

GENERATING HIGH-QUALITY PDF DOCUMENTS FROM XML / HTML / CSS



State of PrintCSS in 2023

MarkupUK 2023, London

Andreas Jung • info@zopyx.com • print-css.rocks

Agenda

(this is not a PrintCSS tutorial)

- Brief introduction to PrintCSS and its standards
- The problems with comparing tools
- Free PrintCSS tools
- Commercial CSS tools
- Role of Javascript in PrintCSS
- The pain points with PrintCSS
- Decision criteria
- Q & A

About me

- Software developer & software architect
- Requirements engineering
- Python developer & Python expert

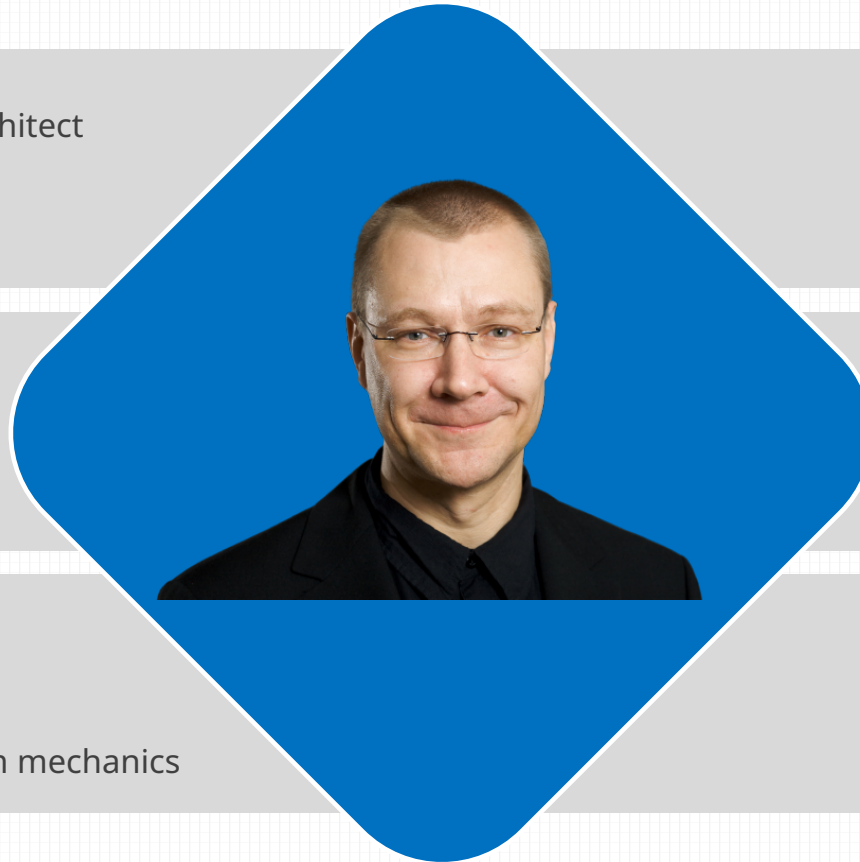
Publishing developer & consultant
since mid-90s

- CMS Solutions
- Publishing Solutions
- Individual software solutions

Funder of print-css.rocks project
Independent PrintCSS consulting

- NGOs, EDU
- medical & pharmaceutical
- energy
- research & development quantum mechanics

<https://print-css.rocks>
<https://andreas-jung.com>
✉ info@zopyx.com



Introduction

GENERATING HIGH-QUALITY PDF DOCUMENTS FROM XML / HTML / CSS



- **Input:** HTML (or XML)
- **Styling:** CSS
- **Formatting:** a free or commercial PrintCSS renderer
- **Output:** PDF

Definition of „PrintCSS“

„PrintCSS“ synonym for „W3C CSS Paged Media“

W3C CSS Paged Media Level 3 (2018)

- the official standard for generated paginated content aka PDF from HTML using cascading stylesheets (CSS)

Related standards:

- CSS Generated Content for Paged Media Module (2014, *GCPM*)

Comparing tools, metrics and their problems

How to compare tools in a reasonable way?

Problems

- we have many PrintCSS tools
- we have hundreds of CSS properties + PrintCSS specific CSS properties
- we can not test all tools against all CSS properties
(not manually, not automatically)
- W3C standards do not cover everything, leaving room for interpretations

Metrics and approach of print-css.rocks

- testing the core PrintCSS features (W3C Paged Media, GCPM)
- focus on high-level topics (e.g. footnotes, CSS grid, RTL)

How we do in print-css.rock's?

Lesson	PDFreactor	Prince	Antenna-house	Weasy-print	PagedJS	Typeset.sh	Vivliostyle	BFO
Basic	✓	✓	✓	✓	✓	✓	✓	✓
Fonts	✓	✓	✓	✓	✓	✓	✓	✓
Images	✓	✓	✓	✓	✓	✓	✓	✓
Footnotes	✓	✓	✓	✓	!	✓	✓	✓
Hyphenation	✓	✓	✓	✓	✗	✓	✗	!
Named pages	✓	✓	✓	✓	!	✓	✓	✓
Page areas	✓	✓	✓	✓	✓	✓	✓	✓
Page numbers	✓	✓	✓	✓	✓	✓	✓	✓
Pagination	✓	✓	✓	✓	✓	✓	✓	✓
Positioning	✓	✓	✓	✓	✓	✓	✓	✓

How we do in print-css.rocks?

Running elements

This example shows how to inject content pieces from the markup into a slot (e.g. header or footer). The @top-left slot is being filled with the title of the book. The @top-right slot is automatically filled with title of the subsection. Running elements allow you to move content piece into one of the sixteen page area. The running element itself is removed from the page flow.

Note: PrinceXML has different view on how to treat a *page-break-before* on the first page (which is an error in my opinion).

Further reading:

- [PrintCSS: Running headers and footers](#)

Repository files

<https://github.com/zopyx/print-css-rocks/tree/master/lessons/lesson-running-elements>

How we do in print-css.ocks?

PDF files		
Converter	Status	PDF Preview
PDFReactor 		
PrinceXML 		
Antennahouse 		
Weasyprint 		
PagedJS 		
Typeset.sh 		 no support for running elements

Standards, non-standards and extensions

Example: variations on footnotes

CSS Paged Media Tutorial by Andreas Jung

Though the body was erect, the head was thrown back so that the closed eyes were pointed towards the needle of the tell-tale that swung from a beam in the ceiling.¹

Though the body was erect, the head was thrown back so that the closed eyes were pointed towards the needle of the tell-tale that swung from a beam in the ceiling.²

Though the body was erect, the head was thrown back so that the closed eyes were pointed towards the needle of the tell-tale that swung from a beam in the ceiling.³

ⁱ The cabin-compass is called the tell-tale, because without going to the compass at the helm, the Captain, while below, can inform himself of the course of the ship.

ⁱⁱ This is another footnote

Multi-column with footnotes (column-fill: balance)

Pulvinar non pellentesque pellentesque nisi nascetur elementum platea turpis, eu, pulvinar et, a mauris, etiam vel, facilisis mus, egestas in dapibus sagittis ac mattis. Phasellus massa cum, habitasse augue sagittis, turpis penatibus auctor sagittis aenean in! Aenean porta tincidunt ultricies est lorem, facilisis dignissim tempor auctor ut sociis pulvinar eros egestas eros placerat habitasse, mid sed mattis velit platea aliquam, purus lorem, lacus et integer, natoque augue a, facilisis augue, mattis cras! A pid mattis lorem urna rhoncus. Elementum tincidunt, mus elementum turpis quis ridiculus eu odio egestas, elit auctor turpis mus platea sed, rhoncus tempor, augue dictumst phasellus tortor, ac. Massa sagittis sociis integer mid magna a turpis pellentesque, placerat lundium! Lacus, enim mus odio placerat? Nec cursus augue.

Though the body was erect, the head was thrown back so that the closed eyes were pointed towards the needle of the tell-tale that swung from a beam in the ceiling.⁴

^{vii}The cabin-compass is called the tell-tale, because without going to the compass at the helm, the Captain, while below, can inform himself of the course of the ship.

^{viii}This is another footnote

^{ix}The cabin-compass is called the tell-tale, because without going to the compass at the helm, the Captain, while below, can inform himself of the course of the ship.

Though the body was erect, the head was thrown back so that the closed eyes were pointed towards the needle of the tell-tale that swung from a beam in the ceiling.⁵

Ac non eros turpis ultrices elementum placerat augue risus in augue elementum! Cras facilisis cras, cras porttitor diam arcu, dictumst enim! Augue! Massa, tempor augue proin etiam eros odio. Magna. Diam, velit? Tortor elementum lectus ut sed odio tincidunt, odio est, turpis sociis magnis magna arcu augue? Tincidunt, risus lundium lorem, vut et cras ridiculus, tempor, dis sed ac? Ac magnis tempor pellentesque, mattis dictumst? Sit arcu montes purus tincidunt natoque? Eu scelerisque, ut augue elementum, dapibus nascetur a, amet lundium! Et, enim? Elementum, facilisis etiam ac augue etiam tincidunt a a ac magna vel? Nisi dapibus enim mattis et scelerisque sed elementum ac. Urna rhoncus? Augue! Pulvinar eros, diam. Mauris tortor amet augue nunc, quis, lacus vel ultricies.

Pulvinar non pellentesque pellentesque nisi nascetur elementum platea turpis, eu, pulvinar et, a mauris, etiam vel, facilisis mus, egestas in dapibus sagittis ac mattis. Phasellus massa cum, habitasse augue sagittis, turpis penatibus auctor sagittis aenean in! Aenean porta tincidunt ultricies est lorem, facilisis dignissim tempor auctor ut sociis pulvinar eros egestas eros placerat habitasse, mid sed mattis velit platea aliquam, purus lorem, lacus et integer, natoque augue a, facilisis augue, mattis cras! A pid mattis lorem urna rhoncus. Elementum tincidunt, mus elementum turpis quis ridiculus eu odio egestas, elit auctor turpis mus platea sed, rhoncus tempor, augue dictumst phasellus tortor, ac. Massa sagittis sociis integer mid magna a turpis pellentesque, placerat lundium! Lacus, enim mus odio placerat? Nec cursus augue.

Though the body was erect, the head was thrown back so that the closed eyes were pointed towards the needle of the tell-tale that swung from a beam in the ceiling.⁶

Multi-column with footnotes (column-fill: auto)

Pulvinar non pellentesque pellentesque nisi nascetur elementum platea turpis, eu, pulvinar et, a mauris, etiam vel, facilisis mus, egestas in dapibus sagittis ac mattis. Phasellus massa cum, habitasse augue sagittis, turpis penatibus auctor sagittis aenean in! Aenean porta tincidunt ultricies est lorem, facilisis dignissim tempor auctor ut sociis pulvinar eros egestas eros placerat habitasse, mid sed mattis velit platea aliquam, purus lorem, lacus et integer, natoque augue a, facilisis augue, mattis cras! A pid mattis lorem urna rhoncus. Elementum tincidunt, mus elementum turpis quis ridiculus eu odio egestas, elit auctor turpis mus platea sed, rhoncus tempor, augue dictumst phasellus tortor, ac. Massa sagittis sociis integer mid magna a turpis pellentesque, placerat lundium! Lacus, enim mus odio placerat? Nec cursus augue.

Though the body was erect, the head was thrown back so that the closed eyes were pointed towards the needle of the tell-tale that swung from a beam in the ceiling.⁷

ⁱThe cabin-compass is called the tell-tale, because without going to the compass at the helm, the Captain, while below, can inform himself of the course of the ship.

ⁱⁱThis is another footnote

...not even talking of

- endnotes 🐱
- sidenotes 🐱

Though the body was erect, the head was thrown back so that the closed eyes were pointed towards the needle of the tell-tale that swung from a beam in the ceiling.⁸

Ac non eros turpis ultrices elementum placerat augue risus in augue elementum! Cras facilisis cras, cras porttitor diam arcu, dictumst enim! Augue! Massa, tempor augue proin etiam eros odio. Magna. Diam, velit? Tortor elementum lectus ut sed odio tincidunt, odio est, turpis sociis magnis magna arcu augue? Tincidunt, risus lundium lorem, vut et cras ridiculus, tempor, dis sed ac? Ac magnis tempor pellentesque, mattis dictumst? Sit arcu montes purus tincidunt natoque? Eu scelerisque, ut augue elementum, dapibus nascetur a, amet lundium! Et, enim? Elementum, facilisis etiam ac augue etiam tincidunt a a ac magna vel? Nisi dapibus enim mattis et scelerisque sed elementum ac. Urna rhoncus? Augue! Pulvinar eros, diam. Mauris tortor amet augue nunc, quis, lacus vel ultricies.

ⁱThe cabin-compass is called the tell-tale, because without going to the compass at the helm, the Captain, while below, can inform himself of the course of the ship.

ⁱⁱThis is another footnote

Pulvinar non pellentesque pellentesque nisi nascetur elementum platea turpis, eu, pulvinar et, a mauris, etiam vel, facilisis mus, egestas in dapibus sagittis ac mattis. Phasellus massa cum, habitasse augue sagittis, turpis penatibus auctor sagittis aenean in! Aenean porta tincidunt ultricies est lorem, facilisis dignissim tempor auctor ut sociis pulvinar eros egestas eros placerat habitasse, mid sed mattis velit platea aliquam, purus lorem, lacus et integer, natoque augue a, facilisis augue, mattis cras! A pid mattis lorem urna rhoncus. Elementum tincidunt, mus elementum turpis quis ridiculus eu odio egestas, elit auctor turpis mus platea sed, rhoncus tempor, augue dictumst phasellus tortor, ac. Massa sagittis sociis integer mid magna a turpis pellentesque, placerat lundium! Lacus, enim mus odio placerat? Nec cursus augue.

Though the body was erect, the head was thrown back so that the closed eyes were pointed towards the needle of the tell-tale that swung from a beam in the ceiling.⁹

ⁱⁱⁱThe cabin-compass is called the tell-tale, because without going to the compass at the helm, the Captain, while below, can inform himself of the course of the ship.

^{iv}This is another footnote

no standard

Standard

Free PrintCSS renderers

Free PrintCSS renderers

(Feature selection is highly opinionated)

WeasyPrint	Vivliostyle	PagedJS
Python	Javascript/ Chromium	Javascript/ Chromium
✓ solid core ✓ Flexbox ✓ Hyphenation ✓ simple footnote support	✓ solid core ✓ Javascript support ✓ MathML ✓ simple footnote support	
✗ Grid ✗ Javascript ✗ XML	✗ Hyphentation ✗ XML	✗ many unfixed issues and bugs
good community support	good community support	community support 🧑
 actively maintained	 actively maintained	 badly maintained

Commercial PrintCSS renderers

Commercial PrintCSS renderers

(Feature selection is highly opinionated)

Typeset.sh	Prince	PDFreactor	BFO Publisher	Antennahouse Formatter
PHP		Java	Java	
✓ solid core ✓ Grid & Flexbox	✓ very mature ✓ Javascript ✓ Flexbox ✓ excellent docs/examples (https://css4.pub/) ✓ XML	✓ very mature ✓ Javascript ✓ Grid & Flexbox ✓ XML	✓ solid core	✓ very mature ✓ most advanced typography features ✓ most complete ✓ XML
• young project • has some rough edges here and there	✗ Grid		• young product • has some rough edges • not much experience	✗ Grid ✗ Flexbox ✗ Javascript
✓ excellent support	✓ excellent support	✓ excellent support		✓ excellent support
 actively maintained	 actively maintained	 actively maintained	 actively maintained	 actively maintained
89 € (annual) or 499 € (perpetual licence)	3800 USD (4 CPUs)	2.950 € (4 CPUs)	n/a	5000 USD (1st CPU)

* check pricing for educational license or „personal“ usage licenses

Role of Javascript?

Why Javascript?

Javascript is the „Franca Lingua“ of the web (besides HTML)

Javascript on the web:

- no modern web solutions without Javascript
- omnipresent

Javascript in Print/PrintCSS

- provide enhanced visualizations (graphs, charts etc) from the web in print
- *for content*: ideally Javascript visualization that work on the web, should work in PrintCSS
- *for internal usage*: influencing the rendering process of the PrintCSS renderer
- generating TOC, listings, indices

Situation of Javascript support in renderers

well...

Renderer	PagedJS	Prince	PDFreactor	Vivliostyle
Implementation	Chromium/native	custom JS engine	custom JS engine	Chromium/native
JS compatibility for PDF generation	unknown	selected JS packages	selected JS packages	best compatibility for JS modules (where they make sense)
JS for internal usage	unknown	• adaptive rendering • adaptive image scaling • table of contents	• adaptive rendering • adaptive image scaling • table of contents	unknown

The pain points with PrintCSS

The dark side of PrintCSS and its renderers

Problems and pain points

- inconsistent implementations and behavior of renderers
(due to incomplete or inconsistent Paged Media specifications)
- different defaults (page size, borders, page area dimensions etc.)
- easy things often work differently across tools
- complex topics
 - floats
 - adaptive rendering of images
 - handling of long or wide tables

Decision criteria

Is your publishing use case suitable for PrintCSS?

**Automatic typesetting
means automatic!!!**

No manual interaction!!!

Know your typographic and layout requirements

...an opinionated selection of criteria

- Is your publication text-oriented or layout-oriented?
- Do you need a grid system or flex box?
- Do you need a baseline grid?
- Support for multiple columns needed?
- Footnotes required?
- Floating objects needed?

Do I need Javascript?

- Do you have to interact with external Javascript modules (charts etc.)?
- Do you need Javascript support inside the renderer or can you prepare Javascript-based assets in a pre-processing step?

Do I have XML or HTML as input?

- only some renderers speak XML natively
- but all renderers handle HTML
- raw XML often not directly suitable unless transformed to publishable XML or HTML
- **PrintCSS is not a transformation framework**
- flat or nested HTML?
 - personal preference: nested HTML but YMMV

Do I have special PDF requirements?

- PDF metadata
- PDF variants
 - PDF/UA (accessibility)
 - PDF/X (print)
 - PDF/A (archiving)
 - *supported by most commercial renderers*
- PDF security (encryption, signing)
 - *supported by most commercial renderers*

Other features needed?

- Scientific context? (formulas, MathML, MathJax)
- integration of renderers in workflows:
 - all tools run on the command line
 - some tools provide a server mode
 - some tools provide APIs for integration with PHP, Python etc.
 - Produce & Publish server (abstraction of renderers)

Recommendations

Recommendation and hints

- be aware of your special requirements (regarding more complex features)
- build prototypes and evaluate against different renderers
- use CSS variables (not widely supported)
- consider using CSS preprocessors (e.g. Less, Sass) for complex workloads
- start with free tools, use commercial tools if needed

Q & A

Thank you for listening

Andreas Jung • info@zopyx.com • print-css.rocks