

Scaffolder – overview of XML output

The XML is a structured representation of the PDF page.

These are the main elements:

Fonts

The font descriptions are given at the top of the XML file:

```
<font>
  <font Id="1.002" font-family="Arial-BoldMT" font-size="14pt" color="rgb(0,0,0)" font-weight="bold" font-style="normal"/>
  <font Id="2.002" font-family="Arial-BoldMT" font-size="9.9pt" color="rgb(0,0,0)" font-weight="bold" font-style="normal"/>
  <font Id="3.002" font-family="ArialMT" font-size="9.9pt" color="rgb(0,0,0)" font-weight="normal" font-style="normal"/>
  <font Id="4.002" font-family="ArialMT" font-size="9.9pt" color="rgb(0,0,0)" font-weight="normal" font-style="normal"/>
</font>
```

Text

Each word occurs three times – in the block, in the line, and as individual words:

The full block of text has <text .../> tags

Position on page (in points from top left corner)

Rotation - 0 is normal horizontal

Number of lines in block

```
<text top="113.65" left="72.02" bottom="136.24" right="557.98" rot="0" numlines="2">A full review of the charity's main areas of risks has been undertaken. This includes all health and safety policies and procedures. The Trustees are happy with the processes that have been established to mitigate these risks.
<line top="113.65" left="72.02" width="485.58" height="11.11">A full review of the charity's main areas of risks has been undertaken. This includes all health and safety policies
  <word rot="0.00" top="113.65" left="72.02" width="6.64" height="11.11" font="4.005">A</word>
  <word rot="0.00" top="113.65" left="80.79" width="12.87" height="11.11" font="4.005">full</word>
  <word rot="0.00" top="113.65" left="95.77" width="28.98" height="11.11" font="4.005">review</word>
  <word rot="0.00" top="113.65" left="126.92" width="8.31" height="11.11" font="4.005">of</word>
```

First line in block <line.../> tags

First word in first line; note: words have fonts

Tables

Tables are given in a similar way to HTML. The following example table is extracted:

During the year the Group incurred the following exceptional items:	
	2016 £'000
Settlement costs to M L Lewis	188
Acquisition costs	155
Diminution in value of unlisted investment	25
	368

The output XML for the table starts:

The diagram illustrates the XML output for the table, with annotations for the first three rows:

- Table row:** Points to the opening `<table>` tag.
- 1st cell in row (empty):** Points to the first `<td>` tag in the first row.
- 2nd cell in row ("2016"):** Points to the text content of the second cell in the first row, which includes a line and a word element for the year "2016".
- 2nd row of table:** Points to the second row of the table, which contains the unit "£'000".
- 3rd row of table:** Points to the third row of the table, which contains the text "Settlement costs to M L Lewis" and the value "188".

```

<table top="162.00" left="56.00" bottom="244.00" right="539.00" rot="0" data="./images/30_02.xls" image="./images/30_02.png">
  <tr>
    <td></td>
    <td>
      <text top="162.31" left="521.53" bottom="168.36" right="538.58" rot="0" numlines="1">2016
        <line top="162.31" left="521.53" width="17.06" height="6.05">2016
          <word rot="0.00" top="162.31" left="521.53" width="17.06" height="6.05" font="3.030">2016</word>
        </line>
      </text>
    </td>
  </tr>
  <tr>
    <td></td>
    <td>
      <text top="172.31" left="519.14" bottom="178.36" right="538.58" rot="0" numlines="1">£'000
        <line top="172.31" left="519.14" width="19.44" height="6.05">£'000
          <word rot="0.00" top="172.31" left="519.14" width="19.44" height="6.05" font="3.030">£'000</word>
        </line>
      </text>
    </td>
  </tr>
  <tr>
    <td>
      <text top="188.48" left="56.69" bottom="196.04" right="188.41" rot="0" numlines="1">Settlement costs to M L Lewis
        <line top="188.48" left="56.69" width="131.72" height="7.56">Settlement costs to M L Lewis
          <word rot="0.00" top="188.48" left="56.69" width="48.25" height="7.56" font="2.030">Settlement</word>
          <word rot="0.00" top="188.48" left="107.48" width="23.05" height="7.56" font="2.030">costs</word>
          <word rot="0.00" top="188.48" left="133.08" width="8.40" height="7.56" font="2.030">to</word>
          <word rot="0.00" top="188.48" left="144.03" width="8.18" height="7.56" font="2.030">M</word>
          <word rot="0.00" top="188.48" left="154.75" width="5.72" height="7.56" font="2.030">L</word>
          <word rot="0.00" top="188.48" left="163.02" width="25.39" height="7.56" font="2.030">Lewis</word>
        </line>
      </text>
    </td>
    <td>
      <text top="188.48" left="522.59" bottom="196.04" right="538.58" rot="0" numlines="1">188
        <line top="188.48" left="522.59" width="15.99" height="7.56">188
          <word rot="0.00" top="188.48" left="522.59" width="15.99" height="7.56" font="4.030">188</word>
        </line>
      </text>
    </td>
  </tr>

```

As mentioned previously, each word or number occurs three times (in block, line and word).