

# Revealing the Detailed Lineage of Script Outputs using Hybrid Provenance

Qian Zhang, Yang Cao, Qiwen Wang,  
Duc Vu, Priyaa Thavasimani, Timothy  
McPhillips, Paolo Missier, Peter  
Slaughter, Christopher Jones, Matthew  
B. Jones, Bertram Ludäscher  
School of Information Sciences, University  
of Illinois at Urbana-Champaign, USA;  
Department of Computer Science,  
University of Illinois at  
Urbana-Champaign, USA; Department of  
Electrical and Computer Engineering,  
University of Illinois at Chicago, USA;  
School of Computing Science, Newcastle  
University, UK; National Center for  
Ecological Analysis and Synthesis, UCSB,  
Santa Barbara, USA

## Abstract

We illustrate how combining retrospective and prospective provenance can yield scientifically meaningful hybrid provenance representations of the computational histories of data produced during a script run. We use scripts from multiple disciplines (astrophysics, climate science, biodiversity data curation, and social network analysis), implemented in Python, R, and MATLAB, to highlight the usefulness of diverse forms of retrospective provenance when coupled with prospective provenance. Users provide prospective provenance (i.e., the conceptual workflows latent in scripts) via simple YesWorkflow annotations, embedded as script comments. Runtime observables, hidden in filenames or folder structures, recorded in log-files, or automatically captured using tools such as noWorkflow or the DataONE RunManagers can be linked to prospective provenance via relational views and queries. The YesWorkflow toolkit, example scripts, and demonstration code are available via an open source repository.

*Received 4 July 2013 | Revision received 10 December 2013 | Accepted 1 January 2014*

Correspondence should be addressed to Qian Zhang, School of Information Sciences (iSchool), UIUC, 112 LIS, 501 E. Daniel Street, Champaign IL 61820-6211 Email: [zhangqian06@gmail.com](mailto:zhangqian06@gmail.com)

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution 4.0 International Licence. For details please see <http://creativecommons.org/licenses/by/4.0/>



## Introduction

The L<sup>A</sup>T<sub>E</sub>X class `ijdc-v9` produces camera-ready papers and articles suitable for inclusion in the International Journal of Digital Curation, with applicability from volume 9 onwards. The similar `idcc` class can be used for submissions to the International Digital Curation Conference, beginning with the 2015 conference. This document explains how to use these classes.

## Dependencies

Certain aspects of the template design have been implemented using third-party packages, aside from those that are required parts of the L<sup>A</sup>T<sub>E</sub>X system. Therefore you should ensure that you have these packages installed on your system before attempting to use the class.

- `atbegshi` is used for switching geometry between pages.
- Tables in your document must be formatted according to the design principles promoted and supported by the `booktabs` package.
- `caption` is used to format the figure and table captions.
- `etoolbox` is used behind the scenes for patching commands.
- `footmisc` is used to format the footnotes.
- `titlesec` is used to format the section headings.

In some cases the class prefers to use packages that are not part of the base installation (but are nevertheless commonly available in T<sub>E</sub>X distributions), but will fall back to their base equivalents if necessary.

- `newtx` will be used if available in place of `times`.
- `xcolor` will be used if available in place of `color`.

For referencing, you are encouraged to use either `biblatex` using the styles provided by `biblatex-apa` (preferred), or `apacite`.

## Loading the Classes

### International Journal of Digital Curation

The class is loaded in the usual way with `\documentclass[<options>]{ijdc-v9}`. The following options are available:

**paper** Use this for peer-reviewed papers.

**article** Use this for general (non-peer-reviewed) articles if you like, but you do not have to as the class defaults to this state.

**editorial** Use this for an editorial.

## International Digital Curation Conference

The class is loaded in the usual way with `\documentclass[<options>]{idcc}`. Two types of option are available. The first relates to the type of submission:

**research** Use this for full research papers.

**practice** Use this for extended abstracts of practice papers, and full practice papers. The class defaults to this state, making this option even more optional than normal.

**data** Use this for data paper abstracts and full data papers.

**poster** Use this for poster abstracts.

**demo** Use this for demonstration proposals.

**bof** Use this for Birds of a Feather session abstracts.

**workshop** Use this for workshop proposals.

The second type of option relates to the conference to which the submission will be made:

**15** Use this for the 2015 conference.

**16** Use this for the 2016 conference.

**17** Use this for the 2017 conference.

## Preamble Commands

The following commands should be given in the preamble to fill out the document metadata.

The following command should be used in all submissions.

`\title[<short version>]{<full version>}` The long version of the title is shown on the cover page of the submission, while the short version appears in the (even page) headers.

The following commands should be given in general articles and IDCC submissions. They should *not* be given in peer-reviewed IJDC papers until after the peer review process is complete.

`\author{<name>}` The name of one author. Repeat the command for each additional author.

It is customary in IJDC and IDCC papers to group authors by institution. Within each institution, the authors are ordered by the level of contribution (or alphabetically where this is equal), and the institutional groups are ordered by the level of contribution of the first author in the group (or alphabetically by first author where this is equal). A different convention may be used if appropriate.

`\affil{<name>}` The affiliation (institution, company) of the immediately preceding author(s). This command may be repeated as necessary.

`\correspondence{<name, postal address.> Email: \email{<email address>}}` Name, address and email address of the corresponding author. This information appears in the footer of the cover page.

If an IJDC submission is a reworked conference paper (that has not already been formally published), for reasons of transparency the name of the conference should be given.

`\conference{<name of conference>}` The conference at which the earlier version of the paper was presented, e.g. ‘the 10th International Digital Curation Conference’.

For IDCC papers, authors are invited to record the date on which they made their submission.

`\submitted{<date>}` The date on which the initial submission was made to the conference by the authors.

Some additional commands are used by the editorial team when preparing a submission for publication. Though authors would not normally need to use them, here they are for completeness.

`\received{<date>}` The date on which the initial submission was received by the editorial team (IJDC papers only).

`\revised{<date>}` The date on which the latest revision was received by the editorial team.

`\accepted{<date>}` The date on which the submission was accepted for publication.

`\subno{<number>}` The submission number allocated by the IJDC Open Journal System.

`\volume{<number>}` The number of the IJDC volume in which the submission will be published.

`\issue{<number>}` The number of the IJDC issue in which the submission will be published.

`\date{<year>}` The year in which the submission will be published.

## Document Body

When it comes to writing the body of the submission, the template should allow you to use the usual  $\text{\LaTeX}$  markup without much adaptation. So, for example, you would start as in Figure 1.

```
\begin{document}
\maketitle

\begin{abstract}
Text of the abstract\dots
\end{abstract}

\section{Introduction}

The text of the introduction starts here\dots
```

**Figure 1.** Sample code for the beginning of an IJDC submission or IDCC full paper.

Please note that if submitting an abstract or proposal to the IDCC instead of a full paper, you should *not* use the abstract environment. Instead, start with a section headed ‘Abstract’ as in Figure 2. Further guidance on how to write such submissions is given on the conference website.

```
\begin{document}
\maketitle

\section{Abstract}

The text of the abstract or proposal starts here\dots
```

**Figure 2.** Sample code for the beginning of an IDCC abstract or proposal.

IJDC and IDCC papers follow the formatting conventions specified by the American Psychological Association (APA, 2010), with a few minor changes. There are some instances where this affects how you write your submission.

### Headings

Five levels of heading are defined ( $\text{\section}$  down to  $\text{\subparagraph}$ ) but most authors only need the first one or two levels.  $\text{\section}$  and  $\text{\subsection}$  headings should be written in title case, that is, with Each Significant Word Given an Initial Capital, while the remaining headers should be written in sentence case as if running text. Do not end your heading names with full stops/periods.

## Quotations

Quotations should be put in a quote environment, wrapped in inverted commas, with the citation placed in parentheses at the end.

‘Cras porttitor dictum lacus. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. In consectetur, diam at volutpat elementum, libero lectus pulvinar sem.’ (Borgman, 2007)

## Tables

- Table text should be in the `\small` font size.
- Tables should not use vertical lines to separate columns, and ideally should not use horizontal lines to separate rows in the body of the table; white space and text alignment should be sufficient. The top and bottom rules should be drawn with `\toprule` and `\bottomrule` respectively, with other rules drawn with `\midrule` or `\cmidrule`. See the documentation of the `booktabs` package for more information.
- Text in the body of tables should normally be left-aligned. Numeric data should be aligned at the decimal point among itself but centred with respect to the heading; the D column type from the `dcolumn` package and the S column type from the `siunitx` package are particularly useful for this.
- Where decked (subdivided) headings are used, there should be a border beneath the upper-level heading (column spanner) indicating to which of the lower-level headings it applies.
- Empty cells can either be left blank or represented by an em dash. A blank cell indicates non-applicability, while an em dash signifies that the data was not collected or has been omitted.
- Captions should end in a full stop/period.

Table 1 demonstrates these features. The code used to produce the table is shown in Figure 3 (note that the `dcolumn` package was loaded in the preamble).

Note the different relative positions of the table and figure captions.

## Reference List and Citations

To use `biblatex` for your reference list, add the following to your preamble:

```
\usepackage[style=apa]{biblatex}  
\addbibresource{<bib file>.bib}  
\DeclareLanguageMapping{british}{british-apa}
```

and include `\printbibliography` at the end of the document.

To use `apacite` for your reference list, add the following to your preamble:

**Table 1.** Papers and articles published in the IJDC in 2008 and 2009.

Issue	Frequency		Percentage	
	Peer-reviewed	General	Peer-reviewed	General
3(1)	9	7	56.3	43.8
3(2)	5	7	41.7	58.3
4(1)	10	4	71.4	28.6
4(2)	8	6	57.1	42.9
4(3)	3	15	16.7	83.3

```
\usepackage{apacite}
\bibliographystyle{apacite}
```

and include `\bibliography{< bib file >}` at the end of the document.

In-text citations are given parenthetically in author–date format. If author forms part of the narrative, as with Rinaldo, Warnement, Baione, Kalfatovic, and Fraser (2011), only the date is added in parenthesis, otherwise both author and date are given (Smith et al., 2003). Where multiple citations are given at once, the order should be the same as in the reference list, i.e. alphabetically by author, with co-authored works coming after singly-authored works, then chronologically (Borgman, Wallis, & Enyedy, 2006; Digital Curation Centre, 2005; Mazairac & Beetz, in press; Santini, 2004a, 2004b; Smith et al., 2003; Witten & Frank, 2005). Please consult the documentation of the package you are using for how to achieve this.

Please do not cite entire websites through the reference list mechanism. Instead, provide the title of the website (in English) and the URL in a footnote.<sup>1</sup> If the title of the website is not clear from the visible pages, the contents of the HTML title element may be used. Other explanatory notes, whether about the body text or cited items, should also be given as footnotes rather than as endnotes or reference list annotations.<sup>2</sup>

Please provide digital object identifiers (DOIs) for referenced items where available.

The data underlying the results presented in the submission should be placed in an appropriate custodial environment and cited (Waterton, Watson, & Norton, 2013), with the reference placed in the reference list. The biblatex-apa package provides a data entry type which should be used for datasets; the entrysubtype value should be ‘Data set’ or similar. If using apacite, use the misc entry type with a type value of ‘Data set’ or similar.

## Acknowledgements

Any acknowledgements should be placed in a section immediately before the references.

<sup>1</sup> Digital Curation Centre: <http://www.dcc.ac.uk/>  
<sup>2</sup> This avoids unnecessary page turning or scrolling.

```
\begin{table}
\caption{Papers and articles published in the IJDC in 2008 and 2009.}
\label{tab:issues}
\centering\small
\begin{tabular}{1D..{2.0}D..{2.0}D..{2.1}D..{2.1}}
\toprule
& \multicolumn{2}{c}{Frequency} & \multicolumn{2}{c}{Percentage} \\
\cmidrule(lr){2-3}\cmidrule(l){4-5}
Issue
& \multicolumn{1}{c}{Peer-reviewed} & \multicolumn{1}{c}{General} \\
& \multicolumn{1}{c}{Peer-reviewed} & \multicolumn{1}{c}{General} \\
\midrule
3(1) & 9 & 7 & 56.3 & 43.8 \\
3(2) & 5 & 7 & 41.7 & 58.3 \\
4(1) & 10 & 4 & 71.4 & 28.6 \\
4(2) & 8 & 6 & 57.1 & 42.9 \\
4(3) & 3 & 15 & 16.7 & 83.3 \\
\bottomrule
\end{tabular}
\end{table}
```

**Figure 3.** Code used to typeset Table 1.

## References

- American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.
- Borgman, C. L. (2007). *Scholarship in the digital age: Information, infrastructure, and the Internet*. Cambridge, MA: MIT Press.
- Borgman, C. L., Wallis, J. C., & Enyedy, N. (2006). Building digital libraries for scientific data: An exploratory study of data practices in habitat ecology. In J. Gonzalo, C. Thanos, M. F. Verdejo, & R. C. Carrasco (Eds.), *Lecture Notes in Computer Science: Vol. 4172. Research and Advanced Technology for Digital Libraries* (pp. 170–183). doi:10.1007/11863878\_15
- Consultative Committee for Space Data Systems. (2012). *Reference model for an Open Archival Information System (OAIS)* (Magenta Book No. CCSDS 650.0-M-2). Retrieved from <http://public.ccsds.org/publications/archive/650x0m2.pdf>
- Digital Curation Centre. (2005). *Digital curation and preservation: Defining the research agenda for the next decade*. Report of the Warwick Workshop, November 7–8, 2005. Retrieved from [http://www.dcc.ac.uk/webfm\\_send/346](http://www.dcc.ac.uk/webfm_send/346)



- Esanu, J., Davidson, J., Ross, S., & Anderson, W. (2004). Selection, appraisal, and retention of digital scientific data: Highlights of an ERPANET/CODATA workshop. *Data Science Journal*, 3, 227–232. Retrieved from <http://www.jstage.jst.go.jp/browse/dsj>
- Mazairac, W. & Beetz, J. (in press). BIMQL: An open query language for building information models. *Advanced Engineering Informatics*. doi:10.1016/j.aei.2013.06.001
- National Science Foundation, Blue-Ribbon Advisory Panel on Cyberinfrastructure. (2003). *Revolutionizing science and engineering through cyberinfrastructure*. Retrieved from [http://www.nsf.gov/publications/pub\\_summ.jsp?ods\\_key=cise051203](http://www.nsf.gov/publications/pub_summ.jsp?ods_key=cise051203)
- Rinaldo, C., Warnement, J., Baione, T., Kalfatovic, M. R., & Fraser, S. (2011, July). Retooling special collections digitisation in the age of mass scanning. *Ariadne*, 67. Retrieved from <http://www.ariadne.ac.uk/issue67/rinaldo-et-al/>
- Santini, M. (2004a). *A shallow approach to syntactic feature extraction for genre classification*. Paper presented at the Seventh Annual Colloquium for the UK Special Interest Group for Computational Linguistics, Birmingham, UK. Retrieved from <ftp://ftp.itri.bton.ac.uk/reports/ITRI-04-02.pdf>
- Santini, M. (2004b). *State-of-the-art on automatic genre identification* (Technical Report No. ITRI-04-03). Information Technology Research Institute. Retrieved from <ftp://ftp.itri.bton.ac.uk/reports/ITRI-04-03.pdf>
- Smith, M., Barton, M., Bass, M., Branschovsky, M., McClellan, G., Stuve, D., & Walker, J. H. (2003). DSpace: An open source dynamic digital repository. *D-Lib Magazine*, 9(1). doi:10.1045/january2003-smith
- Waterton, C., Watson, N., & Norton, L. (2013). *Understanding and acting in Loweswater, 2007–2010* [Data set]. Colchester, UK: UK Data Archive. doi:10.5255/UKDA-SN-7359-1
- Witten, I. H. & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques* (2nd ed.). San Francisco, CA: Morgan Kaufmann.

## Appendix A: Change History

- v1.5** 2016-08-05  
Added details of IDCC 2017. Slightly refactored code to convert dccpaper-base.tex into a package, dccpaper-base.sty.
- v1.4.1** 2015-06-22  
Fixed bug preventing compilation in DVI mode.
- v1.4** 2015-05-22  
Added details of IDCC 2016. Improved whitespace handling. Fixed bug triggered by demo option. Fixed missing use of \Authfont. Added missing DOI tweak for biblatex-apa.
- v1.3.2** 2015-01-21  
Removed dependence on user supplying a title. Fixed bug triggered by numbered sections.
- v1.3.1** 2014-10-14  
Fixed typographical error relating to details of IDCC 2015.
- v1.3** 2014-08-07  
Added documentation of \conference command. Improved display of footnotes, footnote markers and itemize/enumerate lists. Fixed a bug in the handling of author information. Fixed and improved how the transition from first to subsequent page geometry is achieved. Updated the details of IDCC 2015.
- v1.2** 2014-04-11  
Added implementation, installation and licence sections to the Appendix of the documentation. Moved the majority of the file postamble information to the README, and synchronized the latter with the GitHub version. Improved the adaptation of apacite referencing to the house style.
- v1.1** 2014-03-06  
Refactored the source for distribution through CTAN, and to allow the addition of the idcc class.
- v1.0** 2013-12-18  
First public release of ijdc-v9 class.

## Appendix B: Implementation

### ijdc-v9.cls

```

1  %%
2  %% This is file 'ijdc-v9.cls',
3  %% generated with the docstrip utility.
4  %%
5  %% The original source files were:
6  %%
7  %% dccpaper.dtx (with options: 'ijdc9')
8  %%
9  %% -----
10 %% Revealing the Detailed Lineage of Script Outputs using Hybrid Provenance
11 %% Author: Alex Ball
12 %% E-mail: a.ball@ukoln.ac.uk
13 %% License: Released under the LaTeX Project Public License v1.3c or later
14 %% See: http://www.latex-project.org/lppl.txt
15 %% -----
16 %%
17 \def\Version{2016/08/05 v1.5}
18 \NeedsTeXFormat{LaTeX2e}[1999/12/01]
19 \ProvidesClass{ijdc-v9}
20 [\Version\space Class for submissions to the International Journal of Digital
   Curation, volume 9 onwards.]

```

The type of paper is recorded in `\dccp@type`. The possible values are `GeneralArticle`, `Peer-ReviewedPaper`, or `Editorial`. The latter is provided through the constant `\dccp@editorial` so we can test for it later with ease.

```

21 \def\dccp@type{General Article}
22 \def\dccp@editorial{Editorial}

```

These three options set `\dccp@type` appropriately. The `editorial` option additionally switches page numbers to roman numerals. The `article` option is redundant, but included for symmetry.

```

23 \DeclareOption{editorial}{\let\dccp@type=\dccp@editorial%
24   \AtBeginDocument{% Editorials use Roman numerals for page numbers
25     \pagenumbering{roman}%
26     \renewcommand{\thelastpage}{\@roman\c@lastpage}%
27   }%
28 }
29 \DeclareOption{paper}{\def\dccp@type{Peer-Reviewed Paper}}
30 \DeclareOption{article}{\def\dccp@type{General Article}}
31 \ProcessOptions\relax

```

These definitions are used to fill out the headers and footers. They include some details that are defined in `dccpaper-base.sty` and provided by the user: `\thedate`, `\thevolume`, `\theissue`, `\thelastpage`, `\thedoi`.

```

33 \def\dccp@publ@long{International Journal of Digital Curation}
34 \def\dccp@publ@short{IJDC}
35 \def\dccp@publ@msg{The \emph{\dccp@publ@long} is an international journal
36 committed to scholarly excellence and dedicated to the advancement of digital
37 curation across a wide range of sectors. The \dccp@publ@short\ is published by
38 the University of Edinburgh on behalf of the Digital Curation Centre. ISSN:
39 1746-8256. URL: \url{http://www.ijdc.net/}}
40 \def\dccp@titlefoot@bib{%
41   \dccp@publ@long\
42   \thedate, Vol.\ \thevolume, Iss.\ \theissue, \thepage--\thelastpage.%
43 }
44 \def\dccp@titlefoot@doi{%
45   \url{http://dx.doi.org/\thedoi}\
46   DOI: \thedoi
47 }
48 \def\dccp@normhead@doi{doi:\thedoi}

```

`\dccp@subject` will be used for the PDF subject, and will specify the journal and the issue.

```

49 \def\dccp@subject{\dccp@publ@long, \thedate, Vol.\ \thevolume, Iss.\ \theissue}

```

The class is based on the article class, and uses A4 paper.

```

51 \LoadClass[a4paper,12pt,twoside]{article}

```

The rest of the formatting happens in the following file.

```

53 \usepackage{dccpaper-base}
54 %%
55 %% Copyright (C) 2015 Digital Curation Centre, University of Edinburgh
56 %% <info@dcc.ac.uk>
57 %%
58 %% End of file 'ijdc-v9.cls'.

```

## idcc.cls

```

1  %%
2  %% This is file 'idcc.cls',
3  %% generated with the docstrip utility.
4  %%
5  %% The original source files were:
6  %%
7  %% dccpaper.dtx (with options: 'idcc')
8  %%

```

```

9  %% -----
10 %% Revealing the Detailed Lineage of Script Outputs using Hybrid Provenance
11 %% Author: Alex Ball
12 %% E-mail: a.ball@ukoln.ac.uk
13 %% License: Released under the LaTeX Project Public License v1.3c or later
14 %% See: http://www.latex-project.org/lppl.txt
15 %% -----
16 %%
17 \def\Version{2016/08/05 v1.5}
18 \NeedsTeXFormat{LaTeX2e}[1999/12/01]
19 \ProvidesClass{idcc}
20   [\Version\space Class for submissions to the International Digital Curation
    Conference.]

```

These definitions are used to fill out the headers and footers. The ones for `\dccp@publ@short` and `\dccp@publ@msg` are a fallback, defined here but ordinarily overridden using a package option.

```

21 \def\dccp@publ@long{International Digital Curation Conference}
22 \def\dccp@publ@short{IDCC}
23 \def\dccp@publ@msg{The \dccp@publ@long\ takes place on [TBC] in [TBC]. URL:
24 \url{http://www.dcc.ac.uk/events/international-digital-curation-conference-idcc}}

```

`\dccp@subject` will be used for the PDF subject. The value here is again a fallback, ordinarily overridden using a package option.

```

25 \def\dccp@subject{\dccp@publ@long}

```

Paper type is recorded in `\def\dccp@type`. The values can be `Research_Paper`, `Practice_Paper`, `Poster`, `Demonstration`, or `Workshop`.

```

27 \def\dccp@type{Practice Paper}

```

These options set the value of `\def\dccp@type` appropriately. The `practice` option is redundant, but included for symmetry.

```

28 \DeclareOption{research}{\def\dccp@type{Research Paper}}
29 \DeclareOption{data}{\def\dccp@type{Data Paper}}
30 \DeclareOption{poster}{\def\dccp@type{Poster}}
31 \DeclareOption{demo}{\def\dccp@type{Demonstration}}
32 \DeclareOption{bof}{\def\dccp@type{Birds of a Feather}}
33 \DeclareOption{workshop}{\def\dccp@type{Workshop}}
34 \DeclareOption{practice}{\def\dccp@type{Practice Paper}}

```

The details for each conference are slightly different, so they each have an associated option for overriding the fallback definitions for `\dccp@publ@short`, `\dccp@publ@msg` and `\dccp@subject`.

```

35 \DeclareOption{15}{%
36   \def\dccp@publ@short{IDCC15}
37   \def\dccp@publ@msg{The 10th \dccp@publ@long\ takes place on 9--12 February
38     2015 in London. URL: \url{http://www.dcc.ac.uk/events/idcc15/}}
39   \def\dccp@subject{10th \dccp@publ@long, 2015}
40 }
41 \DeclareOption{16}{%
42   \def\dccp@publ@short{IDCC16}
43   \def\dccp@publ@msg{The 11th \dccp@publ@long\ takes place on 22--25 February
44     2016 in Amsterdam. URL: \url{http://www.dcc.ac.uk/events/idcc16/}}
45   \def\dccp@subject{11th \dccp@publ@long, 2016}
46 }
47 \DeclareOption{17}{%
48   \def\dccp@publ@short{IDCC17}
49   \def\dccp@publ@msg{The 12th \dccp@publ@long\ takes place on 20--23 February
50     2017 in Edinburgh. URL: \url{http://www.dcc.ac.uk/events/idcc17/}}
51   \def\dccp@subject{12th \dccp@publ@long, 2017}
52 }
53 \ProcessOptions\relax

```

The `demo` option, if passed, would be passed on to the graphics package wreaking havoc with each and every included image. To avoid this, we clear it from the list of class options using code from Enrico Gregorio.

```

55 \def\@clearglobaloption#1{%
56   \def\@tempa{#1}%
57   \def\@tempb{\@gobble}%
58   \@for\next:=\@classoptionslist\do{%
59     \ifx\next\@tempa
60       \message{Option \next\space has been hidden from subsequent packages}%
61     \else
62       \edef\@tempb{\@tempb,\next}%
63     \fi
64   }%
65   \let\@classoptionslist\@tempb
66   \expandafter\ifx\@tempb\@gobble
67     \let\@classoptionslist\@empty
68   \fi
69 }
70 \@clearglobaloption{demo}

```

The class is based on the article class, and uses A4 paper.

```

72 \LoadClass[a4paper,12pt,twoside]{article}

```

The rest of the formatting happens in the following file.

```

74 \usepackage{dccpaper-base}
75 %%
76 %% Copyright (C) 2015 Digital Curation Centre, University of Edinburgh
77 %% <info@dcc.ac.uk>

```

```

78 %%
79 %% End of file 'idcc.cls'.

```

### dccpaper-base.sty

```

1  %%
2  %% This is file 'dccpaper-base.sty',
3  %% generated with the docstrip utility.
4  %%
5  %% The original source files were:
6  %%
7  %% dccpaper.dtx (with options: 'base')
8  %%
9  %% -----
10 %% Revealing the Detailed Lineage of Script Outputs using Hybrid Provenance
11 %% Author: Alex Ball
12 %% E-mail: a.ball@ukoln.ac.uk
13 %% License: Released under the LaTeX Project Public License v1.3c or later
14 %% See: http://www.latex-project.org/lppl.txt
15 %% -----
16 %%
17 \def\Version{2016/08/05 v1.5}
18 \ProvidesPackage{dccpaper-base.sty}
19   [\Version\space Common class code for IJDC and IDCC papers.]

```

The dccpaper classes are deliberately very similar. This package contains the common code.

We use British English orthography.

```

20 \RequirePackage[british]{babel}

```

The macro patching commands from etoolbox come in useful for handling author and date information, and also for compatibility with apacite.

```

21 \RequirePackage{etoolbox}

```

The dccpaper classes use Times as the main text font. We prefer newtx as it provides support for mathematics, but the standard times package will do. In case they are needed, we also provide Helvetica for the sans serif font and Computer Modern Teletype for the monospaced.

```

22 \RequirePackage[T1]{fontenc}
23 \IfFileExists{newtxtext.sty}%
24   {\RequirePackage{newtxtext,newtxmath}}%
25   {\RequirePackage{times}}
26 \RequirePackage[scaled=0.92]{helvet}
27 \renewcommand{\ttdefault}{cmtt}

```

We will need support for included graphics and colour. The structural elements are a medium turquoise, while the links are slightly darker to make them easier to read on screen.

```
28 \RequirePackage{graphicx}
29 \IfFileExists{xcolor.sty}%
30   {\RequirePackage{xcolor}}%
31   {\RequirePackage{color}}
32 \definecolor[named]{struct}{rgb}{0,0.5,0.5}
33 \definecolor[named]{links}{rgb}{0,0.4,0.4}
```

We will calculate some lengths later.

```
34 \RequirePackage{calc}
```

Ragged right text is easier to read on screen, but fully justified text looks better. The `\raggedyright` layout from Peter Wilson's memoir class (2005/09/25 v1.618) is an excellent compromise. The code below replicates it.

First we save the original definitions of `\` and `\parindent` as `\OrigLineBreak` and `\RaggedParindent` respectively.

```
36 \let\OrigLineBreak\
37 \newdimen\RaggedParindent
38 \setlength{\RaggedParindent}{\parindent}
```

The `\raggedyright` layout more or less lays text out as with full justification, but then lets the shorter lines relax a bit from the right margin. It is the default for DCC papers.

```
40 \newcommand{\raggedyright}[1][2em]{%
41   \let\\\@centercr\@rightskip \z@ \@plus #1\relax
42   \rightskip\@rightskip
43   \leftskip\z@skip
44   \parindent\RaggedParindent}
45 \AtBeginDocument{\raggedyright}
```

The `\flushletright` layout restores full justification, in case it is needed.

```
47 \newcommand*{\flushletright}{%
48   \let\\\OrigLineBreak
49   \leftskip\z@skip
50   \rightskip\leftskip
51   \parfillskip\@flushglue
52   \everypar{}}
```

PdfTeX 1.40.15 introduced the facility to include dummy interword spaces to improve text extraction and reflow. We test for this facility and use it if available.



```

54 \RequirePackage{ifpdf}
55 \ifpdf
56   \ifdef{\pdfinterwordspaceon}{\pdfinterwordspaceon}{}
57   \fi

```

The classes have some special metadata requirements. We start with the author information.

The macro `\thecorrespondence` is used in the title page footer for the name, postal address and email address of the corresponding author.

```

59 \def\thecorrespondence{}
60 \newcommand*{\correspondence}[1]{\def\thecorrespondence{#1}}

```

The handling of authors here is inspired by Patrick W Daly's `authblk`, (2001/02/27 1.3), and defines the familiar user commands. Authors are presented in blocks, one affiliation but perhaps several authors per block.

We make the presentation of the author information configurable (just in case) with some hooks and lengths:

- `\Authfont` is the font used for author names;
- `\Affilfont` is the font used for affiliations;
- `\affilsep` is the line spacing between author names and affiliations;
- `\authblksep` is the line spacing between author name/affiliation blocks.

```

61 \newcommand*{\Authfont}{\normalfont}
62 \newcommand*{\Affilfont}{\normalfont\small}
63 \newlength{\affilsep}\setlength{\affilsep}{0pt}
64 \newlength{\authblksep}\setlength{\authblksep}{1.5\baselineskip}

```

The `\AuthorBlock` command will be used to collect all the author information.

```

65 \newcommand{\AuthorBlock}{}

```

The `\MainAuthor` command collects abbreviated author information for use in the headers.

```

66 \newcommand{\MainAuthor}{}

```

The `\dccp@author` command, meanwhile, collects a full list of authors for the PDF metadata.

```

67 \newcommand{\dccp@author}{}

```

We define counters for

- the total number of authors defined;
- the number of authors in the current block;
- the number of blocks defined.

```
68 \newcounter{authors}
69 \newcounter{authorsinblock}
70 \newcounter{block}
```

The `block` counter will increase once in the preamble, and again when the information is typeset, so we need to reset it at the beginning of the document.

```
71 \AtBeginDocument{\setcounter{block}{0}}
```

The new definition of the author command starts here.

```
72 \renewcommand{\author}[1]{%
```

If this is the first or second `\author` command, we add the name to our abbreviated list of authors. Otherwise, we replace the name of the second and subsequent authors with `et al.` in that list.

At the same time, we use a simpler technique to populate `\dccp@author`.

```
73 \ifnum\theauthors=0
74   \def\dccp@author{#1}%
75   \def\MainAuthor{#1}%
76 \else
77   \appto\dccp@author{, #1}%
78   \ifnum\theauthors=1%
79     \def\OtherMainAuthors{ and #1}%
80     \appto\MainAuthor{\OtherMainAuthors}%
81   \else
82     \ifnum\theauthors=2%
83       \def\OtherMainAuthors{ et al.}%
84     \fi
85   \fi
86 \fi
87 \stepcounter{authors}%
```

Each block has its author names collected in a macro like `\blocki@auth`, and its affiliation collected in a macro like `\blocki@affil` (the `i` is a serial number).

If this is the first author in a block, we need to create the block and add it to `\AuthorBlock`; second and subsequent blocks are preceded by a `\quad` of space.

```
88 \ifnum\theauthorsinblock=0%
89   \stepcounter{block}%
90   \expandafter\def\csname block\roman{block}@auth\endcsname{\Authfont#1}%
91   \ifnum\theblock>1\appto\AuthorBlock{\quad}\fi
92   \appto\AuthorBlock{%
```

```

93     \stepcounter{block}%
94     \begin{minipage}[t]{0.45\textwidth}\centering
95     \csname block\roman{block}@auth\endcsname
96     \ifx\undefined\csname block\roman{block}@affil\endcsname
97     \else
98         \[\affilsep]\csname block\roman{block}@affil\endcsname
99     \fi
100     \end{minipage}%
101     \rule{0pt}{2\baselineskip}%
102 }
103 \else

```

Otherwise we just add the name to the right `\blocki@auth`-style macro.

```

104     \csappto{block\roman{block}@auth}{, #1}%
105     \fi
106     \stepcounter{authorsinblock}%
107 }

```

The `\affil` command adds an affiliation to the current block and closes it by resetting the `authorsinblock` counter.

```

108 \newcommand{\affil}[1]{%
109     \expandafter\def\csname block\roman{block}@affil\endcsname{\Affilfont#1}%
110     \setcounter{authorsinblock}{0}%
111 }

```

The `\HeadTitle` collects the abbreviated title for use in the headers.

```

113 \newcommand*{\HeadTitle}{}

```

We wrap the normal `\title` command with code to populate `\HeadTitle` with the optional argument if provided, or the mandatory one otherwise. We also provide a persistent `\thetitle` macro, stripped of any `\thanks`.

```

114 \let\ProperTitle=\title
115 \renewcommand{\title}[2][\empty]{
116     \ifx\empty #1%
117         \renewcommand*{\HeadTitle}{#2}%
118     \else
119         \renewcommand*{\HeadTitle}{#1}%
120     \fi%
121     \begingroup\let\footnote\@gobble
122     \ProperTitle{#2}%
123     \begingroup
124         \renewcommand{\thanks}[1]{}
125         \protected@xdef\thetitle{#2}
126     \endgroup\endgroup
127 }

```

To prevent L<sup>A</sup>T<sub>E</sub>X throwing an error if no title is supplied, we supply a fallback value and issue a warning instead.

```
128 \def\thetitle{%
129   Please supply a title
130   \ClassWarning{dccpaper}{%
131     Please supply a title in the preamble with \noexpand\title{}%
132   }
133 }
```

We make `\thedata` persistent, borrowing the technique used in Peter Wilson’s memoir class (2005/09/25 v1.618).

```
135 \pretocmd{\date}{\begingroup\let\footnote\@gobble}{\}{\}%
136 \apptocmd{\date}{%
137   \begingroup
138   \renewcommand{\thanks}[1]{\%
139     \protected@xdef\thedata{#1}
140   \endgroup\endgroup%
141 }{\}{\}
```

We ensure `\thedata` is defined using a dummy date.

```
142 \date{20xx}
```

IJDC articles have extra bibliographic information:

- `\volume` sets the volume number, `\thevolume`;
- `\issue` sets the issue number, `\theissue`;
- `\subno` sets the submission number, `\thesubno`.

These numbers are used to build the DOI, `\thedoi`.

```
144 \newcommand*{\thevolume}{0}
145 \newcommand*{\volume}[1]{\renewcommand*{\thevolume}{#1}}
146 \newcommand*{\theissue}{0}
147 \newcommand*{\issue}[1]{\renewcommand*{\theissue}{#1}}
148 \newcommand*{\thesubno}{0}
149 \newcommand*{\subno}[1]{\renewcommand*{\thesubno}{#1}}
150 \newcommand*{\thedoi}{10.2218/ijdc.v\thevolume i\theissue .\thesubno}
```

They also display the page range. The following code was borrowed from Peter Wilson’s memoir class (2005/09/25 v1.618). It defines a counter `lastpage` which, on the second run, will contain the number of the last page.

```
151 \newcounter{lastpage}
152 \setcounter{lastpage}{0}
153 \newcommand{\dol@stpage}{%
154   \if@filesw
```

```

155 \addtocounter{page}{-1}%
156 \immediate\write\@auxout%
157   {\string\setcounter{lastpage}{\the\c@page}}%
158 \stepcounter{page}%
159 \fi
160 }
161 \AtBeginDocument{\AtEndDocument{\clearpage\dol@stpage}}

```

DCC papers display some important dates. We collect these in `\dccp@dates`, initially setting the value to something sensible for papers in draft.

```

163 \def\dccp@dates{\emph{Draft from} \today}

```

Several types of date can be added:

- `\submitted` for when the authors submitted the paper (intended for IDCC papers).
- `\received` for when the paper was received by the editorial board (intended for IJDC papers).
- `\revised` for when the most recent version was received by the editorial board.
- `\accepted` for when the paper was accepted by the editorial board.

```

164 \newcommand*{\submitted}[1]{%
165   \def\dccp@dates{\emph{Submitted} #1}}
166 \newcommand*{\received}[1]{%
167   \def\dccp@dates{\emph{Received} #1}}
168 \newcommand*{\revised}[1]{%
169   \appto\dccp@dates{%
170     \space\space\space\textbar\space\space\space
171     \emph{Revision received} #1}%
172   }
173 \newcommand*{\accepted}[1]{%
174   \appto\dccp@dates{%
175     \space\space\space\textbar\space\space\space
176     \emph{Accepted} #1}%
177   }

```

IJDC papers need to say if they had a previous life as a conference paper. This statement goes in `\dccp@conf`, which is initially empty. The user command for setting this text is `\conference`.

```

178 \let\dccp@conf=\empty
179 \newcommand*{\conference}[1]{%
180   \renewcommand*{\dccp@conf}{An earlier version of this paper was presented at
181     #1.}%

```

The macro `\FixTextHeight` will be useful when switching from the first page geometry to the regular geometry for the rest of the paper. It is based on code from Hideo Umeki's geometry package (2002/07/08 v3.2).

```
183 \newcommand{\FixTextHeight}{%
184   \setlength\@tempdima{\textheight}%
185   \addtolength\@tempdima{-\topskip}%
186   \@tempcnta\@tempdima
187   \@tempcntb\baselineskip
188   \divide\@tempcnta\@tempcntb
189   \setlength\@tempdimb{\baselineskip}%
190   \multiply\@tempdimb\@tempcnta
191   \advance\@tempdima-\@tempdimb
192   \global\advance\footskip\@tempdima
193   \multiply\@tempdima\tw@
194   \ifdim\@tempdima>\baselineskip
195     \addtolength\@tempdimb{\baselineskip}%
196     \global\advance\footskip-\baselineskip
197   \fi
198   \addtolength\@tempdimb{\topskip}%
199   \global\textheight\@tempdimb
200 }
```

The width of the textblock (on all pages) is 150mm, which on A4 paper implies margins of 30mm each. (Making both horizontal margins the same in a two-sided context makes the paper more pleasant to read on screen).

```
202 \setlength{\textwidth}{150mm}
203 \setlength{\oddsidemargin}{30mm - \hoffset - 1in}
204 \setlength{\evensidemargin}{30mm - \hoffset - 1in}
```

It is rare to have marginal notes, but in case we ever do, we centre them in the margin.

```
205 \setlength{\marginparwidth}{30mm - 2\marginparsep}
```

We also want a distance of 15mm from the top of the page to the top of the header, and two blank lines between the bottom of the header and the top of the textblock.

```
206 \setlength{\topmargin}{15mm - \voffset - 1in}
207 \setlength{\headsep}{2\baselineskip}
```

IJDC editorials have slightly different headers and footers. This requires testing for `\dccp@editorial` if it exists. In case it doesn't, we provide it.

```
209 \providecommand{\dccp@editorial}{Editorial}
```

The height of the footer can vary a lot. To keep it a fixed distance from the bottom of the page rather than the top, we need to vary the `\textheight` accordingly. This means we need to measure the height of the footer. (The header is more predictable but we may as well measure it while we are at it).

Here we define the header and footer of the title page (i.e. the title page style), making sure we save them to auxiliary macros `\TitleHead` and `\TitleFoot` so we can measure them.

```

210 \def\ps@title{%
211   \def\@oddhead{%
212     \begin{minipage}{\textwidth}%
213     \centering
214     \LARGE\bfseries\color{struct}%
215     \ifx\dccp@type\dccp@editorial
216       \dccp@publ@long
217     \else
218       \dccp@publ@short\space\space\textbar\space\space\emph{\dccp@type}%
219     \fi
220     \par
221   \end{minipage}%
222 }%
223 \let\@evenhead=\@oddhead
224 \let\TitleHead=\@oddhead
225 \def\@oddfoot{%
226   \begin{minipage}[b]{\textwidth}%
227     \fontsize{9pt}{11pt}\selectfont
228     \ifx\dccp@type\dccp@editorial
229     \else
230       {\centering\dccp@dates\par}
231       \bigskip
232       Correspondence should be addressed to \thecorrespondence\par
233       \bigskip
234     \fi
235     \ifx\empty\dccp@conf
236     \else
237       \dccp@conf\par
238       \bigskip
239     \fi
240     \dccp@publ@msg\par
241     \bigskip
242     \begin{minipage}[b]{\linewidth - 25mm}
243       Copyright rests with the authors. This work is released under a Creative
244       Commons Attribution 4.0 International Licence. For details please see
245       \url{http://creativecommons.org/licenses/by/4.0/}%
246     \end{minipage}\hfill
247     \begin{minipage}[b]{19mm}
248       \href{http://creativecommons.org/licenses/by/4.0/}%
249       {\includegraphics[width=\hsize]{dccpaper-by}}%
250     \end{minipage}
251     \par
252     \bigskip
253     \makebox[0pt][l]{\parbox{0.4\hsize}{%
254       \ifx\undefined\dccp@titlefoot@bib\else\dccp@titlefoot@bib\fi
255     }}\hfill
256     \makebox[0pt][c]{\normalsize\thepage}\hfill
257     \makebox[0pt][r]{\parbox{0.4\hsize}{%
258       \raggedleft\ifx\undefined\dccp@titlefoot@doi\else\dccp@titlefoot@doi\fi

```

```

259     }}%
260     \end{minipage}%
261 }%
262 \let\@evenfoot=\@oddfoot
263 \let\TitleFoot=\@oddfoot
264 }

```

We set the normal page style to title here so that `\TitleHead` and `\TitleFoot` are defined, but we will override it with the `dccpaper` page style later.

```

265 \pagestyle{title}

```

The first page should use the title page style, however.

```

266 \AtBeginDocument{\thispagestyle{title}}

```

Here are the normal headers and footers (i.e. the `dccpaper` page style). We save them to `\NormalHead` and `\NormalFoot`, again so we can measure them.

```

268 \def\ps@dccpaper{%
269   \def\@oddhead{%
270     \begin{minipage}{\textwidth}\frenchspacing
271     {%
272       \fontsize{9pt}{11pt}\selectfont
273       \ifx\undefined\dccp@normhead@doi\else\dccp@normhead@doi\fi
274     }\hfill
275     {\MainAuthor}\space\space\space
276     \textcolor{struct}{\textbar}\space\space\space
277     \thepage\par
278     \vskip6pt\color{struct}{\hrule height 1bp}\par
279   \end{minipage}
280 }%
281 \def\@evenhead{%
282   \begin{minipage}{\textwidth}
283     \thepage\space\space\space
284     \textcolor{struct}{\textbar}\space\space\space
285     {\HeadTitle}\hfill
286   {%
287     \fontsize{9pt}{11pt}\selectfont
288     \ifx\undefined\dccp@normhead@doi\else\dccp@normhead@doi\fi
289   }\par
290   \vskip6pt\color{struct}{\hrule height 1bp}\par
291 \end{minipage}
292 }%
293 \let\NormalHead=\@oddhead
294 \def\@oddfoot{\begin{minipage}[b]{\textwidth}
295   \centering\bfseries\normalsize\color{struct}
296   \ifx\dccp@type\dccp@editorial
297     \dccp@publ@long
298   \else
299     \dccp@publ@short\space\space\textbar\space\space\emph{\dccp@type}%
300   \fi

```



```

301 \par
302 \end{minipage}}%
303 \let\@evenfoot=\@oddfoot
304 \let\NormalFoot=\@oddfoot
305 }
306 \pagestyle{dccpaper}

```

We need to wait until the author has supplied the necessary information before we can do our measuring and set the remainder of the geometry, so we do it at the end of the preamble. First we put our saved macros into boxes we can measure (i.e. `\dccp@firstpagehead`, `\dccp@firstpagefoot`, `\dccp@restpagehead`, `\dccp@restpagefoot`).

```

308 \AtEndPreamble{
309 \newsavebox{\dccp@firstpagehead}
310 \sbox\dccp@firstpagehead{\normalfont\TitleHead}
311 \newsavebox{\dccp@firstpagefoot}
312 \sbox\dccp@firstpagefoot{\normalfont
313 \def\email#1{#1}\def\url#1{#1}\def\href#1#2{#2}\TitleFoot}
314 \newsavebox{\dccp@restpagehead}
315 \sbox\dccp@restpagehead{\normalfont\NormalHead}
316 \newsavebox{\dccp@restpagefoot}
317 \sbox\dccp@restpagefoot{\normalfont\NormalFoot}

```

We can now set the geometry of the title page...

```

318 \setlength{\headheight}{\ht\dccp@firstpagehead + \dp\dccp@firstpagehead}
319 \setlength{\footskip}{%
320 2\baselineskip + \ht\dccp@firstpagefoot + \dp\dccp@firstpagefoot
321 }
322 \setlength{\textheight}{%
323 \paperheight
324 - 30mm % 15mm top and bottom
325 - \headheight
326 - \headsep
327 - \footskip
328 }

```

... and provide a macro that will reset the geometry for the remaining pages.

```

329 \def\dccp@resetgeometry{%
330 \setlength{\headheight}{\ht\dccp@restpagehead + \dp\dccp@restpagehead}
331 \global\headheight=\headheight
332 \setlength{\footskip}{%
333 2\baselineskip + \ht\dccp@restpagefoot
334 }
335 \global\footskip=\footskip
336 \setlength{\textheight}{%
337 \paperheight
338 - 30mm % 15mm top and bottom
339 - \headheight
340 - \headsep

```

```

341     - \footskip
342   }
343   \FixTextHeight
344   \global\textheight=\textheight
345 }
346 }

```

The `\maketitle` command is redefined to the correct formatting. At the end it sets a hook that will reset the geometry when the first page is shipped out, i.e. with effect from the second page. It is here rather than at the end of the abstract in case the abstract itself spills over to the second page.

```

348 \RequirePackage{atbegshi}
349 \renewcommand{\maketitle}{%
350   \null\nobreak\vspace*{-0.528\baselineskip}%
351   \begingroup
352     \centering
353     {\Large\thetitle\par}
354     \vspace{0.7\baselineskip}
355     \AuthorBlock\par
356     \vspace{1.7\baselineskip}
357   \endgroup
358   \AtBeginShipoutNext{\dccp@resetgeometry}%
359 }

```

The abstract environment is redefined in terms of an environment `widequote`, which mimics the `quote` environment, but is a bit wider. We also provide a hook, `\afterabstract`, so that if some annotation needs to be appended to the title page after the abstract, we can do that.

```

361 \newenvironment{widequote}{%
362   \list{}{%
363     \setlength{\rightmargin}{2\parindent}%
364     \setlength{\leftmargin}{2\parindent}%
365   }%
366   \flushletright\item[]%
367 }{%
368   \endlist
369 }
370 \def\afterabstract{}
371 \renewenvironment{abstract}{%
372   \vskip1em%
373   \begin{center}%
374     {\bfseries\abstractname\vspace{-.5em}\vspace{\z@}}%
375   \end{center}%
376   \widequote\footnotesize
377 }{%
378   \endwidequote\afterabstract\newpage
379 }

```

We use the `titlesec` package to give headings the correct formatting. The settings below try to space out headings so they occupy an integer number of normal lines (an attempt at grid typesetting). They are a little complicated because we want it to work even if the heading appears at the top of the page.

```

381 \RequirePackage{titlesec}
382 \titlespacing*{\section}{0pt}{0pt}{\baselineskip}
383 \titlespacing*{\subsection}{0pt}{0pt}{0.6\baselineskip}
384 \titlespacing{\subsubsection}{\parindent}{\baselineskip}{0pt}
385 \titlespacing{\paragraph}{\parindent}{\baselineskip}{0pt}
386 \titlespacing{\subparagraph}{\parindent}{\baselineskip}{0pt}

```

An unfortunate side effect of spacing headings like this is that if a `\subsection` immediately follows a `\section` it forms an unsightly gap. To remedy this, we count how many paragraphs there have been since the last `\section`. Note that as we do not normally number the sections, an automatic reset of the `sectionpars` counter within the `section` counter won't work.

```

387 \newcounter{sectionpars}
388 \let\dccp@old@ep\everypar
389 \newtoks\everypar
390 \dccp@old@ep{\the\everypar\stepcounter{sectionpars}}

```

We need to manually reset `sectionpars` when `\section` is called. Also, the normal font size is 12pt/14.5pt, while `\Large` is 17pt/22pt; so the `\Large` line height =  $1.5172 \times$  normal line height. Nevertheless it seems to work better if we let the heading eat  $0.528\text{\baselineskip}$  into the  $2\text{\baselineskip}$  of padding above it.

```

391 \titleformat{\section}
392 [block]
393 {%
394   \vspace{2\baselineskip}%
395   \nobreak
396   \vspace*{-0.528\baselineskip}%
397   \setcounter{sectionpars}{0}%
398   \filcenter\normalfont\Large\bfseries
399 }
400 {\thesection}
401 {1em}
402 {}

```

The others use a `\normalsize` font so that makes life easier. The format for `\subsection` command includes conditional spacing: if the `sectionpars` counter equals 2, this means the heading immediately follows a `\section`, so less white space is needed.

```

403 \titleformat{\subsection}
404 {%
405   \ifnum\thesectionpars>2%
406   \vspace{2\baselineskip}%

```

```
407 \else
408 \vspace{\baselineskip}%
409 \fi\nobreak
410 \vspace*{-0.6\baselineskip}%
411 \normalfont\normalsize\bfseries
412 }
413 {\thesubsection}
414 {1em}
415 {}
416 \titleformat{\subsubsection}
417 [block]
418 {\normalfont\normalsize\bfseries}
419 {\thesubsubsection}
420 {1em}
421 {}
422 \titleformat{\paragraph}
423 [block]
424 {\normalfont\normalsize\bfseries\itshape}
425 {\thesubsubsection}
426 {1em}
427 {}
428 \titleformat{\subparagraph}
429 [block]
430 {\normalfont\normalsize\itshape}
431 {\thesubsubsection}
432 {1em}
433 {}
```

DCC papers do not typically number their sections.

```
434 \setcounter{secnumdepth}{0}
```

To help with the display of tables we load the array and booktabs packages. As we don't like lines between rows in the table body, we stretch them out a bit so that white space does the job instead.

```
436 \RequirePackage{array,booktabs}
437 \renewcommand{\arraystretch}{1.25}
```

We use the caption package to give captions the right format.

```
439 \RequirePackage
440 [ format=hang
441 , labelsep=period
442 , font=small
443 , labelfont=bf
444 , figureposition=bottom
445 , tableposition=top
446 ]{caption}
```

Footnotes should be set right up against the left margin. They should be set hung and in the same half-ragged style as the main text. They should also, for neatness, be at the bottom of the page regardless of how short it is. The footmisc package helps here.

```
448 \RequirePackage[hang,bottom]{footmisc}
449 \settowidth{\footnotemargin}{\footnotesize\textsuperscript{99}\space}
450 \renewcommand{\footnotelayout}{\raggedyright}
```

Also, if multiple footnotes are set at once, the markers should be separated with superscript commas. The footmisc package should help here but its solution is clobbered by hyperref. So after a footnote is set, we check to see if the next token is also a footnote, and if so, slip a comma in before it.<sup>3</sup> This tweak needs to be done late, `\AtBeginDocument`. Note that the newtx superior figures are a bit lower than normal superscript text.

```
451 \AtBeginDocument{
452   \let\dccp@footnote\footnote
453   \def\dccp@next@token{\relax}%
454   \def\dccp@supercomma{\textsuperscript{,}}%
455   \IfFileExists{newtxtext.sty}%
456     {\def\dccp@supercomma{\raisebox{-0.2ex}{\textsuperscript{,}}}%
457     {}
458   }
459   \newcommand\dccp@check@for@footnote{%
460     \ifx\footnote\dccp@next@token
461       \dccp@supercomma
462     \fi
463   }
464
465   \renewcommand\footnote[1]{%
466     \dccp@footnote{#1}%
467     \futurelet\dccp@next@token\dccp@check@for@footnote
468   }
469 }
```

By default lists are quite loose. These settings help to tighten them.

```
471 \topsep = \z@
472 \partopsep = \z@
473 \appto{\enumerate}{\itemsep = 0.5ex plus 0.25ex minus 0.25ex}
474 \appto{\itemize}{\itemsep = 0.5ex plus 0.25ex minus 0.25ex}
```

A DCC paper should either be using biblatex or apacite for references.

If biblatex is used, we need to ensure that the reference list heading is a normal section rather than a starred one so it appears in the PDF bookmarks.

```
476 \AtBeginDocument{
477   \@ifpackageloaded{biblatex}{%
478     \defbibheading{bibliography}[\refname]{\section{#1}}%
```

<sup>3</sup> This solution was provided at <http://tex.stackexchange.com/q/40072>

We also move the ‘doi:’ portion of a DOI inside the hyperlink.

```
479 \DeclareFieldFormat{doi}{%
480   \ifhyperref{%
481     \href{http://dx.doi.org/#1}{\nolinkurl{doi:#1}}%
482   }{%
483     \nolinkurl{doi:#1}%
484   }%
485 }
486 }
```

If apacite is used, there are a few other adaptations we need to make.

```
487 \@ifpackageloaded{apacite}{%
```

With hyperref loaded, apacite makes the whole of a citation a link to the reference list item. We patch `\@ifauthorsunequalc@de` so only the year portion gets linked.

```
488 \def\@ifauthorsunequalc@de#1{%
489   \if@F@cite
490     \@F@citefalse
491   \else
492     \if@Y@cite
493       {\@BAY}%
494     \fi
495     {\@BBC}%
496   \fi
497   \edef\@cite@undefined{?}%
498   \def\BBA{\@BBA}%
499   \if@A@cite
500     %\hyper@natlinkstart{#1}% We remove this line...
501     {\csname b@\@citeb\APAC@extra@b@citeb\endcsname}%
502     %\hyper@natlinkend% ...and this one.
503     \if@Y@cite
504       {\@BBAY}%
505     \fi
506   \fi
507   \if@Y@cite
508     \hyper@natlinkstart{#1}%
509     {\csname Y@\@citeb\APAC@extra@b@citeb\endcsname}%
510     \hyper@natlinkend
511   \fi
512   \let\BBA\relax
513 }
```

The Spanish language support file defines a different version of `\@ifauthorsunequalc@de`, which might override the patch we have just introduced. So we employ the same test that apacite uses when deciding whether to load that file; if successful, we patch the Spanish version. Note that as apacite loads language support files `\AtBeginDocument`, we have to do our thing after that, `\AfterEndPreamble`.

(Note that as we set the language to British English earlier, this should never be needed, but we try to be resilient to tinkering!)

```

514 \AfterEndPreamble{%
515   \@ifundefined{iflanguage}{%
516     \relax
517   }{%
518     \edef\APAC@tmp{nohyphenation}%
519     \ifx\language\APAC@tmp
520     \else
521       \edef\APAC@tmp{spanish}%
522       \ifx\language\APAC@tmp
523         \def\@ifauthorsunequalc@de#1{%
524           \if@F@cite
525             \@F@citefalse
526           \else
527             \if@Y@cite
528               {\@BAY}%
529             \fi
530             {\@BBC}%
531           \fi
532         \edef\@cite@undefined{?}%
533         \def\BBA{\@BBA}%
534         \@ifundefined{spanishe@\@citeb\APAC@extra@b@citeb}%
535           {}% skip
536           {% Use 'e' instead of 'y' in Spanish
537             \global\let\oldBBA\BBA
538             \global\def\BBA{e\global\let\BBA\oldBBA}%
539           }%
540         \if@A@cite
541           %%\hyper@natlinkstart{#1}% We remove this line...
542           {\csname b@\@citeb\APAC@extra@b@citeb\endcsname}%
543           %%\hyper@natlinkend% ...and this one.
544           \if@Y@cite
545             {\@BBAY}%
546           \fi
547         \fi
548         \if@Y@cite
549           \hyper@natlinkstart{#1}%
550           {\csname Y@\@citeb\APAC@extra@b@citeb\endcsname}%
551           \hyper@natlinkend
552         \fi
553         \let\BBA\relax
554       }%
555     \fi
556   \fi
557 }%

```

Another thing apacite does `\AtBeginDocument` is set the URL style to monospaced. So we reset it back to normal roman type `\AfterEndPreamble`.

```

558   \urlstyle{APACrm}
559 }%

```

We pre-empt apacite's `\providecommand` of `\doi` with our own definition that includes the doi URI scheme label in the link, remembering to remove the one inserted by `\doiprefix`.

```
560 \newcommand{\doi}[1]{\href{http://dx.doi.org/#1}{\nolinkurl{doi:#1}}}%
561 \renewcommand{\doiprefix}{\unskip}%
562 }{}%
563 }
```

Both biblatex and apacite use `\bibitemsep` for the space between bibliography items. Just in case they haven't been loaded, though, we protect our setting of that length with an `\ifx` test.

```
564 \ifx\undefined\bibitemsep
565 \else
566 \setlength{\bibitemsep}{1em plus 1ex minus 1ex}%
567 \fi
568 }
```

As mentioned above, if apacite is used, we can use a package option to ensure that the reference list heading appears in the PDF bookmarks.

```
569 \PassOptionsToPackage{numberedbib}{apacite}
```

We, of course, use hyperref for enhancing the PDF with working links, bookmarks, metadata, etc.

```
571 \usepackage
572 [ colorlinks=true
573 , linkcolor=black
574 , anchorcolor=black
575 , citecolor=links
576 , filecolor=black
577 , menucolor=black
578 , runcolor=black
579 , urlcolor=links
580 ]{hyperref}
```

Links should be in roman type, not monospaced.

```
581 \urlstyle{rm}
```

We provide an `\email` command for displaying the email address of the corresponding author.

```
582 \newcommand*{\email}[1]{\href{mailto:#1}{#1}}
```

Once the user has had a chance to provide the metadata, we can add it to the PDF metadata.



```

583 \AtBeginDocument{%
584   \hypersetup
585     { pdftitle={\thetitle}
586       , pdfauthor={\dccp@author}
587       , pdfsubject={\dccp@subject}
588     }

```

The APA has its own style for line breaks in URLs. The apacite package provides the code for this, but in case biblatex is used instead, we repeat the settings (from 2013/07/21 v6.03) here.

```

589 \ifundefined{Url@force@Tilde}{\def\Url@force@Tilde{\relax}}{}%
590 \def\url@apa@dot{\mathchar"2E}%
591 \def\url@apa@comma{\mathchar"2C}%
592 \def\url@apa@questionmark{\mathchar"3F}%
593 \def\url@apa@exclamation{\mathchar"21}%
594 \def\url@apa@hyphen{\mathchar"2D}%
595 \def\url@apa@underscore{\_}%
596 \def\UrlBreaks{\do\@{\do\|{\do\>{\do\]}{\do\'}{\do+{\do\={\do\#}}}%
597 \def\UrlBigBreaks{\do\/{\do:~{\do@url@hyp}}}%
598 \def\UrlNoBreaks{\do\({\do\{~{\do\<}}{\do\)}%
599 \def\UrlOrds{\do\*{\do\~{\do\'}{\do\"}}}%
600 \def\UrlSpecials{%
601   \do\.{\mathbin{}}\url@apa@dot}%
602   \do\,{\mathbin{}}\url@apa@comma}%
603   \do\~{\mathbin{}}\url@apa@hyphen}%
604   \do\?{\mathbin{}}\url@apa@questionmark}%
605   \do\!{\mathbin{}}\url@apa@exclamation}%
606   \do\_{\mathbin{}}\url@apa@underscore}%
607   \do\ {\Url@space}\do\%{\Url@percent}\do\~M{\Url@space}}%
608   \Url@force@Tilde}%
609 \def\Url@OTnonTT{\do\<{\langle}\do\>{\mathbin{\rangle}}\do
610   \_ {\mathbin{}}\_ \do\|{\mid}\do\{{\lbrace}\do\}{\mathbin{\rbrace}}\do
611   \{\mathbin{\backslash}}\UrlTildeSpecial}
612 }

```

We now embed the Creative Commons licence information in the PDF using an XMP packet. To do this, we employ the same technique as Scott Pakin's hyperxmp (2014/01/02 v2.4). In order to avoid avoid a bug whereby Adobe Acrobat confuses the XMP author information and the regular author information, though, we *only* embed the licence information.

We need to make sure that any characters to appear verbatim in the XMP packet are treated as ordinary characters and not active ones. The likely active characters are symbols and punctuation, so should be treated as other (category 12).

```

614 \begingroup
615 \catcode'\="12
616 \catcode'\&=12
617 \catcode'\#=12
618 \catcode'\<=12

```

```
619 \catcode'\>=12
620 \catcode'\_ =12
```

We construct the XMP packet as the document begins.

```
621 \AtBeginDocument{%
```

For convenience we define `\sp` to be a level of indent, translating to three spaces.

```
622 \def\sp{\space\space\space}
```

The text of the XMP packet is recorded in `\cc@xmp@packet`. We use `^^J` to break lines.

```
623 \long\gdef\cc@xmp@packet{%
624 <?xpacket begin='' id=''?'>^^J%
625 <x:xmpmeta xmlns:x='adobe:ns:meta/'>^^J%
626 <rdf:RDF xmlns:rdf='http://www.w3.org/1999/02/22-rdf-syntax-ns#'>^^J%
627 \sp<rdf:Description rdf:about=''>^^J%
628 \sp\sp xmlns:xapRights='http://ns.adobe.com/xap/1.0/rights/'>^^J%
629 \sp\sp<xapRights:Marked>True</xapRights:Marked>^^J%
630 \sp</rdf:Description>^^J%
631 \sp<rdf:Description rdf:about=''>^^J%
632 \sp\sp xmlns:dc='http://purl.org/dc/elements/1.1/'>^^J%
633 \sp\sp<dc:rights>^^J%
634 \sp\sp\sp<rdf:Alt>^^J%
635 \sp\sp\sp\sp<rdf:li xml:lang='x-default'>This work is licensed under a Creative
Commons Attribution 4.0 International Licence.</rdf:li>^^J%
636 \sp\sp\sp</rdf:Alt>^^J%
637 \sp\sp</dc:rights>^^J%
638 \sp</rdf:Description>^^J%
639 \sp<rdf:Description rdf:about=''>^^J%
640 \sp\sp xmlns:cc='http://creativecommons.org/ns#'>^^J%
641 \sp\sp<cc:license rdf:resource='http://creativecommons.org/licenses/by/4.0/'>^^J%
642 \sp</rdf:Description>^^J%
643 </rdf:RDF>^^J%
644 </x:xmpmeta>^^J%
645 <?xpacket end='r'?'>^^J%
646 }%
647 }
648 \endgroup
```

Different workflows require the XMP packet to be embedded in different ways. PdfTeX can inject objects into PDFs natively.

```
649 \newcommand*{\ccxmp@embed@packet@pdftex}{%
650 \bgroup
651 \pdfcompresslevel=0
652 \immediate\pdfobj stream attr {%
653 /Type /Metadata
654 /Subtype /XML
```

```

655   }\cc@xmp@packet}%
656   \pdfcatalog {/Metadata \the\pdflastobj\space 0 R}%
657   \egroup
658 }

```

The `\pdfmark` command defined by `hyperref` is respected by tools such as `Dvipdf`, `Dvips`, `Dvipsone`, etc.

```

659 \newcommand*\ccxmp@embed@packet@pdfmark{%
660   \pdfmark{%
661     pdfmark=/NamespacePush
662   }%
663   \pdfmark{%
664     pdfmark=/OBJ,
665     Raw={/_objdef \string{ccxmp@packet\string} /type /stream}%
666   }%
667   \pdfmark{%
668     pdfmark=/PUT,
669     Raw={\string{ccxmp@packet\string}
670       2 dict begin
671         /Type /Metadata def
672         /Subtype /XML def
673         currentdict
674       end
675     }%
676   }%
677   \pdfmark{%
678     pdfmark=/PUT,
679     Raw={\string{ccxmp@packet\string} (\cc@xmp@packet)}%
680   }%
681   \pdfmark{%
682     pdfmark=/Metadata,
683     Raw={\string{Catalog\string} \string{ccxmp@packet\string}}%
684   }%
685   \pdfmark{%
686     pdfmark=/NamespacePop
687   }%
688 }

```

`Dvipdfm` has its own `\special` command for inserting PDF objects, but it is a bit basic and requires advance knowledge of how long (in characters) the object is.

The `\ccxmp@count@spaces` macro counts the number of spaces in its parameter through a process of iteration, adding this figure to `\@tempcnta`.

```

689 \def\ccxmp@count@spaces#1 {%
690   \def\ccxmp@one@token{#1}%
691   \ifx\ccxmp@one@token\@empty
692     \advance\@tempcnta by -1
693   \else
694     \advance\@tempcnta by 1
695     \expandafter\ccxmp@count@spaces
696   \fi

```

697 }

The `\ccxmp@count@non@spaces` command counts the number of non-spaces in its argument through a process of iteration, adding this figure to `\@tempcnta`.

```
698 \newcommand*{\ccxmp@count@non@spaces}[1]{%  
699   \def\ccxmp@one@token{#1}%  
700   \ifx\ccxmp@one@token\@empty  
701   \else  
702     \advance\@tempcnta by 1  
703     \expandafter\ccxmp@count@non@spaces  
704   \fi  
705 }
```

The `\ccxmp@string@len` command sets `\@tempcnta` to the number of characters (spaces + non-spaces) in its argument.

```
706 \newcommand*{\ccxmp@string@len}[1]{%  
707   \@tempcnta=0  
708   \expandafter\ccxmp@count@spaces#1 {} %  
709   \expandafter\ccxmp@count@non@spaces#1{}%  
710 }
```

So now, finally, is the command for embedding the packet using Dvipdfm.

```
711 \newcommand*{\ccxmp@embed@packet@dvipdfm}{%  
712   \ccxmp@string@len{\ccxmp@packet}%  
713   \special{pdf: object @ccxmp@packet  
714     <<  
715       /Type /Metadata  
716       /Subtype /XML  
717       /Length \the\@tempcnta  
718     >>  
719     stream~J\ccxmp@packet endstream%  
720   }%  
721   \special{pdf: docview  
722     <<  
723       /Metadata @ccxmp@packet  
724     >>  
725   }%  
726 }
```

X<sub>Y</sub>TeX creates PDFs with Xdvipdfmx, which supports a simpler `\special` for inserting objects that does not require us to count characters.

```
727 \newcommand*{\ccxmp@embed@packet@xetex}{%  
728   \special{pdf:stream @ccxmp@packet (\ccxmp@packet)  
729     <<  
730       /Type /Metadata  
731       /Subtype /XML  
732     >>
```

```

733 }%
734 \special{pdf:put @catalog
735 <<
736   /Metadata @ccxmp@packet
737 >>
738 }%
739 }

```

We rely on `hyperref` to tell us how the PDF will be generated (after all, it may not be done in the current pass) and use the respective technique to embed the XMP packet.

```

740 \AtBeginDocument{%
741   \begingroup
742   \def\ccxmp@driver{hpdfTeX}%
743   \ifx\ccxmp@driver\Hy@driver
744     \ccxmp@embed@packet@pdfTeX
745   \else
746     \def\ccxmp@driver{hdvipdfm}%
747     \ifx\ccxmp@driver\Hy@driver
748       \ccxmp@embed@packet@dviPDFM
749     \else
750       \def\ccxmp@driver{Hxetex}%
751       \ifx\ccxmp@driver\Hy@driver
752         \ccxmp@embed@packet@xetex
753       \else
754         \@ifundefined{pdfmark}{-}{%
755           \ccxmp@embed@packet@pdfmark
756         }%
757       \fi
758     \fi
759   \fi
760 \endgroup
761 }
762 %%
763 %% Copyright (C) 2015 Digital Curation Centre, University of Edinburgh
764 %% <info@dcc.ac.uk>
765 %%
766 %% End of file 'dccpaper-base.sty'.

```

## Appendix C: Installation

### Managed way

The latest stable release of the dccpaper bundle has been packaged for T<sub>E</sub>X Live and MiK<sub>T</sub>E<sub>X</sub>. If you are running T<sub>E</sub>X Live and have tlmgr installed, you can install the bundle simply by running `tlmgr install dccpaper`. If you are running MiK<sub>T</sub>E<sub>X</sub>, you can install the bundle by running `mpm --install=dccpaper`. Both tlmgr and mpm have GUI versions that you might find friendlier.

### Automated way

A makefile is provided which you can use with the Make utility:

- Running `make` generates the derived files
  - README.md
  - dccpaper.pdf
  - ijdc-v9.cls
  - idcc.cls
  - dccpaper-base.sty
  - dccpaper-apacite.bib
  - dccpaper-biblatex.bib
- Running `make inst` installs the files in the user's T<sub>E</sub>X tree.
- Running `make install` installs the files in the local T<sub>E</sub>X tree.

### Manual way

1. Compile `dccpaper.dtx` just as you would a normal L<sup>A</sup>T<sub>E</sub>X file. Bear in mind the documentation requires the biblatex package and the Biber tool rather than regular BibT<sub>E</sub>X. As well as the usual PDF (or DVI) and auxiliary files, several others are generated.
2. Move the files to your T<sub>E</sub>X tree as follows:
  - `source/latex/dccpaper`: `dccpaper.dtx`, `dccpaper.ins`
  - `tex/latex/dccpaper`: `ijdc-v9.cls`, `idcc.cls`, `dccpaper-base.sty`, `dccpaper-by.eps`, `dccpaper-by.pdf`
  - `doc/latex/dccpaper`: `dccpaper.pdf`, `dccpaper-apacite.bib`, `dccpaper-biblatex.bib`, `README.md`
3. You may then have to update your installation's file name database before T<sub>E</sub>X and friends can see the files.

## Appendix D: Sample Document

The following code demonstrates how to use dccpaper to write a IDCC conference paper.

```

1 \documentclass[research,15]{idcc}
2
3 \title{How to write a conference paper}
4 \author{Alex Ball}
5 \affil{Digital Curation Centre}
6 \author{Fred Bloggs}
7 \affil{University of Life}
8
9 \submitted{1 October 2014}
10
11 \usepackage[style=apa]{biblatex}
12 \addbibresource{references.bib}
13 \DeclareLanguageMapping{british}{british-apa}
14
15 \begin{document}
16 \maketitle
17
18 \begin{abstract}
19 Text of the abstract\dots
20 \end{abstract}
21
22 \section{Introduction}
23
24 The text of the introduction starts here\dots
25
26 \section{Conclusions}
27
28 The text of the conclusions starts here\dots
29
30 \section{Acknowledgements}
31
32 Any acknowledgements should be placed here\dots
33
34 \printbibliography
35 \end{document}

```

## Appendix E: Licence

Copyright © 2016 Digital Curation Centre, University of Edinburgh.

This work consists of the image files dccpaper-by.eps and dccpaper-by.pdf, the documented LaTeX file dccpaper.dtx and a Makefile.

The text files contained in this work may be distributed and/or modified under the conditions of the [LaTeX Project Public License \(LPPL\)](#), either version 1.3c of this license or (at your option) any later version.

The image files distributed with this bundle derive from the file [by.eps](#) distributed by Creative Commons. The image is a trademark of Creative Commons and is subject to the [Creative Commons trademark policy](#).

This work is “maintained” (as per LPPL maintenance status) by [Alex Ball](#).

The file dccpaper.pdf, generated by this work, is licensed as shown on page 1.