - Order
 - TLF section used to group files
- Multiple input RTF documents containing Tables, Listings and Figures

Listing 16-013
Listing of Device Deficiencies Prior To Treatment Exposure (All Enrolled Subjects)

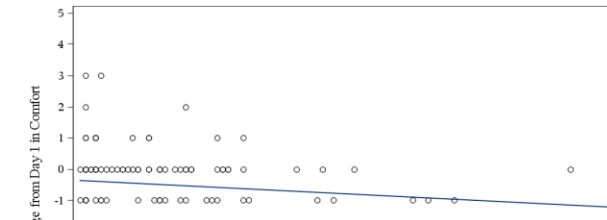
No data satisfied these conditions

SEQ1 = MDACL/DACL; SEQ2 = DACL/MDACL
MDACL = Modified Delefficon A Contact Lens; DACL = Delefficon A Contact Lens

Listing 16-003
Listing of Subjects Discontinued from Study (All Enrolled Subjects)

Inv	Sub	Age	Sex	Lens Sequence Assigned	Lens Sequence Received	Period Discontinued	Visit Discontinued	Reason
4757	00013	19	Female	SEQ2	SEQ2	Period 1	Unscheduled Visit	Withdrawal by Subject
6654	00003	40	Female	NONE	NONE	Period 1	Screen/Baseline	Excl 03: History of refractive surgery or plan to have refractive surgery during the study or irregular cornea in either eye.
6654	00006	28	Female	NONE	NONE	Period 1	Screen/Baseline	Excl 03: History of refractive surgery or refractive surgery during irregular cornea in either

Figure 11-79
Scatter Plot of Change from Day 1 in Comfort and Total Ex Vivo Cholesterol Uptake (µg) at Day 30 (Full Analysis Set)





Solution – Header, Body and Footer

- Each line of RTF documents read in line by line and appended to a data set
- “Subsect” column added to dataset and represent the different sections of the document:
 - **H** for Header
 - **B** for Body
 - **F** for footer

Solution – Header and Body

```
\red0\green128\blue0;  
\red128\green0\blue128;  
\red128\green0\blue0;  
\red128\green128\blue0;  
\red128\green128\blue128;  
\red192\green192\blue192;  
}{\stylesheet{\widctlpar\adjustright\fs20\cgrid\snext0 Normal;}{*\cs10\additive Default Paragraph Font;}  
}{\info{\title Version 9.4 SAS System Output}{\author SAS Version 9.4}{\operatorator SAS Version 9.4}{\version1}{\creatim\yr2020\mol1\dy22\hr22\min47\sec26}}  
\widowctrl\ftnbj\aeenddoc\formshade\viewkind1\viewscale100\pgbrdrhead\pgbrdrfoot\fet0\paperw15840\paperh12240\margl2160\margr1140\margt1701\margb1457  
\sectd\linex0\endnhere\pgwsxn15840\pghsxn12240\lndscpsxn\pgnrestart\pgnstartsl\headery2061\footery1457\marglsxn2160\marginrsxn1140\margtsxn1701\marginbsxn1457  
{\header\pard\plain\qc\pvmrg\phmrg\posxr\posy0{  
\sb0\sa0\qr\f3\fs0\cf8{{\field{\*\fldinst { PAGE }}}}  
{\par}  
\pard{ }}
```

```
/* Start of the body? */  
if substr(inrtf, 1, 6)='\sectd' /*or index(inrtf, " \sectd ")*/  
then  
  
do;  
  /* first row of body */  
  subject='B';
```

- Find the text pattern that showed where the header ends and the body begins
- Each line was added to the data set with the subsect column containing “H” for header
- Until a specific substring was found: **\sectd**
- Then the subsect column would then be set to “B” for Body

Solution – Footer

```
{\par}
\pard\plain\sbl0\sali0\ql\fl\fs20\cf1{}
{\par}
\pard\plain\sbl0\sali0\ql\fl\fs20\cf1{\sars_tools\concatenate_files\pgm\eff\tadva2.sas itl00007.rtf 22NOV20:22:47}
{\par}
\pard\par }
```

- To locate the footer the data set was reversed
- And the first line was set to “F” for footer as this would be the last line in the document
- The final curly bracket was removed from each footer except from the last
- The remaining curly bracket would signify the end of the document

```
/*reverse the table so we can process bottom rows for footer*/
data work.temp2;
  do i = nobs to 1 by -1;
    set temp1 nobs = nobs POINT=i;
    OUTPUT;
  end;

  STOP;
run;
```

```
/*add markup for footer*/
data work.temp3;
  set work.temp2;

  if _n_=1 then
  do;
    subsect='F';

    /* Remove last character from row - it must be squiggle */
    /* This will be conditionally used, i.e. write OUTRTF for all */
    /* files except the last. When writing the last file, use INRTF. */
    outrtf = substr(inrtf,1,length(inrtf)-1);
  end;
run;
```

Solution – Titles and Formatting

1. Once in the body of the text, the title had to be located using a text pattern: `\pard\`

```
\sectd\linex0\endnhere\pgwsxn15840\pghsxn12240\lndscpsxn\pgnrestart
{\header\pard\plain\qc\pvmrg\phmrg\posxr\posy0{
\sb0\sa0\qr\f3\fs0\cf8{{\field{\*\fldinst { PAGE }}}}}
{\par}
\pard{ }}
{\footnote\pard\plain\qc{ }}
\pard\plain\b\sb10\sa10\qc\f4\fs24\cf1{Table 11-007}
{\par}
\pard\plain\b\sb10\sa10\qc\f4\fs24\cf1{Descriptive Statistics.....}
{\par}
\pard{\par}{\*\bkmkstart IDX}{\*\bkmkend IDX}\trowd\trkeep\trhdr\tr
\clbrdrt\brdrs\brdrw15\brdrcf1\clbrdrb\brdrs\brdrw15\brdrcf1\cltxlr
\clbrdrt\brdrs\brdrw15\brdrcf1\clbrdrb\brdrs\brdrw15\brdrcf1\cltxlr
\clbrdrt\brdrs\brdrw15\brdrcf1\clbrdrb\brdrs\brdrw15\brdrcf1\cltxlr
```

```
when('\pard\')
do;
  /*example rtf line: */
  /* \pard\plain\b\i\sb10\sa10\qc\f1\fs100\cf1{This is a really good title} */
```

2. A prefix was added by using `tranwrd` to replace the original text with the new RTF code

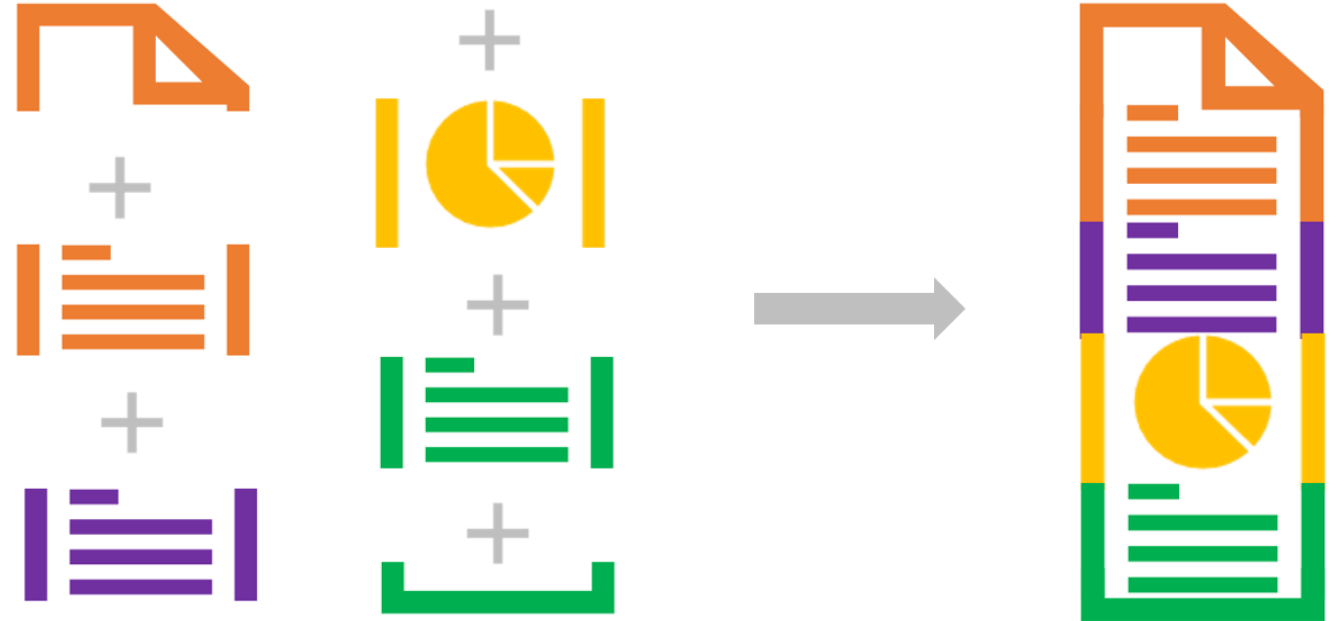
```
/******
/* B. add prefix to title */
/******
outtrtf = tranwrd(tempouttrtf, '{', "\fi-2000{&tprefix." );
```

3. The alignment had to be changed from centre to left and subsequent title lines had to be padded to account for prefix

```
when(index(intrtf, '\qc\'))
do;
  tempouttrtf = tranwrd(intrtf, '\qc\', '\ql\li2000\');
end; /* qc */
```

Solution – Output

- Once all parsing and formatting was complete
 - Old formatting replaced by new RTF code
- The final file was produced
 - with the **header** from the first file
 - The body of all input files
 - and the **footer** of the last file
- Making a complete document



Summary

- RTF is not easy to understand at first inspection
- However by using **SAS macros** all can be made clear
 - Understand the structure
 - Look for keywords
 - Look for patterns
 - A good reference I used for all things RTF:
 - http://www.biblioscape.com/rtf15_spec.htm
- It was a great first experience using SAS
 - Understanding a problem and designing a solution
 - Using SAS
 - Writing code from scratch
 - Using macros



Questions