

CSS or XSL-FO: Which should I use for producing print publications?



A Data Usability Company
ANTENNA HOUSE

Michael Miller
Vice President
Antenna House, Inc.
mike@antennahouse.com
www.antennahouse.com

Xplor[®]
International

Is This A Marketing Talk?



Every Technical Detail?





Paged Output still lives!

CSS and XSL-FO

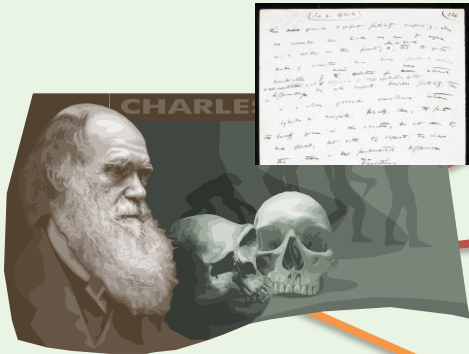
- **Cascading Style Sheets (CSS)**
 - *Used mainly for the web*, is a style sheet language to describe the presentation (that is, the look and formatting) of a document written in HTML, XHTML and XML. CSS can now also produce paged output.
 - **Extensible Stylesheet Formatting Objects (XSL-FO)**
 - *Used for print*, is a pagination markup language describing a rendering vocabulary capturing the semantics of formatting information for paginated presentation.
- Both from W3C

“Once is Enough”

- Separation of content from form
 - Integrated database – text, graphics, etc.
-

- Information is captured once
- Information is stored once

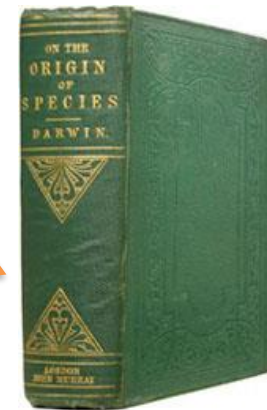
The Challenge



Charles Darwin (1809-1882)



Our focus here is the printed/PDF document



Alphabet Soup for documents and print



Markup

- CSS
- XSL-FO
- troff
- TeX
- LaTeX
- HTML
- HTML5
- XML
- SGML
- XHTML

Page Description

- AFP
- HP-GL
- PDL
- PostScript
- MODCA
- PDF
- XPS

Schemas/DTD/Data Model

- **DITA (Darwin Information Typing Architecture)** – some of the uses include software, product, and training documentation.
- **DocBook** – format for technical (but not only) manuals and documentation.
- **S1000D** – international specification for technical documentation related to commercial or military; aerospace, sea or land; vehicles or equipment.
- **JATS** (Journal Article Tag Suite) – a NISO standard of XML used to describe and publish STEM (scientific/technical/engineering/medical) scholarly journal articles.
- **Text Encoding Initiative (TEI)** – Guidelines for text encoding in the humanities, social sciences and linguistics.

Challenges

- Typesetting, page and document layout is a complex craft that draw from a wide variety of disciplines including art, technology, psychology, etc.
- It has always been difficult to produce high quality print output from XML and SGML
- Historically, this has required a significant financial investment in specialized typesetting tools, (e.g. Advent 3B2, Datalogics Pager/Composer, Miles 33, Arbortext Publisher, Xyvision XPP, TeX, etc.)
- These tools have a steep learning curve due to often cryptic, proprietary stylesheet languages or typesetting codes
- Each of the specialized typesetting tools uses its own markup

The Anatomy of Printed Publications



Front Matter

- Title page
- Table of contents
- List of figures



Body

- Chapters
- Tables
- Graphics



Back Matter

- Index
- Appendices

For Formatting You Need

- portrait and landscape pages
- recto-verso pages
- multiple page sizes
- multiple writing directions
- Unicode
- headers/footers
- floating area
- multiple columns
- generated ToCs
- generated page numbering
- graphics
- re-sorting items for output
- tables
- widow/orphan control
- hyphenation and justification
- change bars

The Driving Forces

- Controlling costs
- Multilingual documents
- More pages and more publications than ever before
- Working in a global community
- Integrating multiple sources of content
- Using information as a competitive advantage

- **Standards**
- **Technology**

How do you know you need it?

If your INFORMATION is...

- Revised between production cycles
- Re-used in different ways
- Published and delivered in different formats
- Boilerplates gets updated with customized information

If your PROCESS has...

- Multiple authors and reviewers
- Multiple departments involved
- Local and remote document creation
- Local and remote distributions
- External deadlines

Motivation: The Business Case for XML, CSS and XSL-FO

What you want:

- | | |
|---|---|
| • High quality, consistent output | ✓ "look and feel", "branding" |
| • Automated production of information products | ✓ Reduced cost, rapid dissemination |
| • Supports re-purposing of content | ✓ "single-source publishing" |
| • Supports the investment in content rather than presentation | ✓ "separation of content from formatting" |
| • Independent of any particular application | ✓ "portability" |
| • Availability of resources | ✓ "non-proprietary skills, languages" |

When to Use CSS or XSL-FO

High Need

- Technical Documentation
- User Manuals
- Training Documents
- Proposals
- Catalogues/Directories
- Parts/Price List
- Insurance Policies
- Legal Contracts
- Statements and Invoices

**Structured Content/Data to
Structured Layouts**

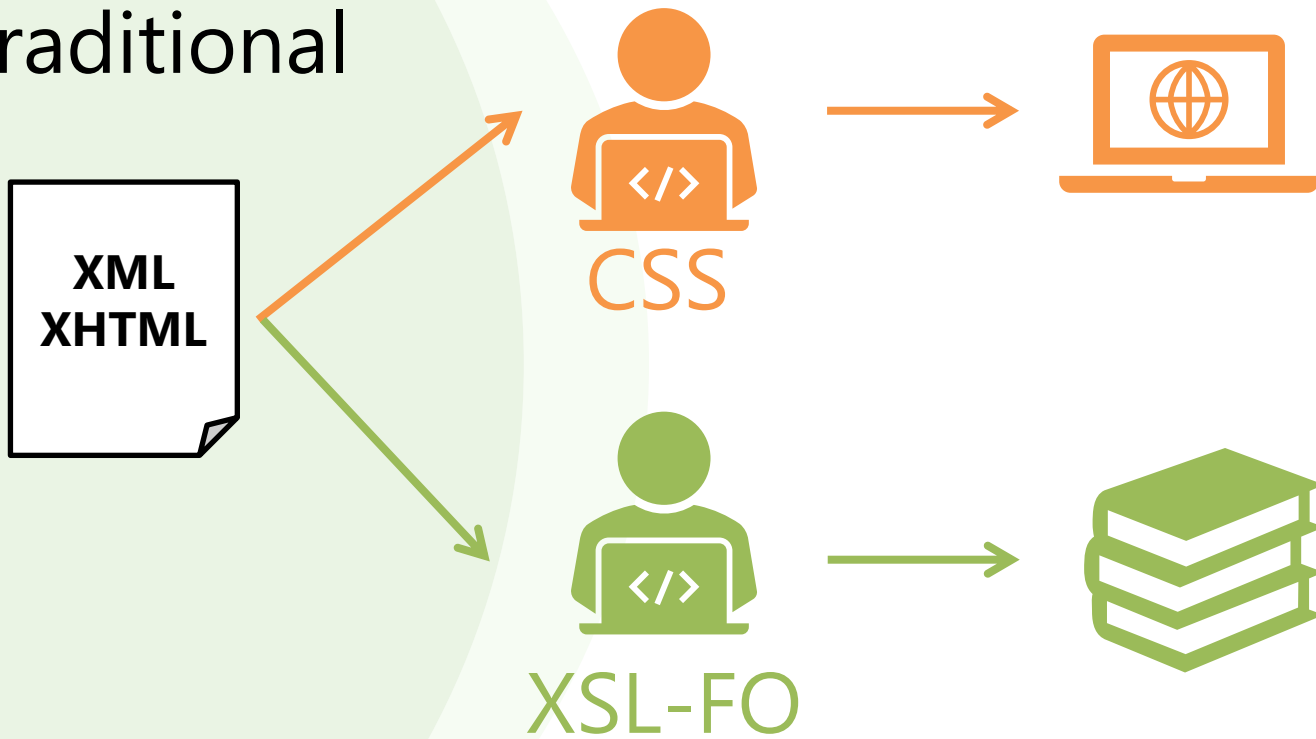
Low Need

- Presentations
- Newsletters
- Flyers
- Display Ads
- One-Time Brochures
- Changing design
- Textbooks

**Content/Data to
Unstructured Layouts**

Content to Web & Print/PDF

- Traditional



XSL-FO

(Extensible Stylesheet Formatting Objects)

This is the part of XSL that deals with how structured information should be presented in print.

XSLT

- **XSLT (Extensible Stylesheet Language Transformations)** is a language for transforming XML documents into other XML documents, or other formats such as HTML for web pages, plain text or **XSL Formatting Objects**, which may subsequently be converted to other formats, such as PDF, PostScript and PNG.

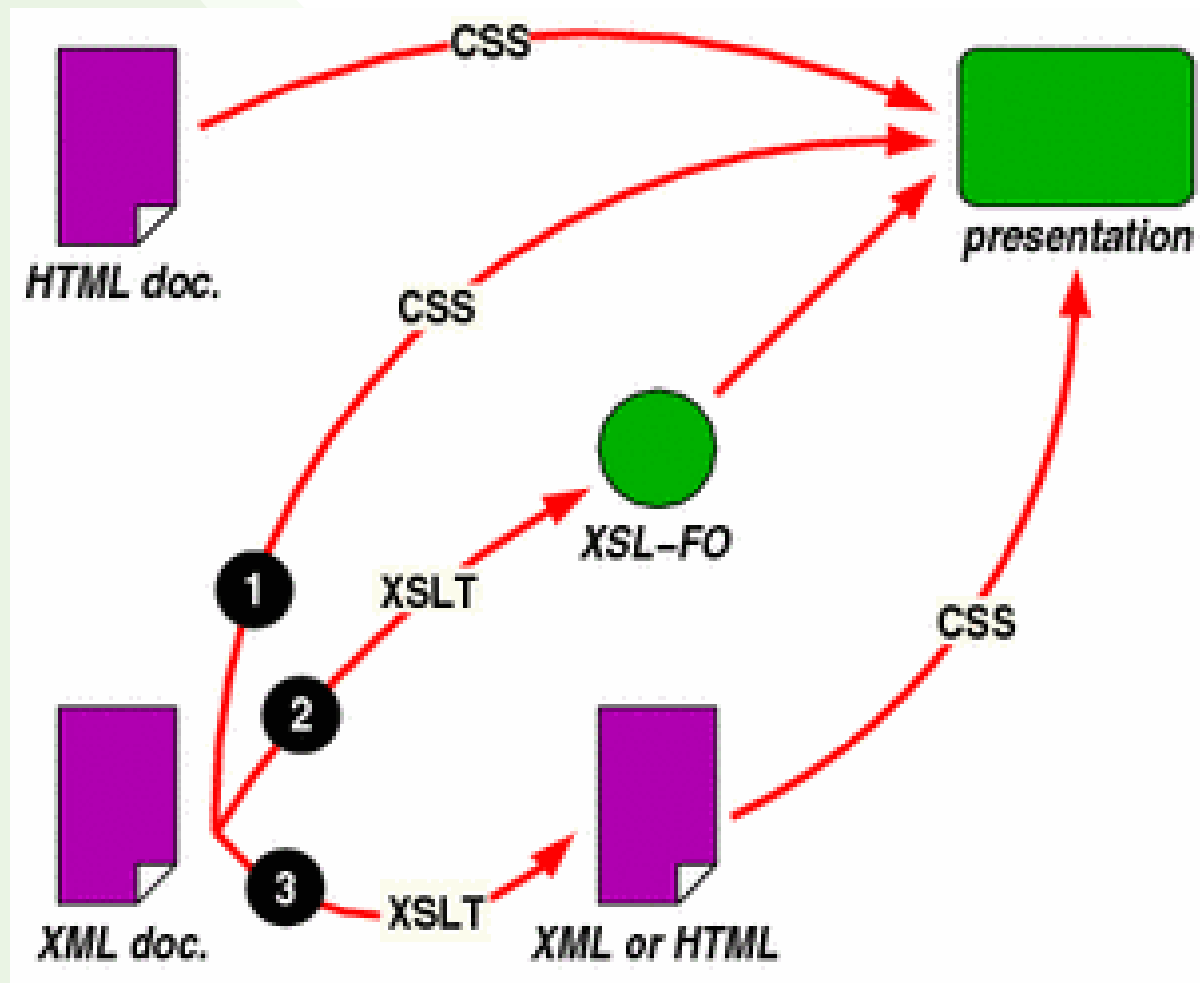
XSLT can transform XML documents into XSL-FO

- Sort lists within documents
- Replace text
- Generate back of book indexes

CSS does not use XSLT

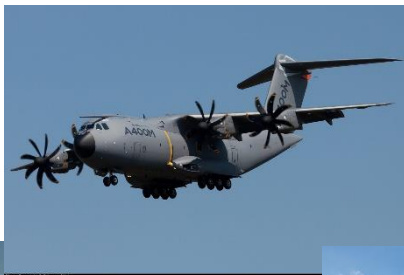
- But it can

Typical publishing data flows



<https://www.w3.org/Style/CSS-vs-XSL.en.html>

What is Documented with XSL-FO



Current Document Formatting for Print

- PDF through XSL-FO
 - Antenna House XSL Formatter
 - Apache FOP
 - RenderX XEP
- Other options
 - Framemaker
 - InDesign

From the beginning

- CSS

- Cascading Style Sheets
- W3C Recommendation
- Originally
 - Simple mechanism for adding style (fonts, color, spacing to web documents)
- Now
 - Mechanism for both web and print documents
 - CSS 3 – CSS Generated Content for Paged Media Module Working Draft May 2014

Wouldn't it be nice if you could also use CSS for print?

CSS3 adds page layout functionality for print

What is CSS3

- Web functionality +
 - CSS3 Module: Paged Media
 - CSS3 module: Generated Content for Paged Media

So Why CSS3?

CSS3 is an acknowledgement that CSS also needs to address page formatting.

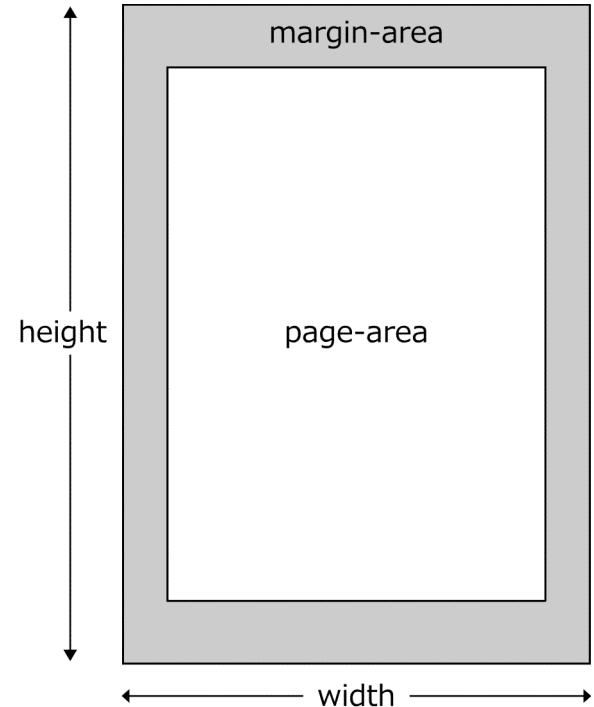
The “Cascading” Stylesheet

- CSS can have up to 3 levels of style information:
 - The default style
 - Style specified in a separate stylesheet
 - Style specified in the actual document or data

Examples of CSS3 within a stylesheet

Page setting @page rule

```
@page {  
  size: A4;  
  margin: 25mm;  
  @top-center {  
    content: "Sample";  
  }  
  @bottom-center {  
    content: counter(page);  
  }  
}
```



```
@page {  
  size: 4in 6in;  
}
```

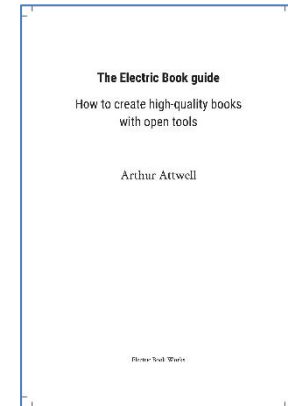
```
@page {  
  size: letter landscape;  
}
```

CSS3 Module: Paged Media adds

- Running headers and footers
- Leaders
- Cross-references
- Footnotes
- Sidenotes
- Named flows
- Hyphenation
- New counter styles
- Character substitution
- Image resolution
- Page floats
- Advanced multi-column layout
- Conditional content
- Crop and cross marks
- Bookmarks
- CMYK colors
- Continuation markers
- Change bars
- Line numbers
- Named page lists
- Generated lists
 - TOC
 - Glossary
 - Index

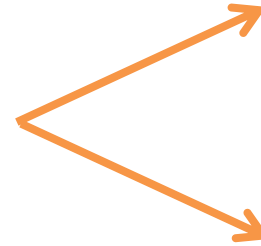
- Page Boxes: the @page Rule
 - Page Terminology and the Page Model
 - Page types
 - Page size
 - Page Selectors and the Page Context
 - Page Properties
 - Left, right, and first pages
 - Content outside the page box
- Margin Boxes
 - At-rules for margin boxes
 - Maximum Dimension of margin boxes
 - Dimension of margin boxes
 - Margin boxes and default values
 - *Margin box examples*
 - Populating margin boxes
- Page Breaks
 - Break before/after elements: 'page-break-before', 'page-break-after', 'page-break-inside'
 - Using named pages: 'page'
 - Breaks inside elements: 'orphans', 'widows'
 - Allowed page breaks
 - Forced page breaks
 - "Best" page breaks
- Page-based Extensions to Counters
 - Page-based counters
- Orienting an Image on the Page
- The 'fit' and 'fit-position' Properties

Documents Formatted with CSS



Content to Web & Print/PDF

- With CSS



CSS Options

- Antenna House CSS Formatter
- Prince XML
- PDFreactor
- Pagination.js
- Simplepagination.js

Now there is a choice!
CSS or XSL-FO

Why Two Stylesheet Languages?

- W3C originally viewed print and web as two distinct function, XSL-FO for print and CSS for web
- XSL-FO is over 20 years old
- Thousands of implementations based on XSL-FO
- A large number of available Stylesheets for XSL-FO
 - DITA, DocBook, JATS, S1000D
- A reasonable open source implementation for XSL-FO
- CSS paged media is relatively new by comparison
- The CSS products that support paged output are now becoming available
- Awareness of CSS for paged media is increasing

	CSS	XSL-FO
Can be used with HTML?	Yes	No
Can be used with XML?	Yes	Yes
Transform language?	No	Yes
Syntax	CSS	XML

Spread produced with XSL-FO

Far far away, behind the word mountains, far from the countries Vokalia and Consonantia, there live the blind texts. Separated they live in Bookmarksgrove right at the coast of the Semantics, a large language ocean. A small river named Duden flows by their place and supplies it with the necessary regelialia. It is a paradisematic country, in which roasted parts of sentences fly into your mouth. Even the all-powerful Pointing has no control about the blind texts it is an almost unorthographic life.

thead
1
2
3
4
5

6
7
8
9
10
tfoot

One day however a small line of blind text by the name of Lorem Ipsum decided to leave for the far World of Grammar. The Big Oxmox advised her not to do so, because there were thousands of bad Commas, wild Question Marks and devious Semikoli, but the Little Blind Text didn't listen. She packed her seven versalia, put her initial into the belt and made herself on the way.

壹 One
貳 Two
參 Three
肆 Four

When she reached the first hills of the Italic Mountains¹, she had a last view back on the skyline of her hometown Bookmarksgrove, the headline of Alphabet Village and the subline of her own road, the Line Lane. Pityful a rethoric question ran over her cheek, then she continued her way.

On her way she met a copy. The copy warned the Little Blind Text, that where it came from it would have been rewritten a thousand times and

everything that was left from its origin would be the word "and" and the Little Blind Text should turn around and return to its own, safe country. But nothing the copy said could convince her and so it didn't take long until a few insidious Copy Writers ambushed her, made her drunk with Longe and Parole and dragged her into their agency, where they abused her for their projects again and again. And if she hasn't been rewritten, then they are still using her.

¹ Far far away, behind the word mountains, far from the countries Vokalia and Consonantia, there live the blind texts. Separated they live in Bookmarksgrove right at the coast of the Semantics, a large language ocean. A small river named Duden flows by

their place and supplies it with the necessary regelialia. It is a paradisematic country, in which roasted parts of sentences fly into your mouth. Even the all-powerful Pointing has no control about the blind texts it is an almost unorthographic life.

Spread produced with XHTML+CSS

Far far away, behind the word mountains, far from the countries Vokalia and Consonantia, there live the blind texts. Separated they live in Bookmarksgrove right at the coast of the Semantics, a large language ocean. A small river named Duden flows by their place and supplies it with the necessary regalia. It is a paradisiac country, in which roasted parts of sentences fly into your mouth. Even the all-powerful Pointing has no control about the blind texts it is an almost unorthographic life.

thead
1
2
3
4
5

6
7
8
9
10
tfoot

One day however a small line of blind text by the name of Lorem Ipsum decided to leave for the far World of Grammar. The Big Oxmox advised her not to do so, because there were thousands of bad Commas, wild Question Marks and devious Semikoli, but the Little Blind Text didn't listen. She packed her seven versalia, put her initial into the belt and made herself on the way.

壹、One
貳、Two
参、Three

四、Four

When she reached the first hills of the Italic Mountains¹, she had a last view back on the skyline of her hometown Bookmarksgrove, the headline of Alphabet Village and the subline of her own road, the Line Lane. Pityful a rethoric question ran over her cheek, then she continued her way.

On her way she met a copy. The copy warned the Little Blind Text, that where it came from it would have been re-

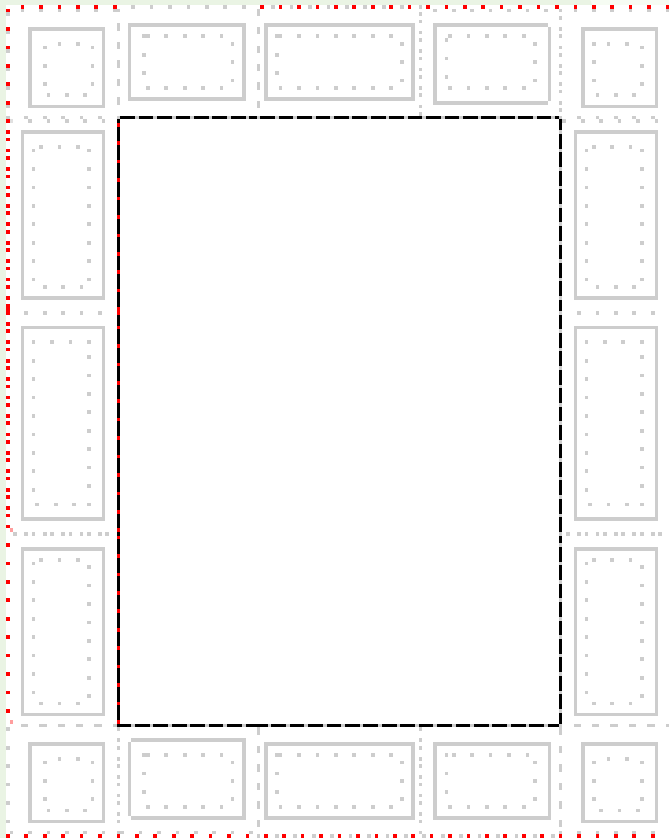
written a thousand times and everything that was left from its origin would be the word "and" and the Little Blind Text should turn around and return to its own, safe country. But nothing the copy said could convince her and so it didn't take long until a few insidious Copy Writers ambushed her, made her drunk with Longe and Parole and dragged her into their agency, where they abused her for their projects again and again. And if she hasn't been rewritten, then they are still using her.

¹ Far far away, behind the word mountains, far from the countries Vokalia and Consonantia, there live the blind texts. Separated they live in Bookmarksgrove right at the coast of the Semantics, a large language ocean. A small river named Duden flows by

their place and supplies it with the necessary regalia. It is a paradisiac country, in which roasted parts of sentences fly into your mouth. Even the all-powerful Pointing has no control about the blind texts it is an almost unorthographic life.

XSL-FO to CSS: Headers & Footers

CSS



XSL-FO



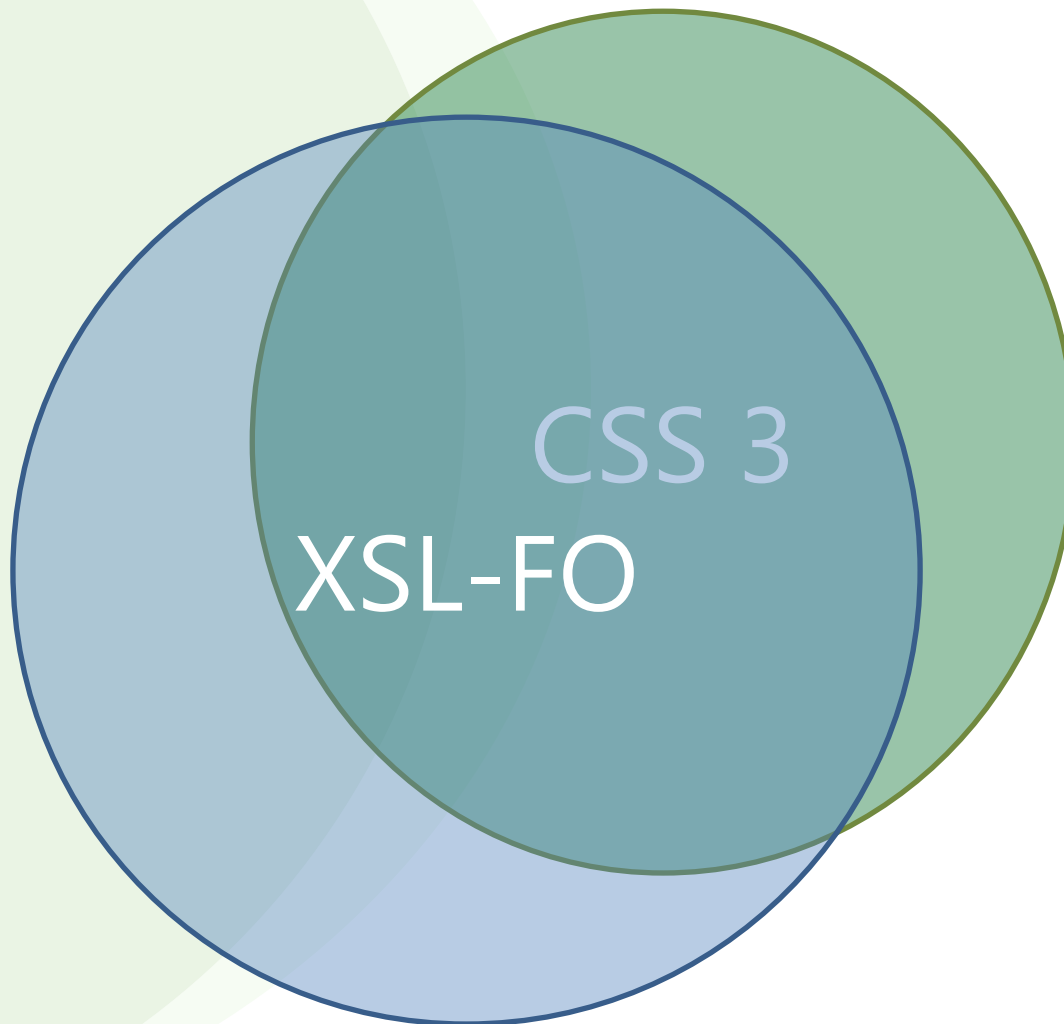
XSL-FO fo:layout-master-set

```
<fo:layout-master-set>
  <fo:simple-page-master master-name="left">
    ...
  </fo:simple-page-master>
  <fo:simple-page-master master-name="right">
    ...
  </fo:simple-page-master>
  <fo:page-sequence-master master-name="psm">
    <fo:repeatable-page-master-alternatives>
      <fo:conditional-page-master-reference
        master-reference="left"
        odd-or-even="even"/>
      <fo:conditional-page-master-reference
        master-reference="right"
        odd-or-even="odd"/>
    </fo:repeatable-page-master-alternatives>
  </fo:page-sequence-master>
</fo:layout-master-set>
```

CSS @page rule

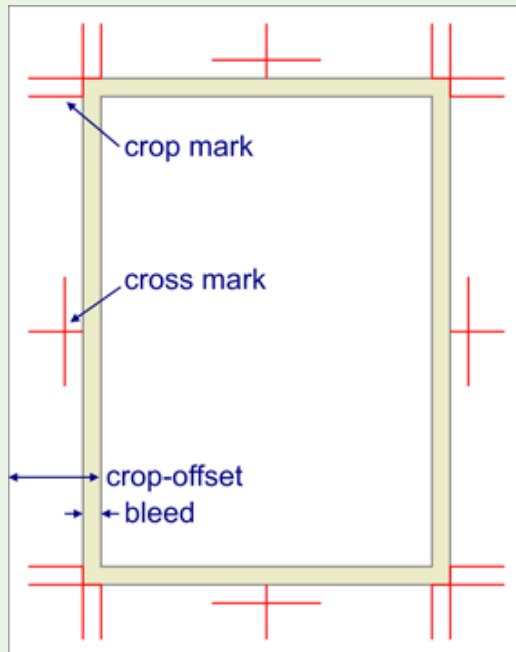
```
@page {  
  size: 120mm 140mm;  
  @footnote {  
    float: bottom page;  
    width: 100%;  
  }  
}  
  
@page :right {  
  @bottom-right {  
    content: counter(page) " of " counter(pages);  
  }  
}  
  
@page :left {  
  @bottom-left {  
    content: counter(page) " of " counter(pages);  
  }  
}
```


XSL-FO and CSS overlap in capabilities



Extensions

- Extensions add features that are not part of the core specification.



XSL-FO / CSS

axf:crop-offset / ah-crop-offset

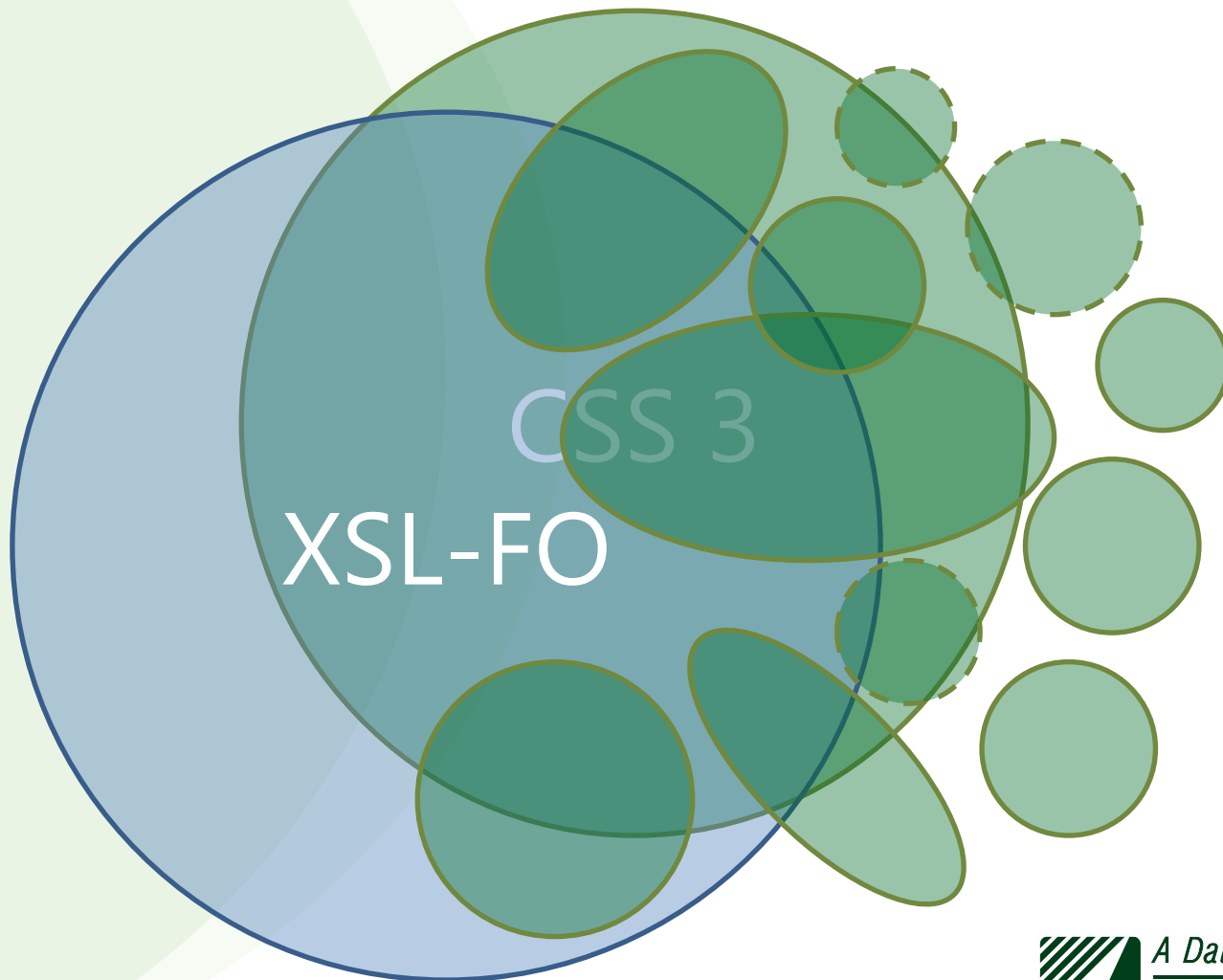
axf:crop-offset-top / ah-crop-offset-top

axf:crop-offset-bottom / ah-crop-offset-bottom

axf:crop-offset-left / ah-crop-offset-left

axf:crop-offset-right / ah-crop-offset-right

XSL-FO and CSS overlap in capabilities + Extensions



XSL-FO and CSS overlap + Extensions



Formatting
Software

XSL/CSS Properties List

The following table shows XSL-FO properties and their corresponding CSS properties. A blank cell indicates that no corresponding CSS property is implemented for the XSL-FO property, or vice-versa. See also [XSL-FO Conformance](#) or [CSS Conformance](#) to learn the current implementation status for each property. [CSS Conformance](#) also defines the CSS module abbreviations, such as [CSS3-GCPM] etc., that are used in the table. Showing properties as corresponding does not always mean that their specifications in XSL-FO and CSS are completely aligned. Some correspondences just indicate that the properties are functionally equivalent or mostly similar.

XSL	CSS	Description
axf:abbreviation-character-count	-ah- abbreviation-character-count	Specifies the minimum number of characters considered to be an
7.6.1 absolute-position	[CSS2.1] position	
axf:action-type	-ah- action-type	
axf:additive-symbols	[CSS3-CounterStyle] (-ah-) additive-symbols	✗ <axf:counter-style>
	[CSS2.1] vertical-align	
		[CSS3-Line] Setting the alignment point: the 'alignment-adjust' property
		[CSS3-Line] Aligning the alignment point of an element: the 'alignment-baseline' property
7.15.1 allowed-height-scale	-ah-allowed-height-scale	
7.15.2 allowed-width-scale	-ah-allowed-width-scale	
axf:alt-glyph	-ah- alt-glyph	Specifies the alternative glyph of a character.
axf:alt		Specifies the alternate text of the image.
axf:author		Specifies the author of the annotation. V6.2MR2 no-LT

'axf:' prefix + '-ah-' prefix
AH Formatter Extension

'axf:' prefix + CSS specification
CSS in XSL-FO

XSL section no. + '-ah-' prefix
XSL-FO in CSS

Organizations can optimize the creation and publishing of documentation to address business requirements

- Pressures
 - Shorter documentation development schedules
- Actions
 - Publish in multiple delivery formats (paper, electronic and web)
 - Provide real time customization at the point of interaction
- Capabilities
 - Single source authoring of documentation
 - Automated publication of documentation through captured logic
 - Localization
 - Graphics
- Enablers
 - Standards
 - Structured authoring applications
 - Content management systems
 - Integration of structured authoring application and content management systems

CSS3 and XSL-FO can be an important component of this!!

The benefits of a standardized format for data and presentation are overwhelming. They include:

- Complete interoperability of both content and style across applications and platforms
- Freedom of content creators from vendor control of production tools
- Freedom of users to choose their own views into content
- Easy construction of powerful tools for manipulating content on a large scale
- Multiple products and multiple vendors to choose from
- Easily integrates with other standards compliant applications
- true international publishing across all media

Standards!

For formatting you now have two standards
to choose from:

CSS and XSL-FO

Pagination for Structured Documents can be completely automated!

Benefits of CSS and XSL-FO as Part of an Implementation

- Reduce the cost of technology investment
- Enable vendors to compete and provide standards-based solutions
- Create efficiency by reducing duplicate processes
- Reduce vendor or IT support
- Eliminate data transformation between different specifications
- Overall cost savings with respect to system maintenance, upgrade, training, etc.
- Enable the use of the same authoring and publishing solution for both web and paper

CSS or XSL-FO

XSL-FO

- Now over 20 years old
- Can do complex document formatting
- Much more effort went into developing XSL-FO
- More resources available
 - Stylesheets available for DITA, DOCBook, S1000D, JATS,
 - Product – Commercial and Open Source

CSS3

- Now over 13 years old
- Can do almost all the same level of complex document formatting
- Easier to use and easier to learn = easier to maintain and cheaper
- CSS simplicity means it has its limitations

CSS or XSL-FO

Use **CSS** or **XSL-FO** when you can

- A lots of resources
- Large community of users
- Thousands of implementations
- Integrated with a large number of systems
- Available for almost all operating systems, cloud, and virtual machines



A Data Usability Company

ANTENNA HOUSE

Questions?

Michael Miller

Vice President

Antenna House, Inc.

mike@antennahouse.com

www.antennahouse.com



CSS or XSL-FO: Which should I use for producing print publications?

- Michael Miller
- Antenna House
- +1 302-427-2456
- mike@antennahouse.com
- www.antennahouse.com