Presentation of XML

Patryk Czarnik

XML and Applications 2013/2014
Week 10 – 9.12.2013

Separation of content and formatting

- According to best XML practices:
  - Documents consist of content / data.
  - Tags are for structure and meaning (semantic tagging).
    - e.g. `<amount>2.99</amount>` rather than `<i>2.99</i>`
  - There is no direct formatting information.
- How to present documents?
  - Generic (and poor) XML presentation methods
    - XML source
    - document tree
    - unformatted text content
  - Custom application handling a particular known class of documents
  - Importing XML to text editors or DTP tools
  - External style sheets

Idea of stylesheet

```
<person position="expert" id="102103">
  <fname>Dawid</fname><surname>Paszkiewicz</s
  <phone type="office">+48223213203</phone>
  <phone type="mobile">+48501502503</phone>
  <email>paszkiewicz@example.com</email>
</person>
```

```
<person position="expert" id="102105">
  <fname>Marek</fname><surname>Kącki</surname>
  <phone type="office">+48223213212</phone>
  <phone type="mobile">+48501502524</phone>
  <email>kacki@example.com</email>
</person>
```

- white background, blue frame
- font 'Times 10pt'
- 12pt for name
- abbreviation before phone number
- email in italic
- position before name

- yellow background, blue 3D frame
- font 'Bookman 12pt'
- phone numbers in single line
- email in typewriter font
- no too much details

expert
Dawid Paszkiewicz
tel. +48223213203
mob. +48501502503
paszkiewicz@examp

assistant
Marek Kącki
tel. +48223213212
mob. +48501502524
kacki@example.com

Dawid Paszkiewicz
+48223213203 +48501502503
paszkiewicz@example.com

Marek Kącki
+48223213212 +48501502524
kacki@example.com

Benefits of content and formatting separation

- With semantic tagging – source data analysis easier and more reliable (than reverse-engineering of formatted text)
- Ability to easily present
  - the same document after modifications
  - other documents from the same class
- Changes in formatting applied easily
  - modifications in one place – the stylesheet
  - whole class of documents formatted consistently
- Alternative styles for the same class of documents, depending on
  - media type (screen, printout, voice)
  - details level
  - reader preferences (or disabilities...)

Standards related to XML presentation

- Assigning style to document:
  - *Associating Style Sheets with XML documents*
- Stylesheet languages:
  - DSSSL (historical, used for SGML)
    Document Style Semantics and Specification Language
  - CSS
    Cascading Style Sheets
  - XSL
    Extensible Stylesheet Language

Associating style with document

- Using `xml-stylesheet` processing instruction
- Defined in W3C recommendation
  Associating Style Sheets with XML documents

```
<?xml version="1.0" encoding="utf-8"?>

<?xml-stylesheet
      type="text/css" href="blue.css" ?>

<?xml-stylesheet  title="Yellow"  alternate="yes"
      type="text/css"  href="yellow.css" ?>

<?xml-stylesheet  title="Pink"  alternate="yes"
      type="text/css"  href="pink.css" ?>

<person>
  <fname>Arkadiusz</fname><name>Gierasimczyk</name>
  ...
</person>
```

Cascading Style Sheets – history

- Roots of stylesheet idea – 1970s:
  - translation of markup documents to (different) printer languages
- Beginning of CSS: 1994
- CSS Level 1: December 1996
- CSS Level 2: May 1998
- CSS Level 3: work in progress for about 10 years...
  - modularisation
  - new features (browsers already support most of them)
- CSS 2.1: June 2011
  - restricts CSS 2 and makes it more precise

Applications of CSS

- First and major one: style for Web sites
- Separation of content and style for HTML
- (Simple) style sheets for XML
- CSS 2 and the idea of “accessibility”:
  - support for different media
  - support for alternative presentation means (e.g. voice generation)
  - enabling **reader** to override style proposed by author (reader rules)

Example stylesheet (fragment)

```
<company>
  <name>Extremely professional staff</name>
  <department id="acc">
    <name>Accountancy</name>
    <person position="expert" id="102103">
      <fname>Dawid</fname><surname>Paszkiewicz</surname>
      <phone type="office">+48223213203</phone>
      <phone type="mobile">+48501502503</phone>
      <email>paszkiewicz@example.com</email>
    </person>
    <person position="chief" id="102104">
      <fname>Monika</fname><surname>Domżałowicz</surname>
      <phone type="office">+48223213200</phone>
      <email>mdom@example.com</email>
    </person>
    </department>
    <main-office> ... </main-office>
</company>
```

Example stylesheet (fragment)

```
person {
  display: block;
  margin: 10px auto 10px 30px;
  padding: 0.75em 1em;
  width: 200px;
  border: solid 2px #002288;
  background-color: #FFFFFF;
}

person[position='chief'] {
  background-color: #DDFFDD;
}

fname, surname {
  display: inline;
  font-size: larger;
}

person[position='chief'] surname {
  font-weight: bold;
}
```

Resulting visualisation

Extremely professional staff

Accountancy

Dawid Paszkiewicz
tel. +48223213203
mob. +48501502503
paszkiewicz@example.com

Monika Domżałowicz
tel. +48223213200
mob. +48501502513
mdom@example.com

CSS selectors (representative examples)

- `surname` – element of the given name
- `fname, surname` – both elements
- `company name` – name being descendant of `company`
- `company > name` – name being direct child of `company`
- `surname + phone` – phone directly succeeding `surname`
- `phone:first-child` – phone being first child of its parent
- `person[position]` – `person` owing `position` attribute
- `person[position='manager']` – `person` with `position` attribute equal to `manager`
- `person [roles~='manager']` – `person` with attribute `role` containing word `manager` (attribute as space-separated list)
- `ol.staff` – equivalent to `ol[class~='staff']` (HTML only)
- `person#k12` – `person` with ID (in DTD meaning) equal to `k12`

Medium-dependent style

```
@media print {
  person {
    background-color: white;
    font-family: serif;
  }
}

@media screen {
  person {
    background-color: yellow;
    font-family: sans-serif;
  }
}

@media all {
  person {
    border-style: solid;
  }
}
```

The display property

- What kind of object is rendered for given source element
- Available values: `inline`, `block`, `list-item`, `run-in`, `inline-block`, `table`, `table-cell`, ..., `none`,
- Rarely used for HTML, since already set for HTML elements
- Primary property to be set for XML elements (`inline` by default)

```
person {
   display: block;
}

last-name {
   display: inline;
}

summary point {
   display: list-item;
}
```

Blocks and layout

- Visual blocks nested basing on nesting of source elements
  - tree of blocks and cascading inheritance of properties
- By default, (normal) blocks expanding to whole available width, laid out one under another
- Relative or absolute positioning available
- `position` (`static`, `relative`, `absolute`, `fixed`) – positioning policy
- `left`, `right`, `top`, `bottom` – position
- `width`, `height`, `min-width`, `max-height`, ... – block size
- `border-style`, `border-color`, `border-width` – borders
- `margin`, `padding` – external and internal margin

Text and font

- `color`, `background-color`, `background-image` – colour and background
- `font-family` – concrete font name or generic font family as `serif`, `sans-serif`, `monospace` . . .
- `font-size` – font size (sizes given in `px`, `pt`, `mm`, `in`, `em`, ...)
- `font-style`, `font-weight` – font variant
- `text-decoration` – underline, strike, etc.
- `text-align` – alignment of text in paragraph
  - does not apply to block in block (e.g. table in body) – use margins

Generated content

- With `:before` and `:after` pseudoclasses and `content` property
- Static text (labels etc.) not present in source document
- Attribute values
- Automatic numbering

```
person:before {
    content: attr(position);
}
phone[type='office']:before {
    content: 'tel. ';
}
phone[typ='mobile']:before {
    content: 'mob. ';
}
```

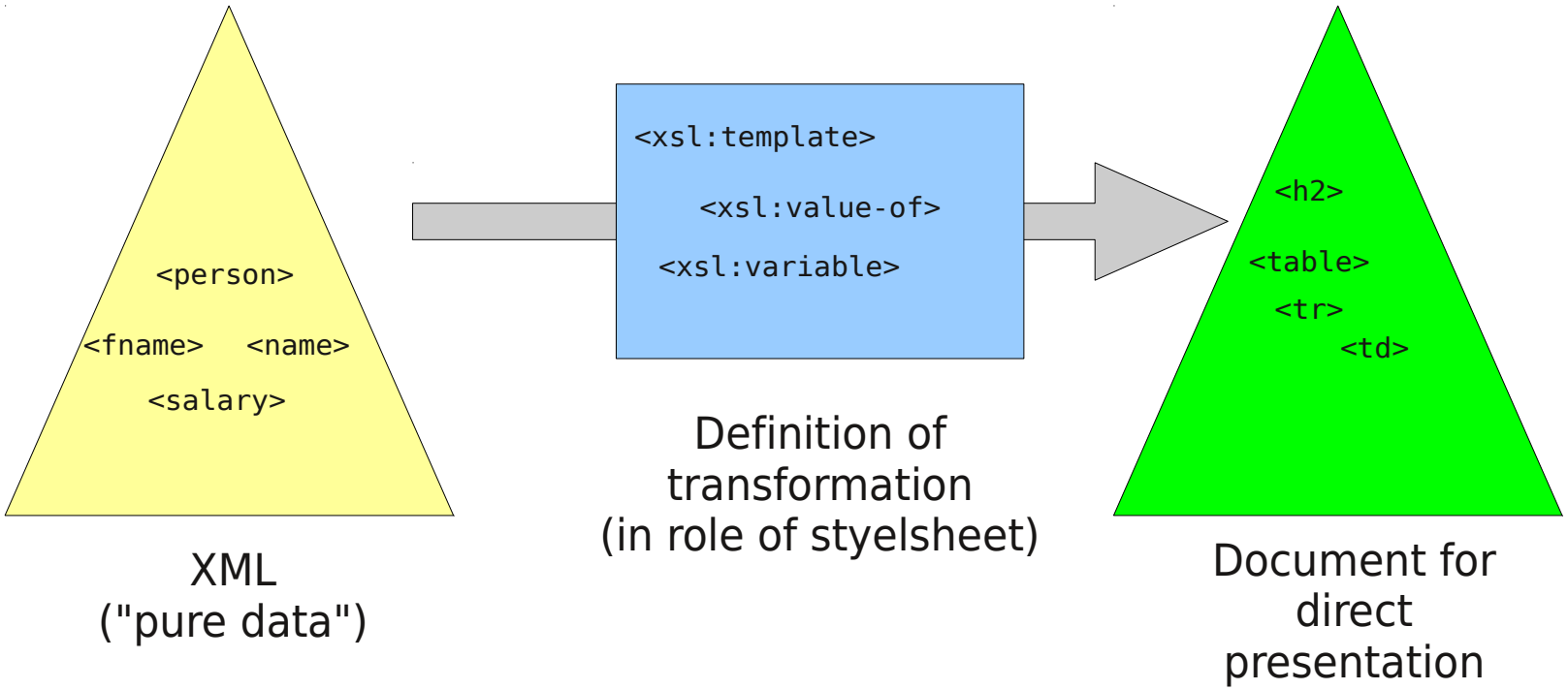
CSS capabilities and advantages

- Rich visual formatting features
- Selecting elements by
  - name
  - location in document tree
  - attribute existence
  - attribute values
- Good support
  - internet browsers
  - authoring tools
- Easy to write simple stylesheets :)

CSS shortcomings

- Only visualisation, not translation to different formats
- Selectors relatively weak. Conditions not expressible in CSS:
  - checking content of element, e.g.:
    - element `A` that contains element `B`
    - element `A` that contains text `abc`
  - logical composition of many conditions (available to some extent, but inconvenient)
  - value comparison (e.g. negative amount in red)
- Blocks structure directly based on source structure
  - reordering of elements hard (and not possible in general way)
  - not possible to show one element several times on page
- No data processing. Not available for example:
  - number calculations (summing etc.)
  - operations on text (shortening, regular expression matching, etc.)

Presentation by transformation



Extensible Stylesheet Language (XSL)

- Defined in W3C recommendations (v1.0 in 1999 and 2001):
  - **XSL** – general framework and **XSL Formatting Objects**
  - **XSLT** – language for defining XML transformations
  - **XPath** – expression language, including paths for document fragments addressing
- Original approach:
  - Transformation definition (XSLT), in the role of stylesheet, specifies how a source document is translated into FO document.
  - Presentation of result FO is specified by XSL-FO standard and available through a rendering engine.
- Practice:
  - HTML result format used more often (although FO also used)
  - XSLT and XPath also used for purposes other than presentation

Transformation to HTML – example (1)

```
<xsl:stylesheet version="1.0"
    xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
  <xsl:output method="html" encoding="utf-8" />
  <xsl:template match="/">
    <html>
      <head>
        <title>
         Employees of <xsl:value-of select="/company/name"/>
        </title>
        <style type="text/css">
          body { background-color: #FFFFDD; ... }
          div.person { margin: 10px auto 10px 30px; ... }
          ...
        </style>
      </head>
      <body>
        <xsl:apply-templates />
      </body>
    </html>
  </xsl:template>
```

Transformation to HTML – example (2)

```
<xsl:template match="person">
  <xsl:variable name="mgr">
    <xsl:if test="@position='manager'">manager</xsl:if>
  </xsl:variable>
  <div class="person {$mgr}">
    <div class="name">
      <xsl:apply-templates select="fname" />
      <xsl:text> </xsl:text>
      <xsl:apply-templates select="surname" />
    </div>
    <div class="phone">
      <xsl:apply-templates select="phone" />
    </div>
    <div class="email">
      <xsl:apply-templates select="email" />
    </div>
  </div>
</xsl:template>
```

Transformation to HTML – example (3)

```
  <xsl:template match="company/name">
    <h1>
      <xsl:apply-templates />
    </h1>
  </xsl:template>
  <xsl:template match="department/name">
    <h2>
      <xsl:apply-templates />
    </h2>
  </xsl:template>
  <xsl:template match="phone">
    <xsl:apply-templates />
    <xsl:if test="position()!=last()">
      <xsl:text> </xsl:text>
    </xsl:if>
  </xsl:template>
</xsl:stylesheet>
```

Resulting HTML code (fragments)

```
<html>
<head>
  <META http-equiv="Content-Type" content="text/html; charset=utf-8">
  <title>Employees of Extremely professional company</title>
  <style type="text/css">
    body {
    ...
  </style>
</head>
<body>
  <h1>Extremely professional company</h1>
  <h2>Accountancy</h2>
  <div class="person ">
    <div class="name">Dawid Paszkiewicz</div>
    <div class="phone">+48223213203 +48501502503</div>
    <div class="email">paszkiewicz@example.com</div>
  </div>
  <div class="person manager">
    <div class="name">Monika Domżałowicz</div>
    <div class="phone">+48223213200 +48501502513</div>
    <div class="email">mdom@example.com</div>
  </div>
...
```

HTML – resulting formatting

Extremely professional company

Accountancy

Dawid Paszkiewicz
+48223213203 +48501502503
paszkiewicz@example.com

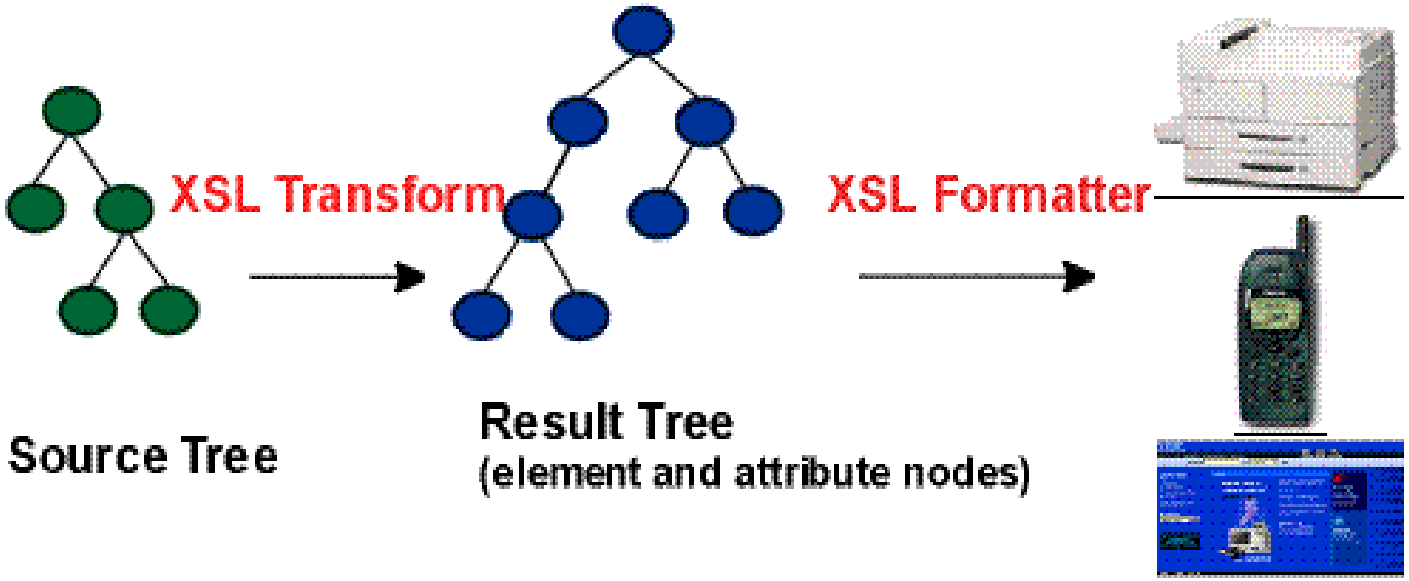
Monika Domżałowicz
+48223213200 +48501502513
mdom@example.com

Marek Kącki
+48223213212 +48501502524
kacki@example.com

HTML – basic facts

- Origin – format for software documentation
- Main application – WWW
- Nowadays recommended approach (well..., HTML 4 approach):
  - use structural (`h1`-`h6`, `p`, `ul`, `li`, `table`, ...) and semantic (`strong`, `q`, `dfn`, `blockquote`, ...) tags where appropriate
  - use general grouping elements (`div`, `span`) and custom classes when necessary
  - do not use deprecated formatting tags (`b`, `i`, `center`, ...)
  - use external CSS stylesheets when possible
  - use embedded CSS formatting rules when necessary
- XHTML ~= HTML in XML format
- (X)HTML as XSLT output format:
  Especially useful for Web and on-screen publication

Original idea of XSL



Result XML tree is the result of XSLT processing.

Source: *Extensible Stylesheet Language (XSL) Version 1.0*, W3C Recommendation 15 October 2001 (http://www.w3.org/TR/xsl/)

Transformation to XSL-FO – example (1)

```
<xsl:stylesheet version="1.0"
    xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
    xmlns:fo ="http://www.w3.org/1999/XSL/Format">
  <xsl:output method="xml" encoding="utf-8" />
  <xsl:template match="/">
    <fo:root>
      <fo:layout-master-set>
        <fo:simple-page-master  master-name="A4"
             page-width="210mm" page-height="297mm" margin="1cm">
          <fo:region-body margin="16pt 0" />
          <fo:region-before extent="16pt" />
        </fo:simple-page-master>
      </fo:layout-master-set>
      <fo:page-sequence master-reference="A4">
        <fo:static-content flow-name="xsl-region-before">
          <fo:block>
            Employees of <xsl:value-of select="/company/name"/>.
          </fo:block>
        </fo:static-content>
        <fo:flow flow-name="xsl-region-body">
          <xsl:apply-templates />
      </fo:flow></fo:page-sequence>
  </fo:root></xsl:template>
```

Transformation to XSL-FO – example (2)

```
<xsl:template match="person">
  <fo:block font-family="Verdana, sans-serif"
      space-before.minimum="12pt" padding="0.5em"
      border-width="1.5pt"        border-style="solid"
      border-color="#664400"      background-color="#FFFFCC">
    <fo:block font-size="14pt">
      <xsl:apply-templates select="fname" />
      <xsl:text> </xsl:text>
      <xsl:apply-templates select="surname" />
    </fo:block>
    <fo:block margin-top="0.5em">
      <xsl:apply-templates select="phone" />
    </fo:block>
    <fo:block margin-top="0.5em">
      <xsl:apply-templates select="email" />
    </fo:block>
  </fo:block>
</xsl:template>
<xsl:template match="person[@position='manager']/surname">
  <fo:inline font-weight="bold">
    <xsl:apply-templates />
  </fo:inline>
</xsl:template>
```

Transformation to XSL-FO – example (3)

```
  <xsl:template match="phone">
    <fo:block>
      <xsl:choose>
        <xsl:when test="@type='mobile'">mob. </xsl:when>
        <xsl:otherwise>tel. </xsl:otherwise>
      </xsl:choose>
      <xsl:apply-templates />
    </fo:block>
  </xsl:template>
  <xsl:template match="email">
    <fo:block font-style="italic">
      <xsl:apply-templates />
    </fo:block>
  </xsl:template>
  <xsl:template match="department/name">
    <fo:block font-size="16pt" font-weight="bold" font-style="italic"
              text-align="left" margin-bottom="6pt">
      <xsl:apply-templates />
    </fo:block>
  </xsl:template>
<!-- Some more templates... -->
```

Resulting XSL-FO code (fragments)

```
<?xml version="1.0" encoding="utf-8"?>
<fo:root xmlns:fo="http://www.w3.org/1999/XSL/Format">
  <fo:layout-master-set>
    <fo:simple-page-master master-name="A4" page-width="210mm" ...>
  ...</fo:layout-master-set>
  <fo:page-sequence master-reference="A4">
    <fo:static-content flow-name="xsl-region-before">
      <fo:block>Employees of Extremely professional
              company.</fo:block>
    </fo:static-content>
    <fo:flow flow-name="xsl-region-body">
      <fo:block ...> ...
        <fo:block space-before.minimum="12pt" padding="0.5em" ...>
          <fo:block font-size="14pt"> Monika
            <fo:inline font-weight="bold">Domżałowicz</fo:inline>
          </fo:block>
          <fo:block margin-top="0.5em">
            <fo:block>tel. +48223213200</fo:block>
            <fo:block>mob. +48501502513</fo:block>
          </fo:block>
          <fo:block margin-top="0.5em">
            <fo:block font-style="italic">mdom@example.com</fo:block>
        </fo:block></fo:block>...
```

XSL-FO – resulting formatting

Employees of Extremely professional company.

Extremely professional company

Accountancy

Dawid Paszkiewicz

tel. +48223213203
mob. +48501502503

paszkiewicz@example.com

Monika **Domżałowicz**

tel. +48223213200
mob. +48501502513

mdom@example.com

XSL-FO – basic facts

- Presentation-oriented XML application
- Elements for different kinds of visual objects (`block`, `inline`, `table`, and so on)
- Attributes for formatting, based on CSS properties
- Especially useful for printed publications
- Focused on paged media type:
  - master pages (templates), page areas (header, footer, etc.)
  - automatic text flow and repeated (“static”) content
  - Practice: intermediate format in
    XML → XSL-FO → PDF transformation
  - Not supported by web browsers
- Designed as part of XSL framework
  - result of XSLT transformation
  - not intended to be used standalone

Basic structure of XSL-FO document

- Two main parts: declarations and actual content

```
<fo:root xmlns:fo="http://www.w3.org/1999/XSL/Format">

  <fo:layout-master-set>
    <fo:simple-page-master master-name="my-page">
      <fo:region-body />
    </fo:simple-page-master>
  </fo:layout-master-set>

  <fo:page-sequence master-reference="my-page">
    <fo:flow flow-name="xsl-region-body">
      <fo:block>Hello World!</fo:block>
    </fo:flow>
  </fo:page-sequence>

</fo:root>
```

Formatting objects

XSL-FO elements relate to resulting formatting objects.

- Block level
  - `block`
  - `list-block`, `list-item`, `list-item-label`
  - `table`, `table-row`, `table-cell`, ...
- Inline level
  - `inline`, `character`
  - `external-graphics`
- Special features
  - `basic-link`, `bookmark`, `marker`
  - `footnote`
  - `page-number`

List – example

```
<fo:list-block>
  <fo:list-item>
    <fo:list-item-label>
      <fo:block>First name: </fo:block>
    </fo:list-item-label>
    <fo:list-item-body>
      <fo:block margin-left="15em">Dawid</fo:block>
    </fo:list-item-body>
  </fo:list-item>
  <fo:list-item>
    <fo:list-item-label>
      <fo:block>Surname: </fo:block>
    </fo:list-item-label>
    <fo:list-item-body>
      <fo:block margin-left="15em">Paszkiewicz</fo:block>
    </fo:list-item-body>
  </fo:list-item>
</fo:list-block>
```

Table – example

```
<fo:table border="solid 2pt black">
 <fo:table-header>
  <fo:table-row>
    <fo:table-cell><fo:block font-weight="bold">Surname
                            </fo:block></fo:table-cell>
    <fo:table-cell><fo:block font-weight="bold">First name
                            </fo:block></fo:table-cell>
    ...
  </fo:table-row>
 </fo:table-header>
 <fo:table-body>
   <fo:table-row>
     <fo:table-cell><fo:block>Paszkiewicz</fo:block></fo:table-cell>
     <fo:table-cell><fo:block>Dawid</fo:block></fo:table-cell>
    ...
   </fo:table-row>
```

Formatting properties

Most of XSL-FO attributes relate to style properties analogue to CSS properties.

- `margin`, `padding`, `border-style`
- `background-color`, `background-image`
- `font-family`, `font-weight`, `font-style`, `font-size`
- `text-align`, `text-align-last`, `text-indent`, `start-indent`, `end-indent`, `wrap-option`,
- `break-before`
- and much more (almost 300 properties in XSL 1.1)

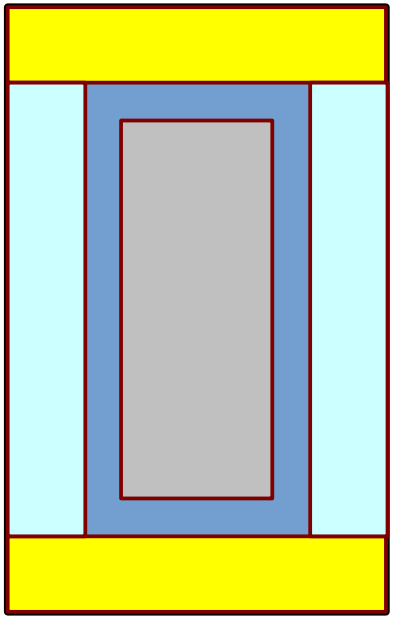
“Page master” – page template

- Single page layout
- A document may be split in many such pages
- One or more body regions
- Four predefined (but optional to use) edge regions

```
<fo:simple-page-master master-name="A4"
  page-width="297mm"   page-height="210mm"
  margin-top="1cm"     margin-bottom="1cm"
  margin-left="1cm"    margin-right="1cm">

    <fo:region-body   margin="3cm"/>
    <fo:region-before extent="2cm"/>
    <fo:region-after  extent="2cm"/>
    <fo:region-start  extent="2cm"/>
    <fo:region-end    extent="2cm"/>

</fo:simple-page-master>
```



Distributing content to pages

- `page-sequence` – results in a number of pages
- `flow` – content split into pages
- `static-content` – content repeated on all pages
- `flow-name` – page region reference

```
<fo:page-sequence master-reference="A4">
  <fo:static-content flow-name="xsl-region-before">
    Employees of <xsl:value-of select="company/name" />
  </fo:static-content>

  <fo:flow flow-name="xsl-region-body">
    <xsl:apply-templates />
  </fo:flow>
</fo:page-sequence>
```

Page sequence master

- Using different page layouts within one page-sequence
- Simple page masters referred to be used in order (repetitions available)

```
<fo:layout-master-set>
  <fo:simple-page-master master-name="first">...</fo:simple...>
  ...
  <fo:page-sequence-master master-name="seq_master">
    <fo:single-page-master-reference master-reference="first"/>
    <fo:repeatable-page-master-reference
       master-reference="starting" maximum-repeats="3"/>
    <fo:repeatable-page-master-reference
       master-reference="default"/>
  </fo:page-sequence-master>
</fo:layout-master-set>

  <fo:page-sequence master-reference="seq_master">
    <fo:flow flow-name="xsl-region-body">...
```

Alternative page references

- Choosing page depending on conditions
  - odd/even, blank?, first?

```
<fo:page-sequence-master master-name="rich_master">
  <fo:repeatable-page-master-alternatives>
    <fo:conditional-page-master-reference
       master-reference="first" page-position="first"/>
    <fo:conditional-page-master-reference
       master-reference="right" odd-or-even="odd"/>
    <fo:conditional-page-master-reference
       master-reference="left" odd-or-even="even"/>
  </fo:repeatable-page-master-alternatives>
</fo:page-sequence-master>
```

Multiple flows

- Added in XSL 1.1, not available in Apache FOP

```
<fo:layout-master-set>
  <fo:simple-page-master master-name="two-bodies"
     page-width="210mm" page-height="297mm" margin="10mm">
    <fo:region-body margin="27mm 0 100mm 0" region-name="top"/>
    <fo:region-body margin="177mm 0 0 0" region-name="down"/>
    <fo:region-before extent="27mm" />
  </fo:simple-page-master>
</fo:layout-master-set>

<fo:page-sequence master-reference="two-bodies">
  <fo:static-content flow-name="xsl-region-before">
    <fo:block>Header</fo:block>
  </fo:static-content>
  <fo:flow flow-name="top">
    <xsl:apply-templates select="main-content"/>
  </fo:flow>
  <fo:flow flow-name="down">
    <xsl:apply-templates select="bottom-content"/>
  </fo:flow>
</fo:page-sequence>
```

Custom flow maps – by example

- Here: two text flows merged together and placed into two page regions

```
<fo:flow-map flow-map-name="E4">
  <fo:flow-assignment>
    <fo:flow-source-list>
      <fo:flow-name-specifier flow-name-reference="A"/>
      <fo:flow-name-specifier flow-name-reference="B"/>
    </fo:flow-source-list>
    <fo:flow-target-list>
      <fo:region-name-specifier
            region-name-reference="R1"/>
      <fo:region-name-specifier
            region-name-reference="R2"/>
    </fo:flow-target-list>
  </fo:flow-assignment>
</fo:flow-map>
```

Markers

- Typical application – running header/footer

In flow (main content)

```
<xsl:template match="section">
  <fo:block ...>
    <fo:marker marker-class-name="sec-name">
      <xsl:value-of select="title"/>
    </fo:marker>
...
```

In static content (header)

```
<fo:block>
  Page <fo:page-number />
  Section <fo:retrieve-marker retrieve-class-name="sec-name"/>
</fo:block>
```

XSL-FO – discussion

- Main XSL-FO advantages
  - “in line” with XSL
  - easy and direct way to obtain printout (e.g. PDF) from XML data
  - general advantages of stylesheets over “hard-coded” formatting
  - automatised process when compared to manual formatting
- Main XSL-FO disadvantages
  - too complex for simple needs (see e.g. lists)
  - too limited for advanced needs
    - lack of pagination feedback, cannot say “if these two elements occur on the same page then...”
    - hard to format particular elements in a very special way (thus this is a general drawback of stylesheets comparing to manual formatting)