



Accessible Tables using Tagged PDF

TUG-2022, 22–24 JULY 2022 — ROSS MOORE



What makes a table Accessible?

This page gives a good summary of what is needed in an electronic document: <https://www.w3.org/WAI/tutorials/tables/>

- ▶ header and data cells are marked as such;
- ▶ each data cell needs to be associated with the relevant header cells that tell what the data means;
- ▶ the scope of each header cell is given; that is, to which rows or columns, or groups thereof, does it refer;
- ▶ a caption and/or summary to give a brief overview.

Relying on visual cues alone is not sufficient to create an accessible table. With structural markup, headers and data cells can be programmatically determined using software ...



Example: Calendar/schedule

Tags X

Day/Date Time Activity Lead

Day/Date	Time	Activity	Lead
Monday, September 13	9:00 am	Welcome/Introductions	Michele Traver
	9:15 am	Background/AOP Review	Russ Brown
	10:00 am	Gulf of Maine Cod	Charles Perretti
	11:00 am	Review/Discussion	Review Panel
	11:15 pm	Public Comment	Public
	11:30 pm	Lunch	
	12:30 pm	Gulf of Maine Cod cont.	Charles Perretti
	2:30 pm	Public comments	Public
	3:00 pm	Adjourn	
Tuesday, September 14	9:00 am	Welcome	
	9:15 am	George	
	10:30 am	Break	
	1:00 pm	George	
	2:30 pm	Adjourn	
Wednesday, September 15	10:30 am	Review	
	12:00 pm	George	

Path: Attribute Objects
Type: Attribute

Attribute Objects

```
/Attribute Object 1 <<Dictionary>>
  /Headers [Array]
    [0] (TH.tabular.[02,01])
    [1] (TH.tabular.[01,02])
  /O /Table
  /Scope /Row
```

Change Item Delete Item New Item Cancel Ok

Examples: more table heads

Table 1: Comparison of reference points estimated in an earlier assessment and from the current update. The overfishing threshold is the $F_{MSY\text{page}}$ (F_{MSY}). The biomass target, ($SSB_{MSY\text{page}}$) is long-term stochastic projections of fishing at the $F_{MSY\text{page}}$. Median recruitment reflects the media age-1 recruitment from 1982–2017. Intervals shown reflect the 5P and 95P percentiles.

	2019 $M=0.2$	2019 $M\text{-ramp}$	$M=0.2$	$M\text{-ramp}$
F_{MSY}	0.132	0.125	0.132	0.125
SSB_{MSY} (mt)	42,602	63,567	39,912	60,116
($27,916$ - $62,785$)	($46,144$ - $84,098$)	($25,372$ - $59,589$)	($41,914$ - $58,012$)	($44,391$ - $66,103$)
MSY (mt)	7,580	11,20	7,11	10,73
($4,743$ - $13,366$)	($8,149$ - $15,268$)	($4,462$ - $11,023$)	($7,439$ - $14,181$)	($7,439$ - $14,181$)
Median recruits age-1) (000s)	4,377	8,464	4,677	9,299
($1,161$ - $4,434$)	($2,853$ - $5,934$)	($1,064$ - $16,392$)	($2,129$ - $18,185$)	($2,129$ - $18,185$)
Overfishing	Yes	Yes	Yes	No
Overshared	Yes	Yes	Yes	Yes

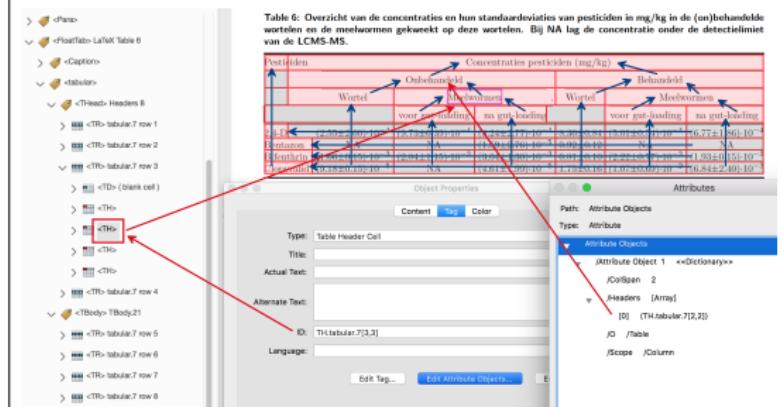
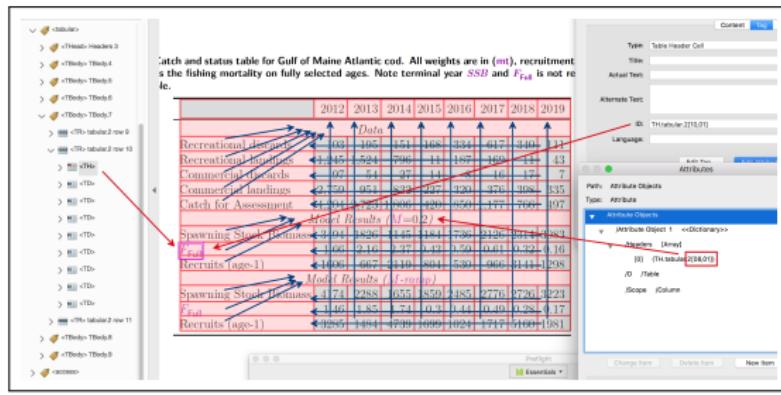
The screenshot shows the LaTeX code for the table and the corresponding table structure. A red box highlights the first column of the table. A red arrow points from the 'Attribute Objects' panel to the first column of the table, indicating the mapping between the attribute objects and the table cells.

Table 1: Comparison of reference points estimated in an earlier assessment and from the current update. The overfishing threshold is the $F_{MSY\text{page}}$ (F_{MSY}). The biomass target, ($SSB_{MSY\text{page}}$) is long-term stochastic projections of fishing at the $F_{MSY\text{page}}$. Median recruitment reflects the media age-1 recruitment from 1982–2017. Intervals shown reflect the 5P and 95P percentiles.

Year	Catch (mt)	SSB (mt)	F_{Full}	Catch (mt)	SSB (mt)	F_{Full}
2020	409	2,635	0.162	499	3,925	0.119
2021	523	3,590	0.177	613	4,750	0.133
2022	531	4,508	0.173	619	5,251	0.175
2023	570	5,488	0.173	1,017	5,707	0.175
2024	1,244	7,279	0.173	1,366	6,802	0.175

The screenshot shows the LaTeX code for the table and the corresponding table structure. A red box highlights the first column of the table. A red arrow points from the 'Attribute Objects' panel to the first column of the table, indicating the mapping between the attribute objects and the table cells.

Examples: hierarchical table heads





Examples in HTML

REAL-WORLD TABLES — FISHERY DATA

Normal Table

The basic style adopted for tabular material is in accordance with [t1](#). Supported rules are `\toprule`, `\midrule`, `\bottomrule` which can inherit instead of being for aesthetic appearance only. This extra semantics takes:

- `\toprule` indicates the start of a (`THead`) block of rows;
- `\midrule` indicates the end of a block, with the following row starting;
- `\bottomrule` indicates the end of the tabular; closing the (`TBody`) block.

Table 1: Comparison of reference points estimated in an earlier assessment and from update. The overfishing threshold is the $F_{MSY, \text{max}} (F_{40\%})$. The biomass target, (S_0) long-term stochastic projections of fishing at the $F_{MSY, \text{proj}}$. Median recruitment refl age-1 recruitment from 1982–2017. Intervals shown reflect the 5th and 95th percent

	2019 $M=0.2$	2019 $M=\text{temp}$	$M=0.2$
F_{MSY}	0.173	0.175	0.173
SSB_{MSY} (mt)	42,692 (27,916–62,785)	63,867 (46,144–84,098)	39,912 (25,472–59,589)
MSY (mt)	7,580 (4,853–11,366)	11,420 (8,149–15,268)	7,171 (4,462–11,023)
Median recruits (ago-1) (000s)	4,377 (1,161–14,434)	8,464 (2,353–15,934)	4,677 (1,064–16,392)
Overfishing	Yes	Yes	Yes
Overfished	Yes	Yes	Yes

Tagged tabulars from Chapter 5

Alignment using preamble commands

§5-2-1	A	B	C
	100	10	1

§5-2-2	A	B	C	100
	100	10	1	A

Checking alignments:

§5-2-3	1 1 1 1 1 1 1 1 1 1 1 1	2 2 2 2 2 2 2 2 2 2 2 2	3 3 3 3

§5-2-4	1 1 1 1 1 1 1 1 1 1 1 1	2 2 2 2 2 2 2 2 2 2 2 2	3 3 3 3

§5-2-5	1 1 1 1 1 1 1 1 1 1 1 1	2 2 2 2 2 2 2 2 2 2 2 2	

§5-2-13	one	two	three
	1	2	3

Using array package features:

§5-2-6	1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0	1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0

§5-2-7	$10^{10!}$ 10^{-999}	a big number a small number

Super-consciousness is a	Possibilities et espérances	Mogelijkheden en hoop

Super-consciousness is a	Possibilities et espérances

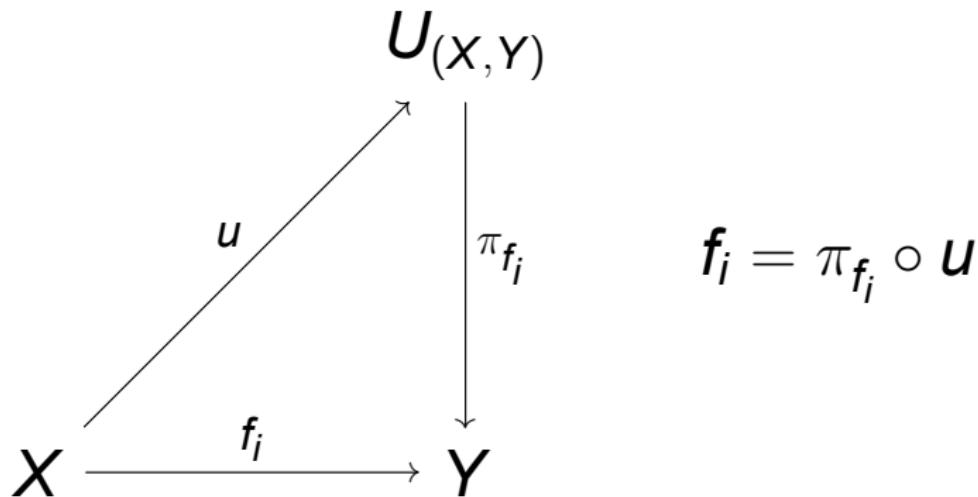
website: <http://science.mq.edu.au/~ross/TaggedPDF/TUG2022/>

more tagged PDFs: <http://science.mq.edu.au/~ross/TaggedPDF/>



Algebra: Universal Morphism

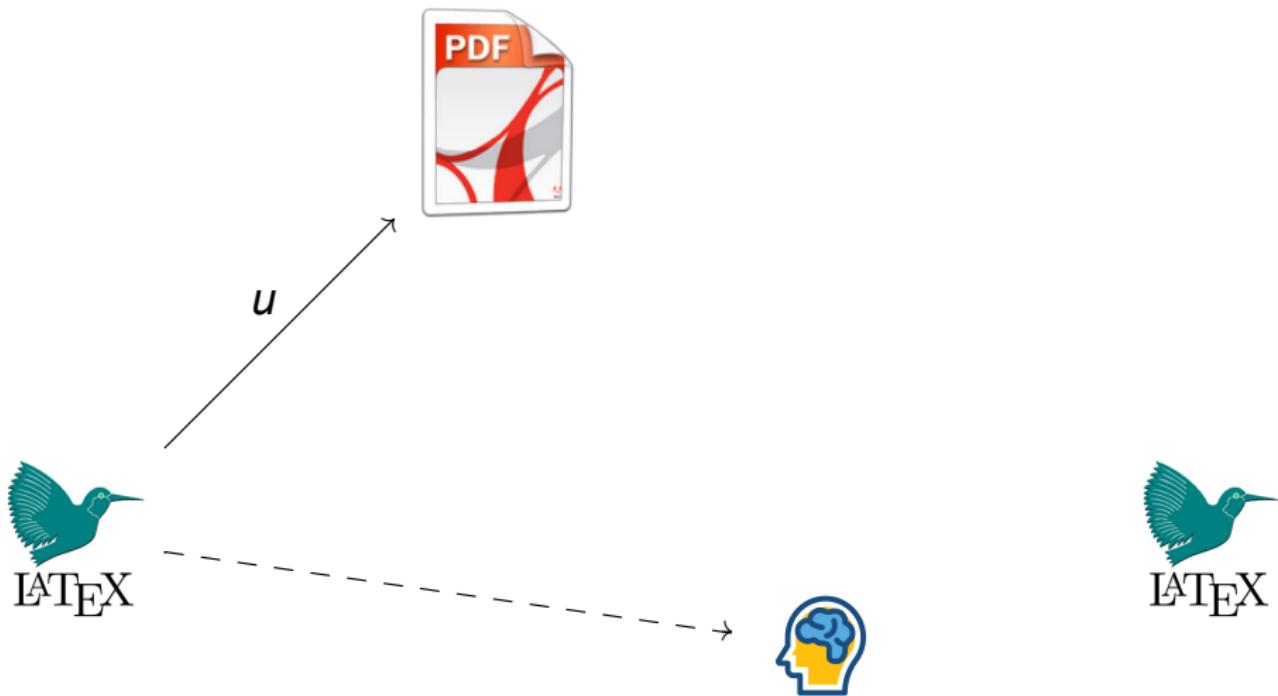
What's in a name? PDF/UA *Universal Accessibility*



Any structure preserving map $f_i : X \rightarrow Y$ factors through the 'Universal Object' $U_{(X,Y)}$, via a projection π_{f_i} .



PDF/UA: Universal Accessibility



Visualising how a Tabular works

Template: generated from the `<cols>`

... # ... & ... # ... & & ... # ... \cr


User data:

... & ... & ... & ... & ... & ... \\
... & ... & ... & ... & ... & ... \\

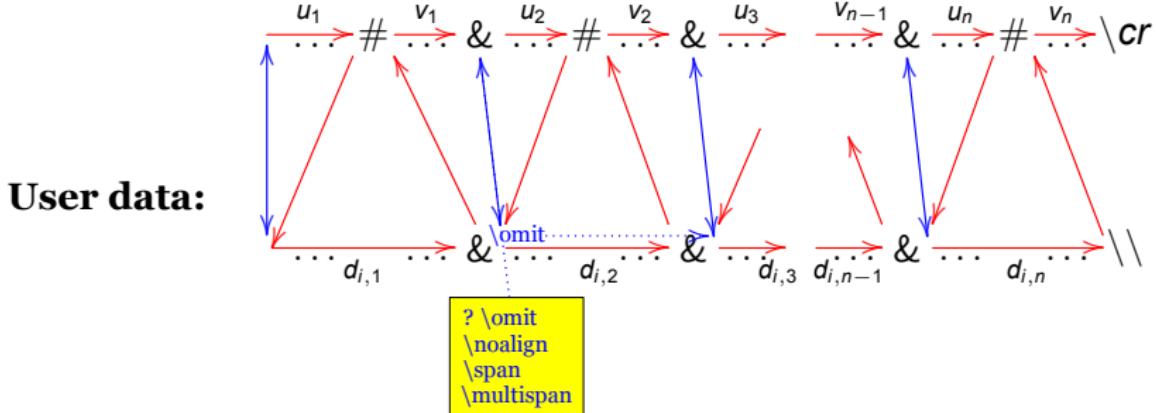
`\end{tabular}`

We conceptualise a table as merging user-supplied data into the preamble/template; both as an author coding the data, and as a reader visualising some or all of the overall layout.

However, this is *not* the way TeX processes these token-streams!

Switching, not merging

Template: generated from the `<cols>`



Summary: $\text{\TeX}/\text{\LaTeX}$ processes tokens for row_i in the order

$u_1 d_{i,1} v_1 u_2 d_{i,2} v_2 u_3 d_{i,3} \dots d_{i,n-1} v_{n-1} u_n d_{i,n} v_n$

with the ability to omit some portions of the template, and to include extra non-aligned material.



A L^AT_EX preamble

p{3.75cm} p{4.5cm} p{7.75cm} produces the following preamble:

```
\unhc@arstrutbox \hskip\col@sep \setbox\z@\hbox\bgroup\bgroup\CT@everycr
{} \vtop@startpbox {3.75cm} \ignorespaces \@sharp \unskip \relax@endpbox
\do@row@strut \egroup\egroup\begin{group}\CT@setup \CT@row@color \CT@cell@color
\CT@do@color \endgroup\@tempdima \ht\z@ \advance\@tempdima
\minrowclearance\LT@vrule \height \@tempdima \width \z@ \unhbox\z@
\hskip\col@sep & \hskip\col@sep \setbox\z@\hbox\bgroup\bgroup\CT@everycr
{} \vtop@startpbox {4.5cm} \ignorespaces \@sharp \unskip \relax@endpbox
\do@row@strut \egroup\egroup\begin{group}\CT@setup \CT@row@color \CT@cell@color
\CT@do@color \endgroup\@tempdima \ht\z@ \advance\@tempdima
\minrowclearance\LT@vrule \height \@tempdima \width \z@ \unhbox\z@
\hskip\col@sep & \hskip\col@sep \setbox\z@\hbox\bgroup\bgroup\CT@everycr {}
\unhc@arstrutbox \hskip\col@sep \setbox\z@\hbox\bgroup\bgroup\CT@everycr
{} \vtop@startpbox{7.75cm} \ignorespaces \@sharp \unskip \relax@endpbox
\do@row@strut \egroup\egroup\begin{group}\CT@setup \CT@row@color \CT@cell@color
\CT@do@color \endgroup\@tempdima \ht\z@ \advance\@tempdima
\minrowclearance\LT@vrule \height \@tempdima \width \z@ \unhbox\z@
\hskip\col@sep \tabskip\z@ \cr
```

Notes on the preamble

- ▶ `\@sharp` becomes `#` during processing of the table's rows; during the creation of the preamble, it is set to `\relax`, to be inert under repeated expansion.
- ▶ Notice that `\ignorespaces` occurs immediately before the data tokens for a cell's contents, as does `\unskip` come immediately after. But as these tokens are commonly used with `input`, and can be redefined within various packages, they are *not useful* as hooks to begin and end content tagging.
- ▶ Instead, the `\tabular` macro is adjusted to perform tagging tasks before calling up its usual L^AT_EX expansion.

The next page illustrates how tagging is started this way, and propagated through the cells and rows, by adjusting the `&` token.



Preamble with some tagging aspects

|>{\large}c|>{\large\bfseries}l |>{\itshape}c| produces the following preamble:

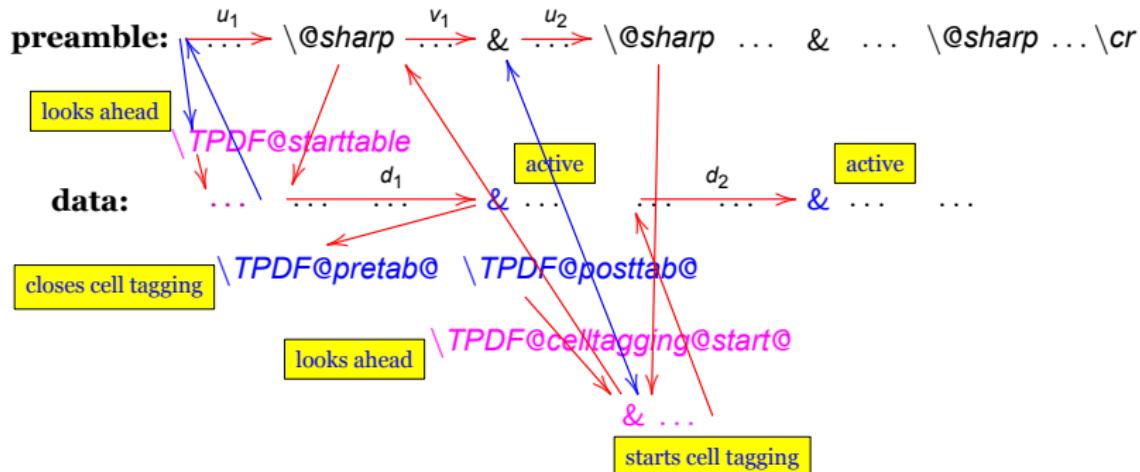
```
\unhcopy\@arstrutbox {\vrule\@width \arrayrulewidth \TPDF@setabcol@prop{0}{9}{>} \hskip\col@sep
\TPDF@setabcol@prop {0}{0}{c} \setbox\z@\hbox\bgroup\bgroup\CT@everycr{} \hskip\z@ plus.5fill \relax
\kern\z@ \d@llarbegin \advance\TPDF@thiscolcount \one \large\ignorespaces \@sharp \unskip\relax
\d@llarend \do@row@strut \hskip\z@ plus.5fill \relax\egroup\egroup\begin{group}\CT@setup \CT@row@color
\CT@cell@color \CT@do@color \endgroup\@tempdima \ht\z@ \advance\@tempdima \minrowclearance\vrule\@height
\@tempdima \@width \z@ \unhbox\z@ \hskip\col@sep {\vrule\@width \arrayrulewidth}
\TPDF@setabcol@prop {0}{9}{>} & \hskip\col@sep \TPDF@setabcol@prop {1}{0}{1} \setbox\z@
\hbox\bgroup\bgroup \CT@everycr{} \d@llarbegin \advance\TPDF@thiscolcount \one \large\bfseries
\ignorespaces \@sharp \unskip\relax\d@llarend \do@row@strut \hfill\egroup\egroup\begin{group}\CT@setup
\CT@row@color \CT@cell@color \CT@do@color \endgroup\@tempdima \ht\z@ \advance\@tempdima
\minrowclearance\vrule\@height \@tempdima \@width \z@ \unhbox\z@ \hskip\col@sep {\vrule\@width
\arrayrulewidth} \TPDF@setabcol@prop {0}{9}{>} & \hskip\col@sep \TPDF@setabcol@prop {0}{0}{c}
\setbox\z@ \hbox\bgroup\bgroup \CT@everycr{} \hskip\z@ plus.5fill \relax\kern\z@ \d@llarbegin
\advance\TPDF@thiscolcount \one \itshape\ignorespaces \@sharp \unskip\relax\d@llarend \do@row@strut
\hskip\z@ plus.5fill\relax\egroup\egroup\begin{group}\CT@setup \CT@row@color \CT@cell@color \CT@do@color
\endgroup\@tempdima \ht\z@ \advance\@tempdima \minrowclearance\vrule\@height \@tempdima \@width \z@
\unhbox\z@ \hskip\col@sep {\vrule\@width \arrayrulewidth}\tabskip\z@ \cr
```

The macro `\TPDF@setabcol@prop` here records the cell's text-alignment, to be passed to HTML as an attribute via CSS rules. When a `<cols>` entry represents vertical alignment, or some semantic information, then this can be handled also.

Tagging a Tabular layout

LaTeX's usual `\tabular` is replaced with `\TPDF@tabular` which, after reading the `(cols)`, does many tagging-related tasks including making `&` into an 'active' character.

A macro `\TPDF@starttable` is inserted immediately before the user data being read from the file-source. This initiates a 'look-ahead' to see what is coming next; e.g., `\toprule` or `\hline`, `\multicolumn` or `\multirow`, a blank cell or content for a normal cell; taking appropriate actions for each.



Further cells are tagged, using the active `&`. When a row is done, its tagging is closed; there is a 'look ahead' to commence tagging the next row's cells, following any `\hrule`, `\midrule`, etc.