

Modélisation et interopérabilité :

Semaine 40, cours 3

Benoît Valiron <benoit.valiron@monoidal.net>

<http://inf356.monoidal.net/>

Question pratique : Evaluation

- Trois choses :
 - Un examen final (coef 2/3)
 - Un examen partiel (coef 1/6)
 - Mardi 10 Novembre, de 12h à 15h
 - Salle à déterminer
 - Contenu : format XML, DTD, Manipulation de formats, Relax NG.
 - Documents autorisés (mais pas de livres...)
 - Un projet (coef 1/6)
 - À rendre la dernière semaine de cours
 - Support numérique : vous assurer que je peux le lire !

Le projet

- Sujet en ligne ce week-end
- Il vous sera demandé :
 - Création d'une base de donnée au format XML
 - Horaires des TER en Rhône-Alpes
 - Fichiers de validation DTD et Relax NG
 - Documentation du format que vous avez développé :
 - Inclus dans le Relax NG
 - En format docbook
 - Des fonctions XPath pour récupérer des infos
 - Des transformations XSLT pour un affichage de résultats
 - Un affichage graphique (SVG et google-map)

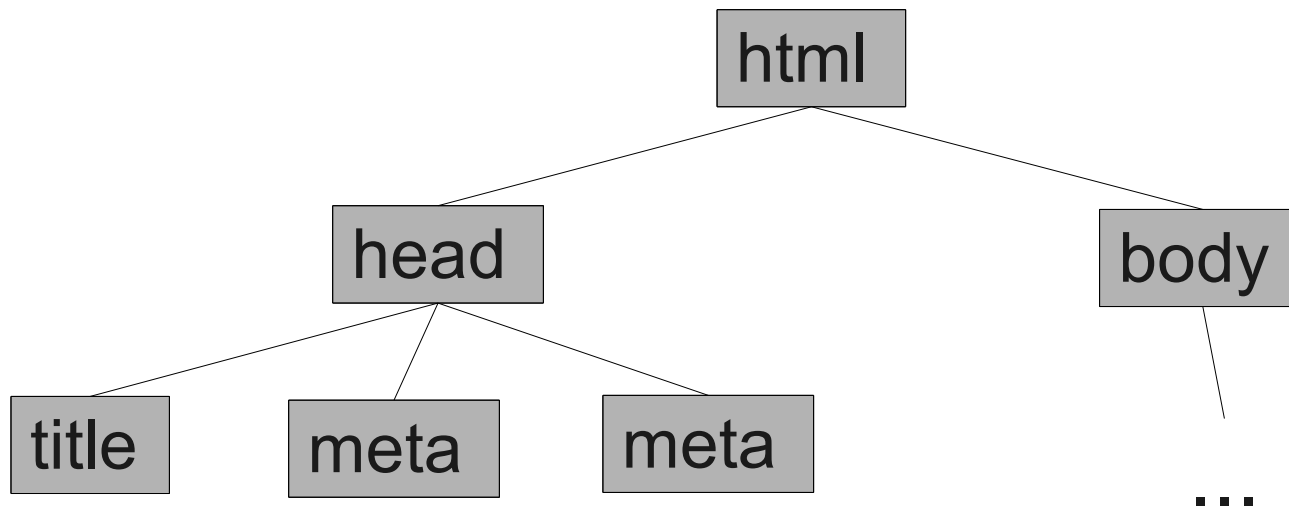
Aujourd'hui, petit tour du
propriétaire.

XML : où le trouver ?

- Pages internet : XHTML
- Images : SVG
- Maths : MathML
- Données textuelles
 - Données GPS : openstreet, GPX, google KML...
 - Flux RSS
- Méta-données
 - RDF et Dublin Core
- Traitements de texte :
 - ODF : OpenOffice
 - OOXML : MS Word
- Protocole d'échange de données entre applications : SOAP
- ...

XHTML

- Application XML lisible par un navigateur.
- Nom qualifié : <http://www.w3.org/1999/xhtml>
- Format (simplifié) :



```
<!ELEMENT html (head, body)>  
<!ELEMENT head (meta*, title, meta*)>  
<!ELEMENT title (#PCDATA)>  
<!ELEMENT meta EMPTY>
```

3 DTDs pour XHTML

- **XHTML Strict :**

```
<!DOCTYPE html PUBLIC
    "-//W3C//DTD XHTML 1.0 Strict//EN"
    "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
```

- **XHTML Transitional :**

```
<!DOCTYPE html PUBLIC
    "-//W3C//DTD XHTML 1.0 Transitional//EN"
    "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd"
```

- **XHTML Frameset :**

```
<!DOCTYPE html PUBLIC
    "-//W3C//DTD XHTML 1.0 Frameset//EN"
    "http://www.w3.org/TR/xhtml1/DTD/xhtml1-frameset.dtd"
```

http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd

... Entités paramètres ...

```
<!--===== Document Structure =====-->
<!ELEMENT html (head, body)>
<!ATTLIST html
  %i18n;
  id          ID          #IMPLIED
  xmlns      %URI;       #FIXED 'http://www.w3.org/1999/xhtml1'
  >

<!--===== Document Head =====-->
<!ENTITY % head.misc "(script|style|meta|link|object)*">
<!ELEMENT head (%head.misc;,
  ((title, %head.misc;, (base, %head.misc;)? ) |
  (base, %head.misc;, (title, %head.misc;))))>
<!ATTLIST head
  %i18n;
  id          ID          #IMPLIED
  profile     %URI;       #IMPLIED
  >

<!ELEMENT title (#PCDATA)>
<!ATTLIST title
  %i18n;
  id          ID          #IMPLIED
  >
```

Le fichier fait 978 lignes de code, dont 200 lignes de commentaires.

- 47 ELEMENT
- 77 ATTLIST
- 54 ENTITY

.... Plus de règles ...

Élément body

- Texte structuré :
 - Titres : h1 h2 h3 h4 h5
 - Paragraphes : p
 - Sauts de ligne : br
 - Style de texte : b i tt
 - Tableaux : table tr td
 - Images : img
 - Listes : ul li
 - ...

Un document XHTML

- Extension : .html ou .xhtml
- Appel à l'espace de nom :

```
<?xml version="1.0"?>
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <title>Titre du document</title>
  </head>
  <body>
    <h1>Titre 1</h1>
    <p>
      Paragraphe
      <br />
      Nouvelle ligne
    </p>
    <h2>Sous titre 1</h2>
    <h3>Sous-sous titre 1</h3>
    <ul>
      <li>élément d'une liste</li>
      <li>autre élément</li>
    </ul>
    <h2>Sous titre 2</h2>
    
  </body>
</html>
```

Visualisation

```
benoit@ben-laptop: /tmp
Titre 1

Paragraphe
Nouvelle ligne

Sous titre 1

Sous-sous titre 1

* élément d'une liste
* autre élément

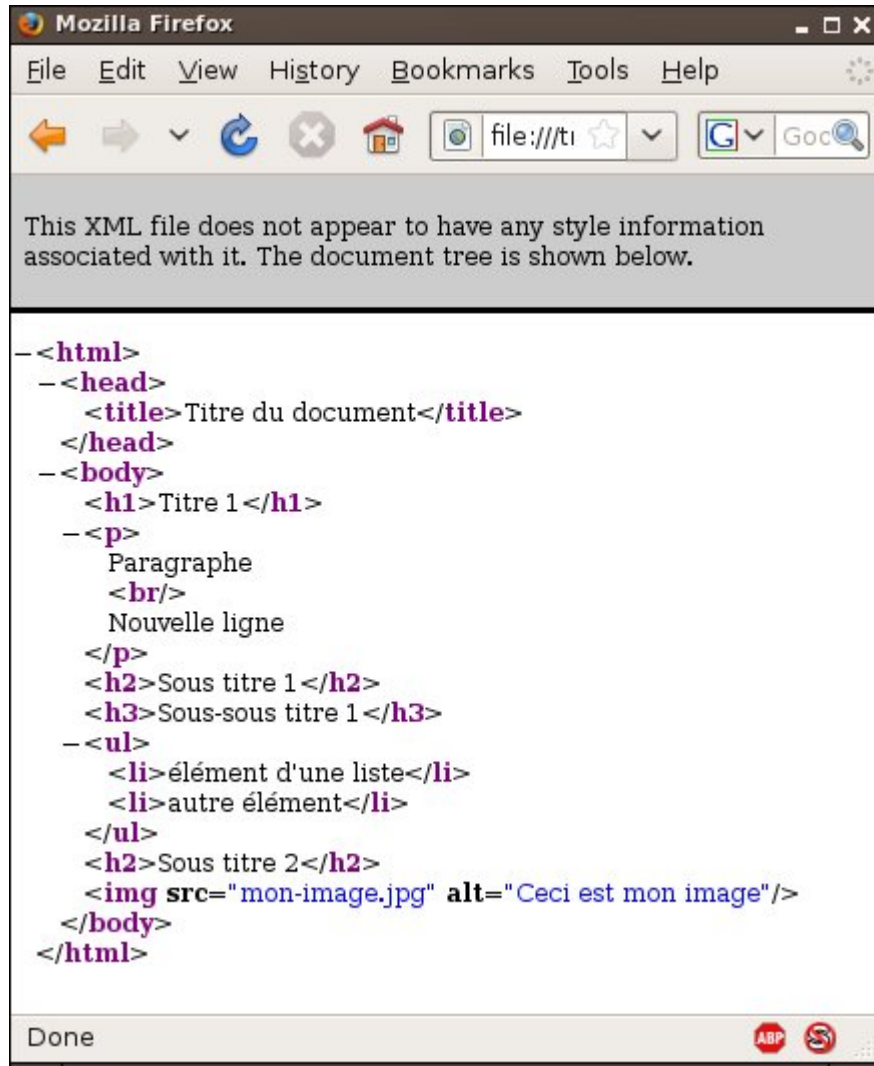
Sous titre 2

Ceci est mon image

Commands: Use arrow keys to move, '?' for help, 'q' to quit, '
Arrow keys: Up and Down to move. Right to follow a link; Le
H)elp O)ptions P)rint G)o M)ain screen Q)uit /=search [delete
```



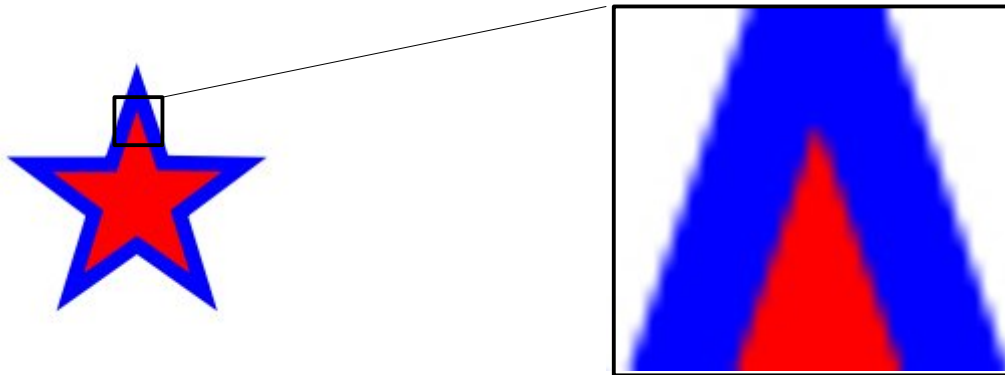
Sans espace de nom...



- Firefox est perdu.
- Lynx est plus malin et infère le format du fichier.

Format d'image

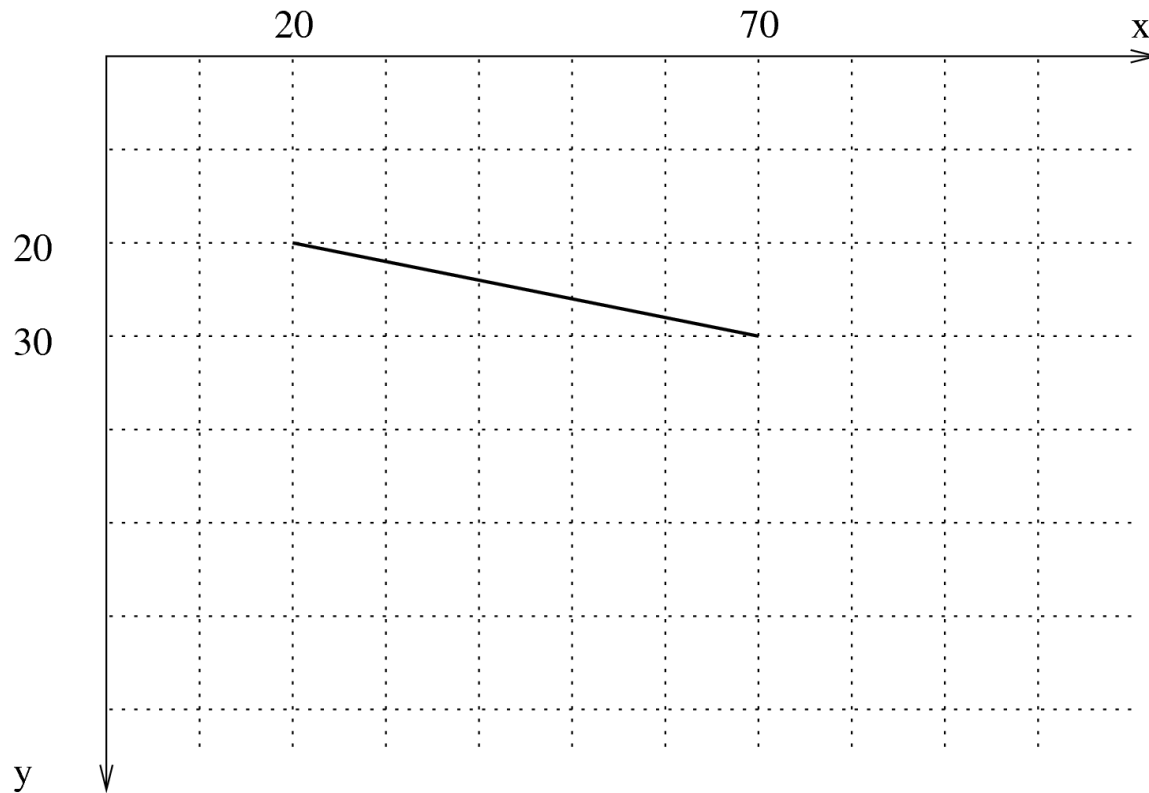
- Les formats JPG, GIF, PNG... sont des formats binaires : description de l'image pixel par pixel.
- Ne supportent pas bien le zoom



- Vectoriel : image décrite par des points et des zones du plan
- Meilleure résistance au zoom ; plus modulaire.

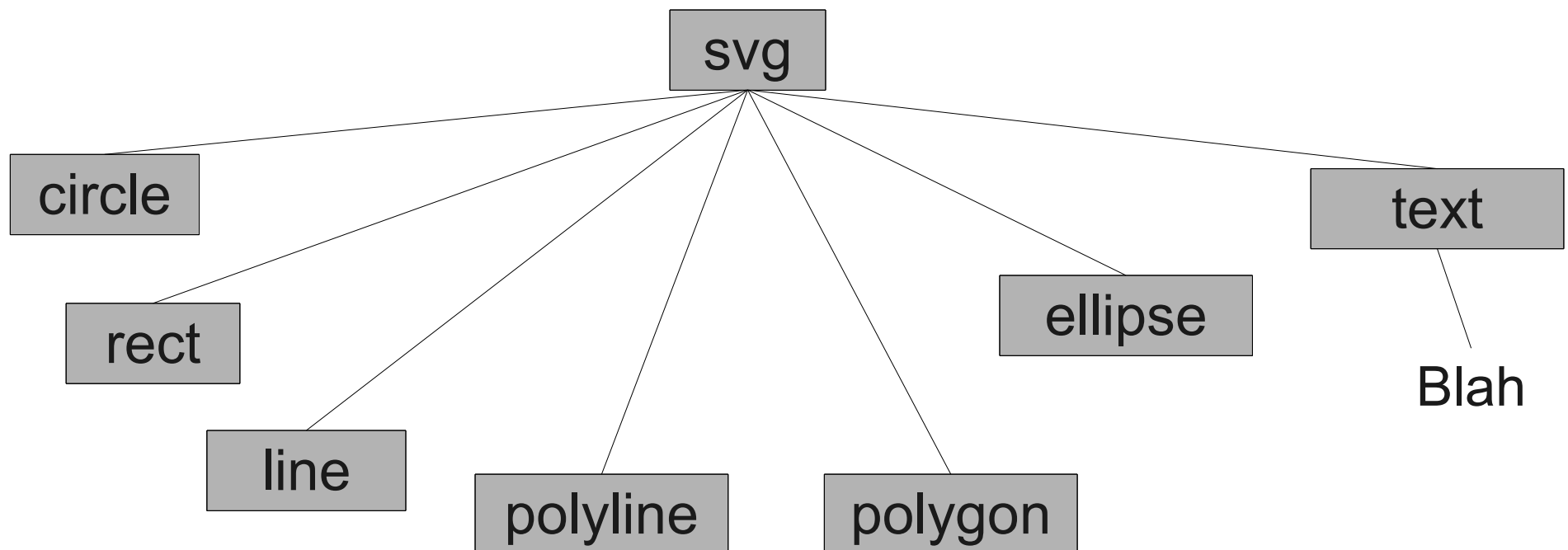
Image vectorielle

- Le trait va être caractérisé par son épaisseur et ses extrémités $(20, 20)$ et $(70, 30)$.



Format SVG

- Application XML.
- Nom qualifié : <http://www.w3.org/2000/svg>
- Exemple d'arbre SVG (sans attribut) :



Attributs SVG

- Pour les formes et le texte:
 - Positions et longueurs :

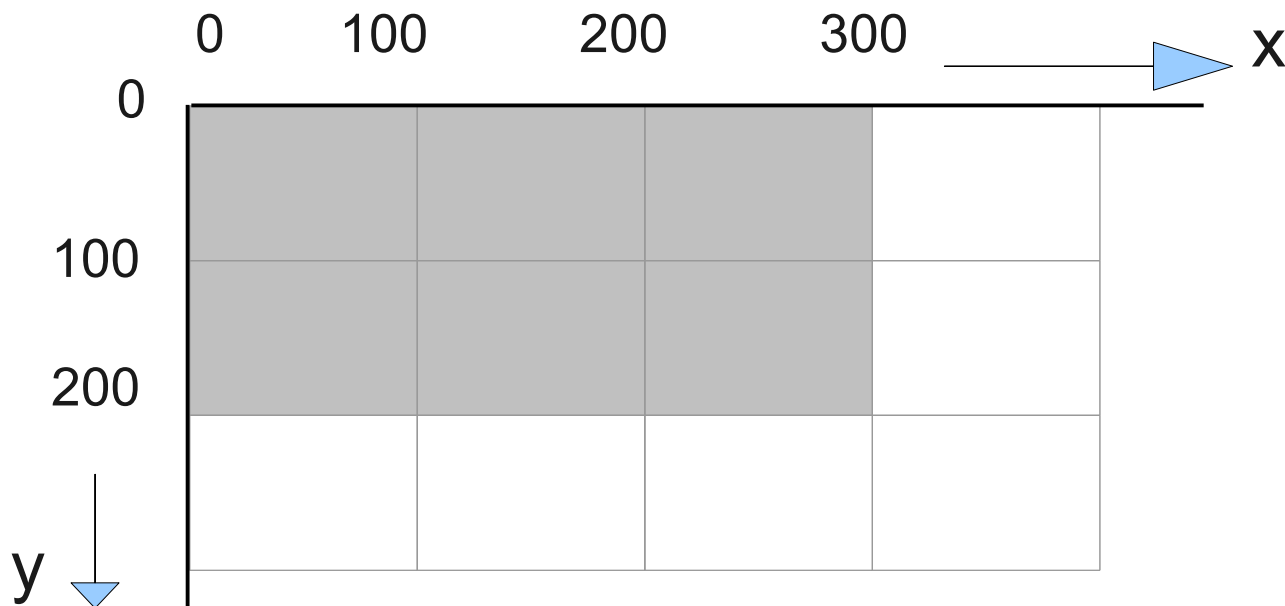
```
<circle cx="70" cy="100" r="50" />  
<rect x="150" y="50" width="135" height="100" />  
<line x1="325" y1="150" x2="375" y2="50" />  
<polyline points="50,250 75,350 100,250 125 350" />  
<polygon points=" 250,250 297,284 279,340" />  
<ellipse cx="400" cy="300" rx="72" ry="50" />  
<texte x="100" y="100">Hello</text>
```

- Trait et couleurs :
 - stroke : couleur du trait
 - fill : couleur de l'intérieur de la forme
 - stroke-width : épaisseur du trait

Attributs SVG (suite)

- Pour la racine :
 - `viewBox` : la plage du plan (ex : “x1 y1 x2 y2”)
 - `width` : la largeur désirée (ex : “2cm”)
 - `height` : la hauteur désirée (ex : “4cm”)

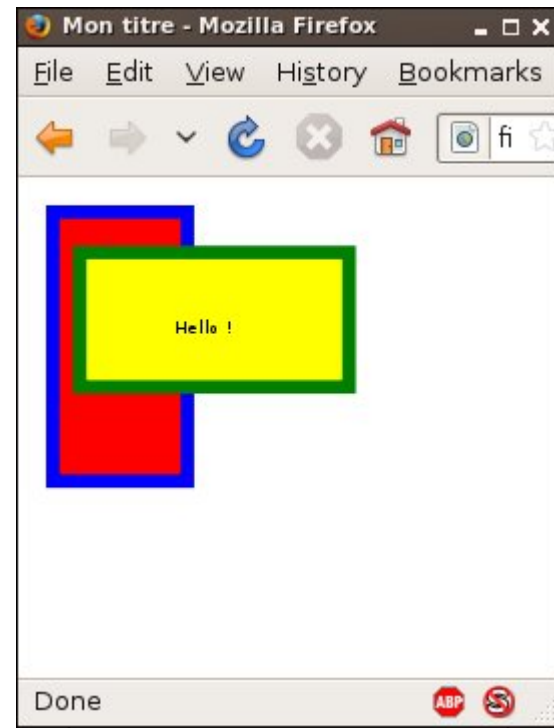
```
<svg viewBox="0 0 200 300" width="4cm"> ... </svg>
```



Exemple : rectangles

- Dans un fichier, extension .xml ou .svg
- Avec espace de nom

```
<svg viewBox="0 0 300 300"  
  width="4cm"  
  xmlns="http://www.w3.org/2000/svg">  
  <rect  
    fill="red"  
    stroke="blue"  
    stroke-width="10"  
    x="10" y="10"  
    width="100"  
    height="200"/>  
  <rect  
    fill="yellow"  
    stroke="green"  
    stroke-width="10"  
    x="30" y="40"  
    width="200"  
    height="100"/>  
  <text x="100" y="100">Hello!</text>  
</svg>
```



Dans un fichier seul

- Extension : .svg ou .xml
- Lisible directement par firefox.
- Internet explorer a besoin d'un plugin.

```
<?xml version="1.0" encoding="utf-8"?>
<svg xmlns="http://www.w3.org/2000/svg"
      viewBox="0 0 300 300" width="4cm">
  <rect fill="red" stroke="blue" stroke-width="10"
        x="10" y='10' width='100' height='200' />
  <rect fill="yellow" stroke="green" stroke-width="10"
        x="30" y='40' width='200' height='100' />
  <text x='100' y='100'>Hello!</text>
</svg>
```

Inclusion dans un fichier XHTML

- Extension du fichier : toujours .xhtml
- Utilisation des espaces de nom

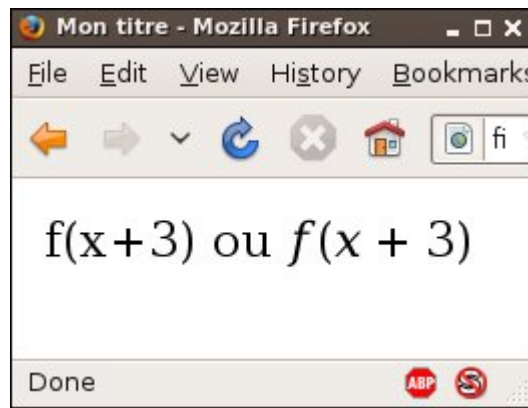
```
<?xml version="1.0" encoding="utf-8"?>
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <title>Mon titre</title>
  </head>
  <body>
    <svg xmlns="http://www.w3.org/2000/svg"
         viewBox="0 0 300 300" width="4cm">
      <rect fill="red" stroke="blue" stroke-width="10"
            x="10" y='10' width='100' height='200' />
      <rect fill="yellow" stroke="green" stroke-width="10"
            x="30" y='40' width='200' height='100' />
      <text x='100' y='100'>Hello!</text>
    </svg>
  </body>
</html>
```

Avec des préfixes...

```
<?xml version="1.0" encoding="utf-8"?>
<html:html xmlns:html="http://www.w3.org/1999/xhtml">
  <html:head>
    <html:title>Mon titre</html:title>
  </html:head>
  <html:body>
    <svg:svg xmlns:svg="http://www.w3.org/2000/svg"
      viewBox="0 0 300 300" width="4cm">
      <svg:rect fill="red" stroke="blue" stroke-width="10"
        x="10" y='10' width='100' height='200' />
      <svg:rect fill="yellow" stroke="green" stroke-width="10"
        x="30" y='40' width='200' height='100' />
      <svg:text x='100' y='100'>Hello!</svg:text>
    </svg:svg>
  </html:body>
</html:html>
```

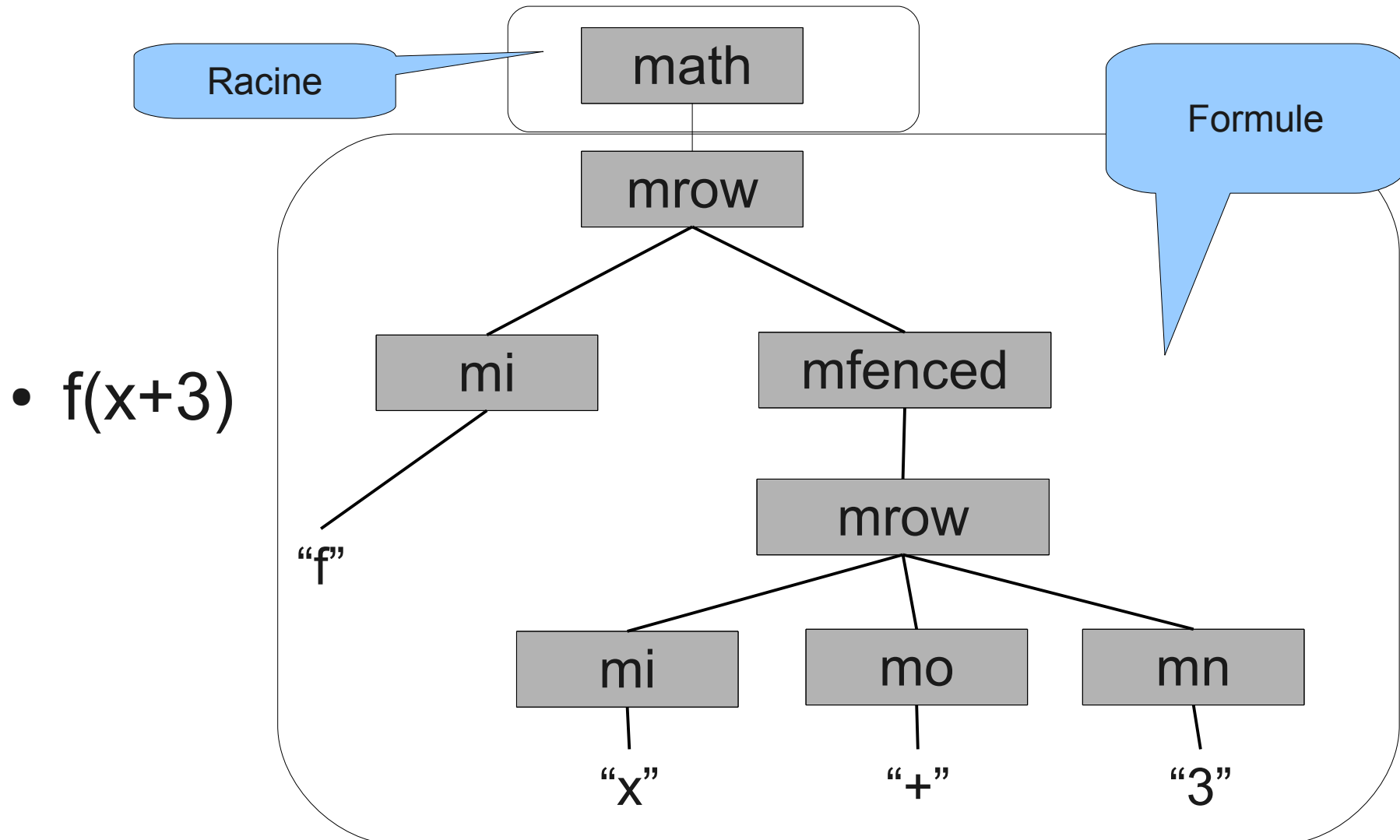
MathML

- Encoder des mathématiques dans du texte
 - Pas juste une suite de lettres
 - Exposants
 - Fractions
 - Racines
 - Espacement entre les symboles particulier suivant leur sens.



Idée derrière MathML

- Concevoir une formule comme un arbre :



La notion de formule

- Une formule est un arbre, attaché à la racine.
- Pour accoler plusieurs éléments en ligne:

`<mrow> ... </mrow>`

- Les formules élémentaires sont :

- Les identifiants (des variables) :

`<mi>x</mi> <mi>f</mi> <mi>y</mi>`

- Les opérateurs :

`<mo>+</mo> <mo>-</mo> <mo>·</mo>`

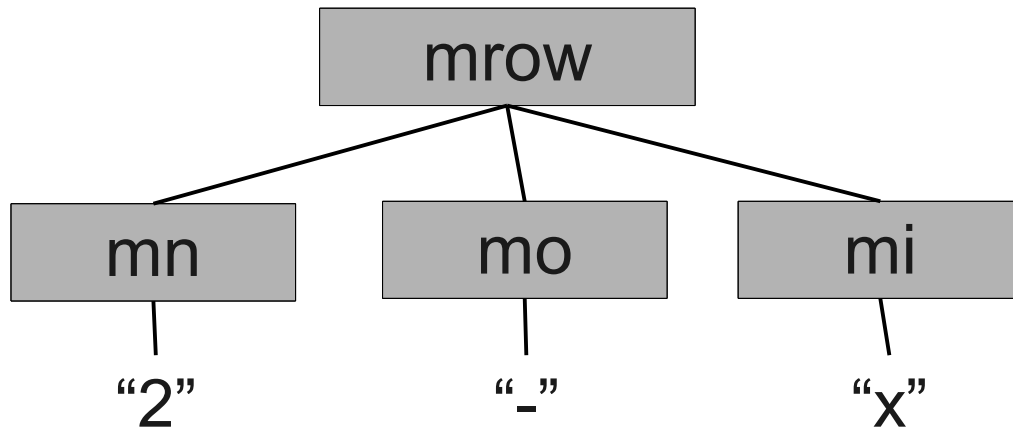
`<mo>=</mo> <mo><</mo> ...`

- Les nombres :

`<mn>1234</mn>`

Example

```
<mrow>  
  <mn>2</mn>  
  <mo>-</mo>  
  <mi>x</mi>  
</mrow>
```



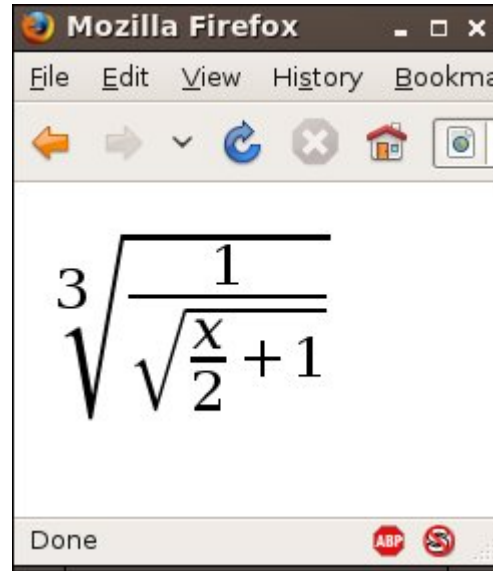
Formules complexes

- **Fractions :**
`<mfrac> numérateur dénominateur </mfrac>`
- **Racine :**
`<msqrt> radical </msqrt>`
`<mroot> radical exposant </mroot>`
- **Exposant :**
`<msup> formule exposant </msup>`
- **Parenthèses :**
`<mfenced> formule </mfenced>`

Exemple

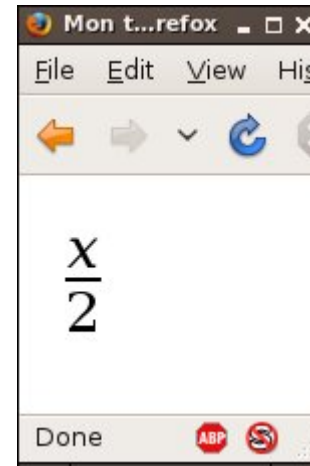
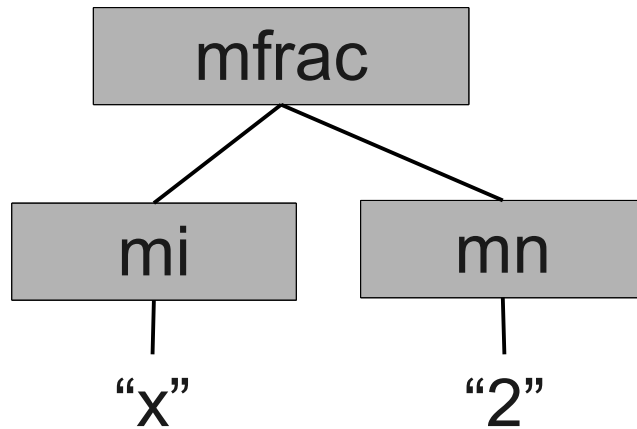
- On veut représenter la formule

$$\sqrt[3]{\frac{1}{\sqrt{\frac{x}{2} + 1}}}$$



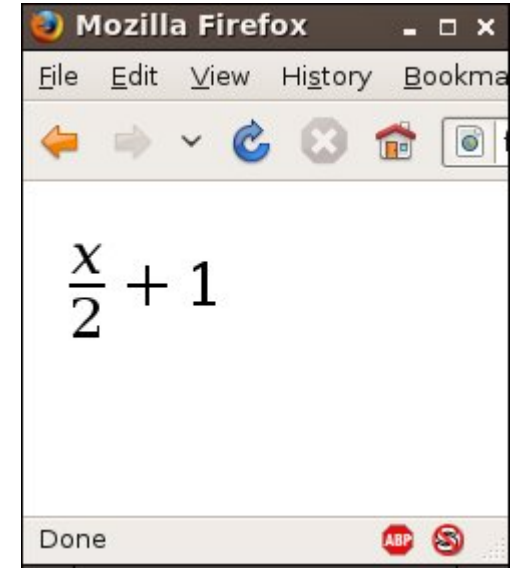
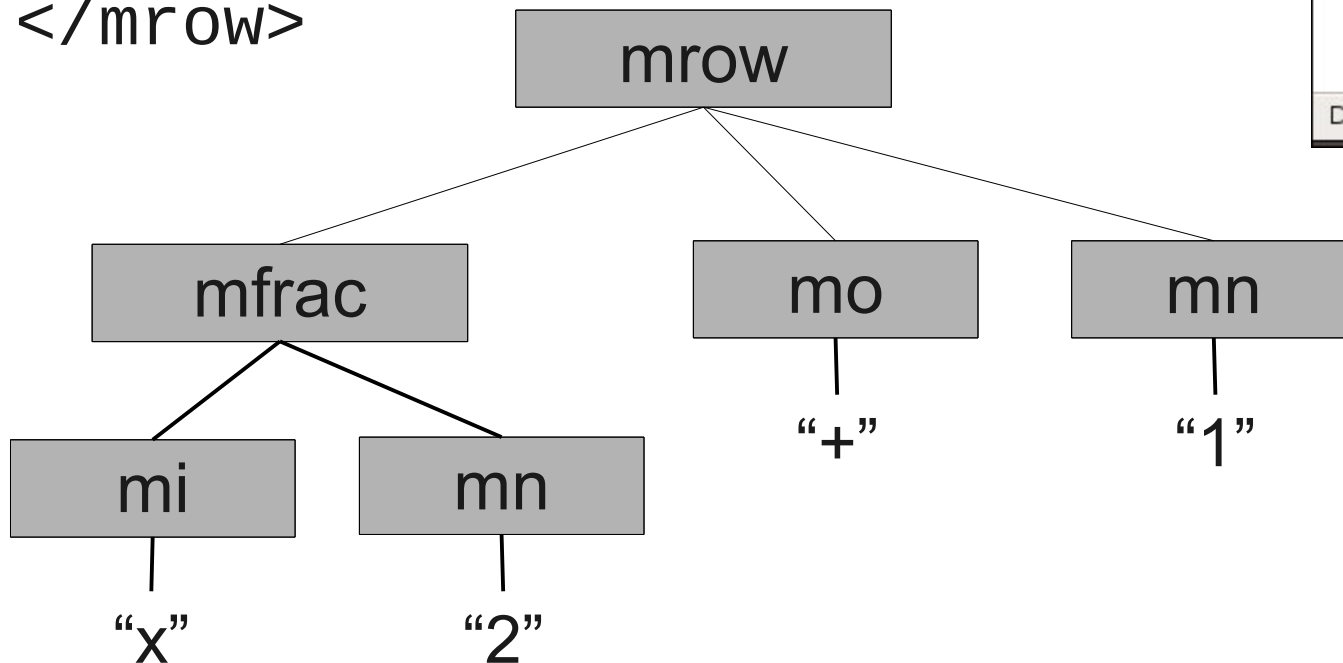
Exemple

```
<mfrac>  
  <mi>x</mi>  
  <mn>2</mn>  
</mfrac>
```



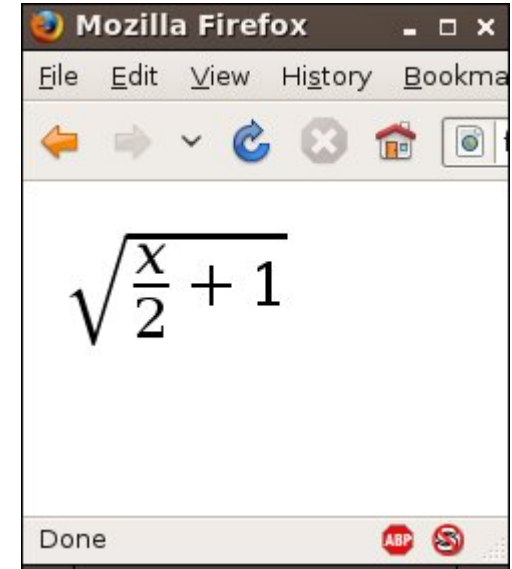
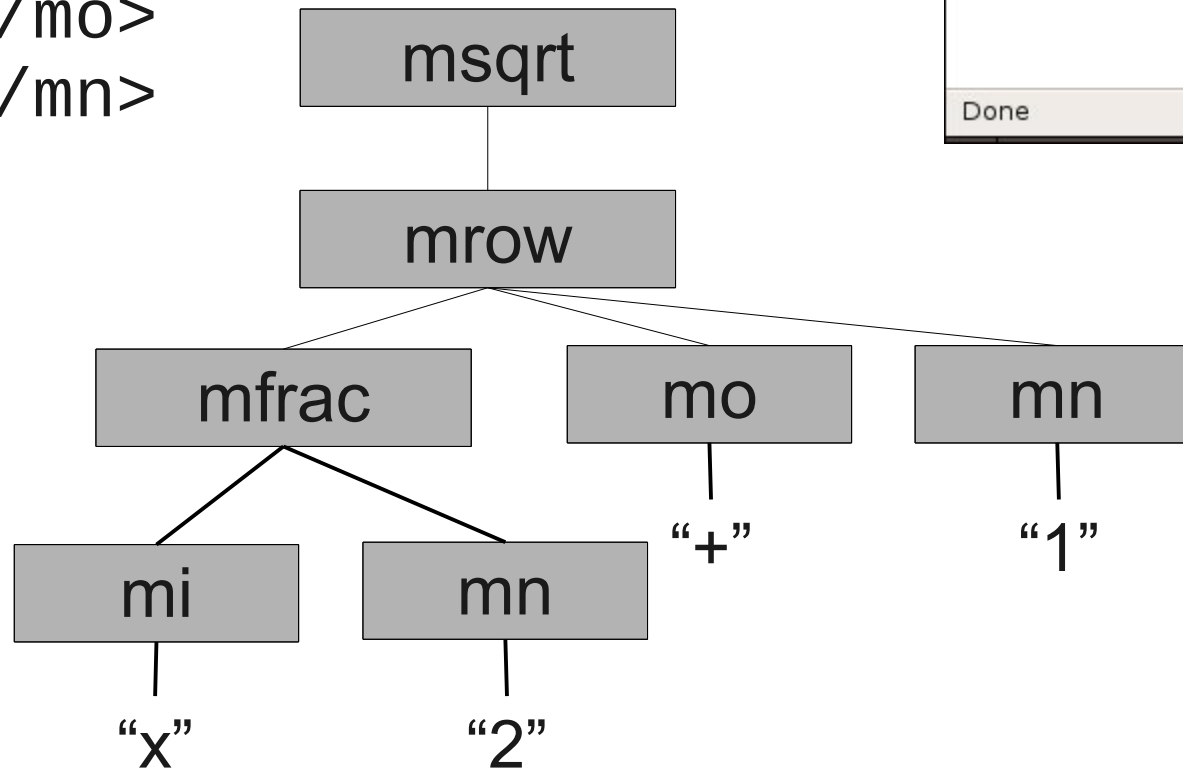
Exemple

```
<mrow>  
  <mfrac>  
    <mi>x</mi>  
    <mn>2</mn>  
  </mfrac>  
  <mo>+</mo>  
  <mn>1</mn>  
</mrow>
```



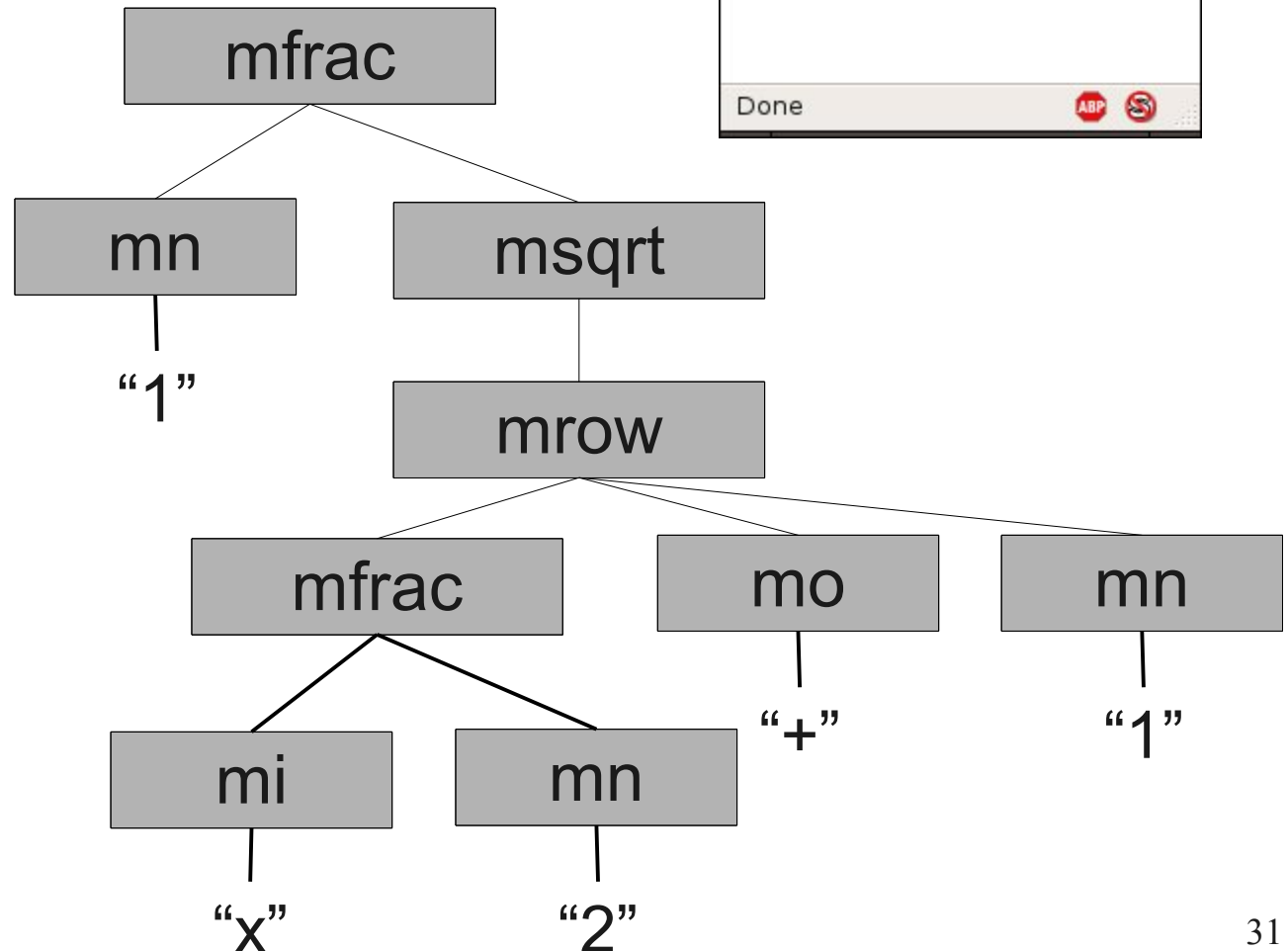
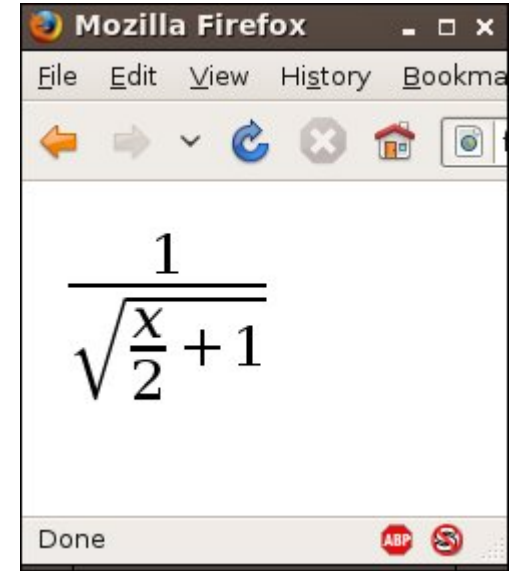
Exemple

```
<msqrt>  
  <mrow>  
    <mfrac>  
      <mi>x</mi>  
      <mn>2</mn>  
    </mfrac>  
    <mo>+</mo>  
    <mn>1</mn>  
  </mrow>  
</msqrt>
```



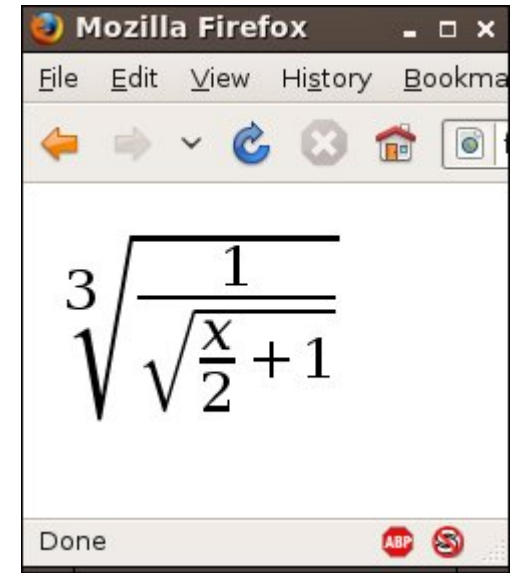
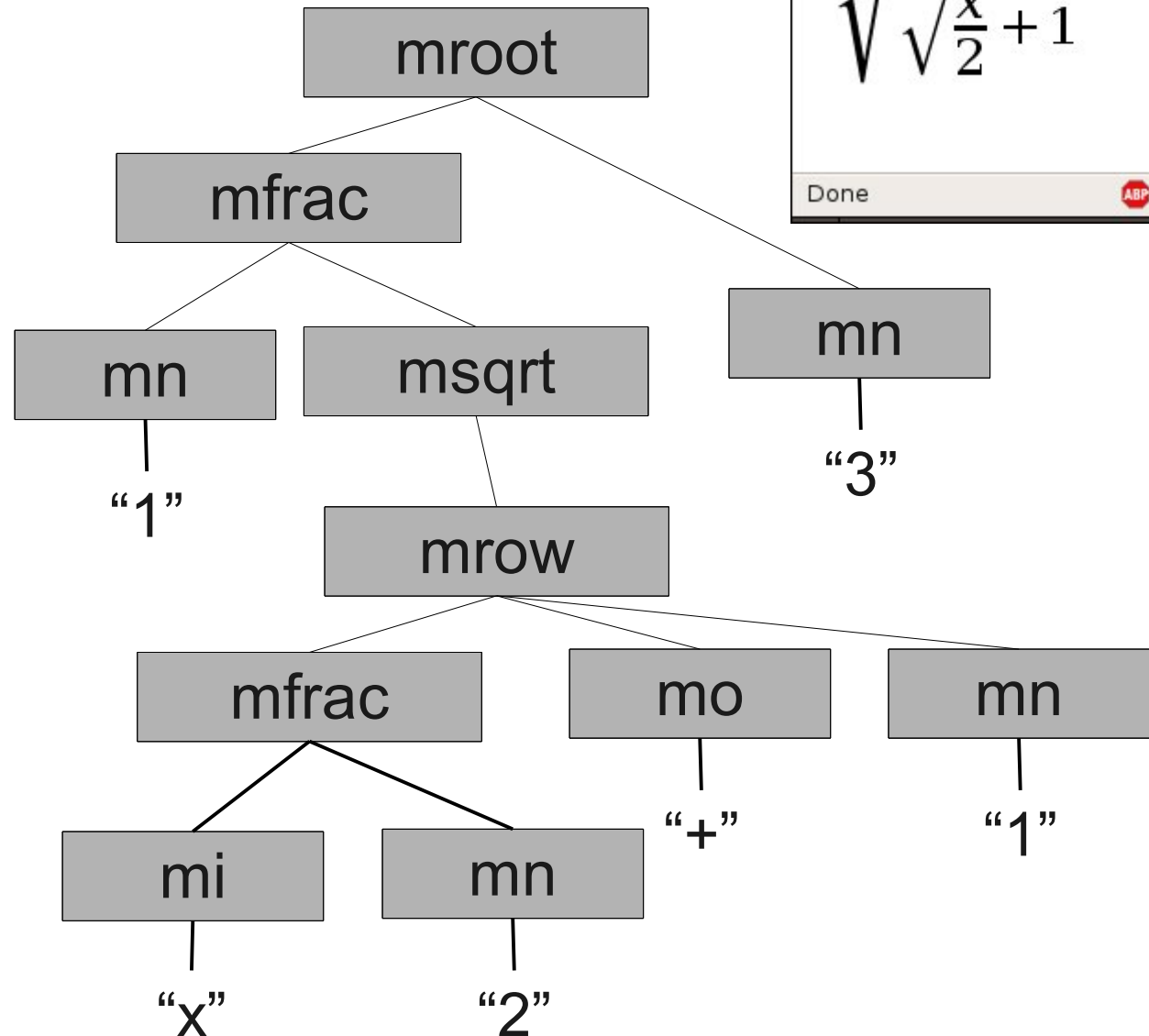
Exemple

```
<mfrac>  
  <mn>1</mn>  
  <msqrt>  
    <mrow>  
      <mfrac>  
        <mi>x</mi>  
        <mn>2</mn>  
      </mfrac>  
      <mo>+</mo>  
      <mn>1</mn>  
    </mrow>  
  </msqrt>  
</mfrac>
```



Exemple

```
<mroot>  
  <mfrac>  
    <mn>1</mn>  
    <msqrt>  
      <mrow>  
        <mfrac>  
          <mi>x</mi>  
          <mn>2</mn>  
        </mfrac>  
        <mo>+</mo>  
        <mn>1</mn>  
      </mrow>  
    </msqrt>  
  </mfrac>  
<mn>3</mn>  
</mroot>
```



Un fichier MathML

- Un document xml
- Element racine `$...$`
- Espace de nom :

<http://www.w3.org/1998/Math/MathML>

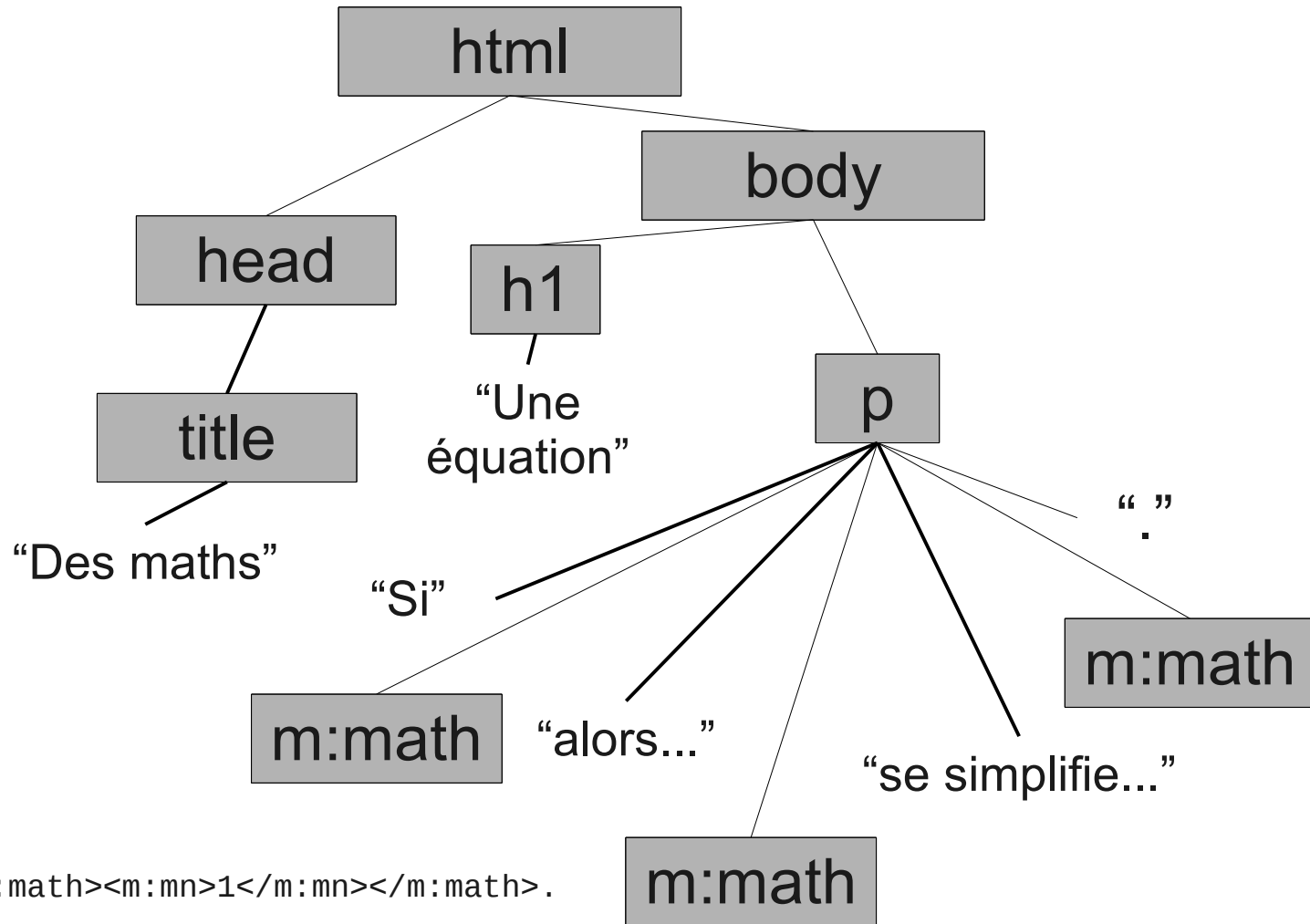
```
<?xml version="1.0"?>
<math xmlns="http://www.w3.org/1998/Math/MathML">
  <mroot>
    <mfrac>
      <mn>1</mn>
      <msqrt>
        <mrow>
          <mfrac>
            <mi>x</mi>
            <mn>2</mn>
          </mfrac>
          <mo>+</mo>
          <mn>1</mn>
        </mrow>
      </msqrt>
    </mfrac>
    <mn>3</mn>
  </mroot>
</math>
```

```

<?xml version="1.0"?>
<html xmlns="http://www.w3.org/1999/xhtml"
      xmlns:m="http://www.w3.org/1998/Math/MathML">
<head><title>Des mathématiques</title></head>
<body>
<h1>Une équation</h1>
<p>
Si
<m:math>
  <m:mi>x</m:mi>
  <m:mo>=</m:mo>
  <m:mn>0</m:mn>
</m:math>,
alors la formule
<m:math>
  <m:mroot>
    <m:mfrac>
      <m:mn>1</m:mn>
      <m:msqrt>
        <m:mrow>
          <m:mfrac>
            <m:mi>x</m:mi>
            <m:mn>2</m:mn>
          </m:mfrac>
          <m:mo>+</m:mo>
          <m:mn>1</m:mn>
        </m:mrow>
        </m:msqrt>
      </m:mfrac>
      <m:mn>3</m:mn>
    </m:mroot>
  </m:math>
se simplifie pour donner <m:math><m:mn>1</m:mn></m:math>.
</p>
</body>
</html>

```

Du XHTML et du MathML

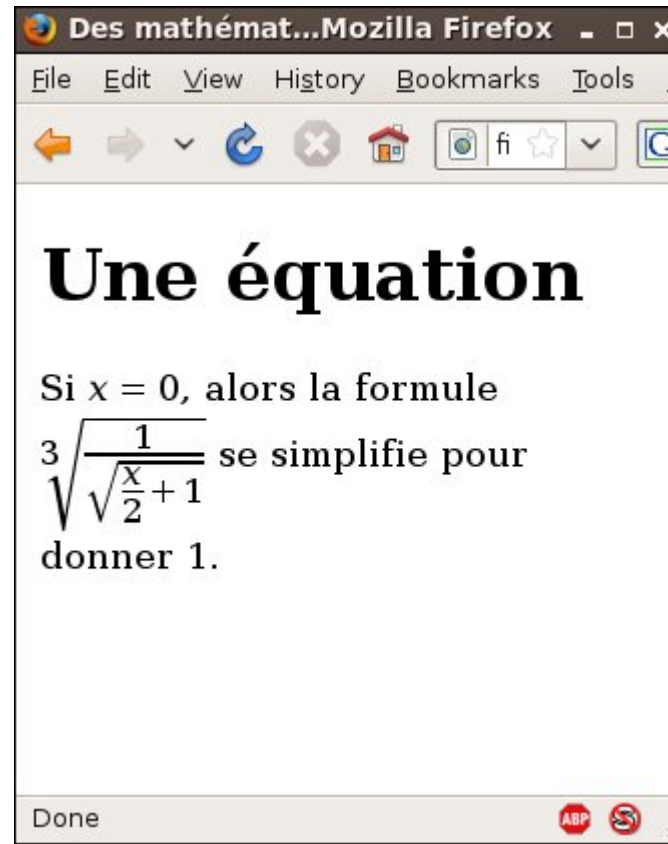


```

<?xml version="1.0"?>
<html xmlns="http://www.w3.org/1999/xhtml"
      xmlns:m="http://www.w3.org/1998/Math/MathML">
<head><title>Des mathématiques</title></head>
<body>
<h1>Une équation</h1>
<p>
Si
<m:math>
  <m:mi>x</m:mi>
  <m:mo>=</m:mo>
  <m:mn>0</m:mn>
</m:math>,
alors la formule
<m:math>
  <m:mroot>
    <m:mfrac>
      <m:mn>1</m:mn>
      <m:msqrt>
        <m:mrow>
          <m:mfrac>
            <m:mi>x</m:mi>
            <m:mn>2</m:mn>
          </m:mfrac>
          <m:mo>+</m:mo>
          <m:mn>1</m:mn>
        </m:mrow>
      </m:msqrt>
    </m:mfrac>
    <m:mn>3</m:mn>
  </m:mroot>
</m:math>
se simplifie pour donner <m:math><m:mn>1</m:mn></m:math>.
</p>
</body>
</html>

```

Du XHTML et du MathML



Flux de données : RSS

- Plusieurs définitions :
 - Really Simple Syndication
 - Rich Site Summary
- Pour un flux d'informations :
 - Sites d'information
 - Liste de contenu
 - Blogs
 - Nouveautés sur un site
- Pas d'espace de nom par défaut.

Format RSS

- Exemple :

```
<rss>
<channel>
<title>Liste des leçons</title>
<item>
  <title>XML et DTDs</title>
  <pubDate>Semaine 38</pubDate>
  <description> ... </description>
</item>
<item>
  <title>Espaces de noms</title>
  <pubDate>Semaine 39</pubDate>
  <description> ... </description>
</item>
...
</channel>
</rss>
```

- DTD (simplifiée):

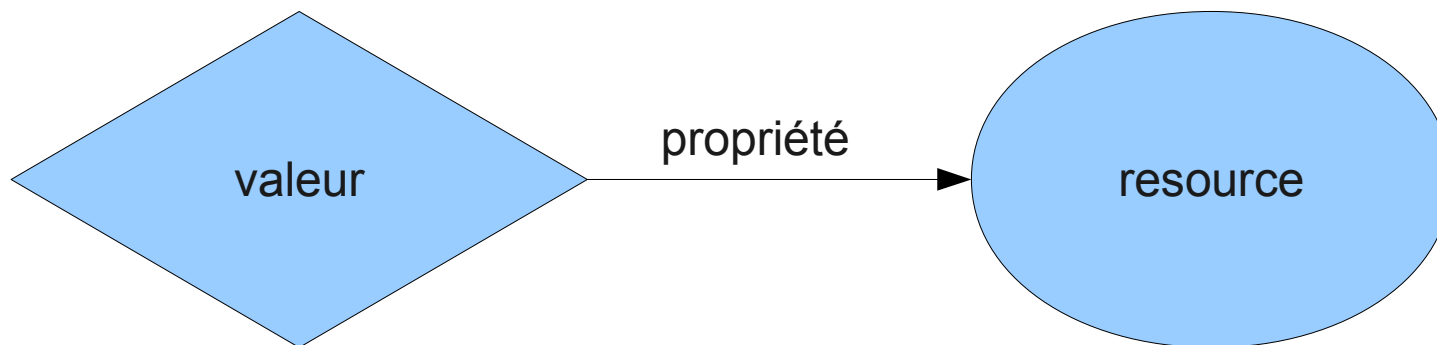
```
<!ELEMENT rss (channel+)>
<!ELEMENT channel
  (title, description?, item*)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT description (#PCDATA)>
<!ELEMENT item
  (title|pubDate|description)*>
<!ELEMENT pubDate (#PCDATA)>
```

Les métadonnées.

- Problème classique : Ajout dans un XML d'informations qui ne sont pas du contenu.
 - Auteur(s)
 - Date de création, de mise à jour
 - Description
 - Lien externe
 - ...

Format RDF

- Modèle de donnée simple :



```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <rdf:Description about="http://inf356.monoidal.net">
    <author>Benoît Valiron</author>
  </rdf:Description>
</rdf:RDF>
```

Quel langage ?

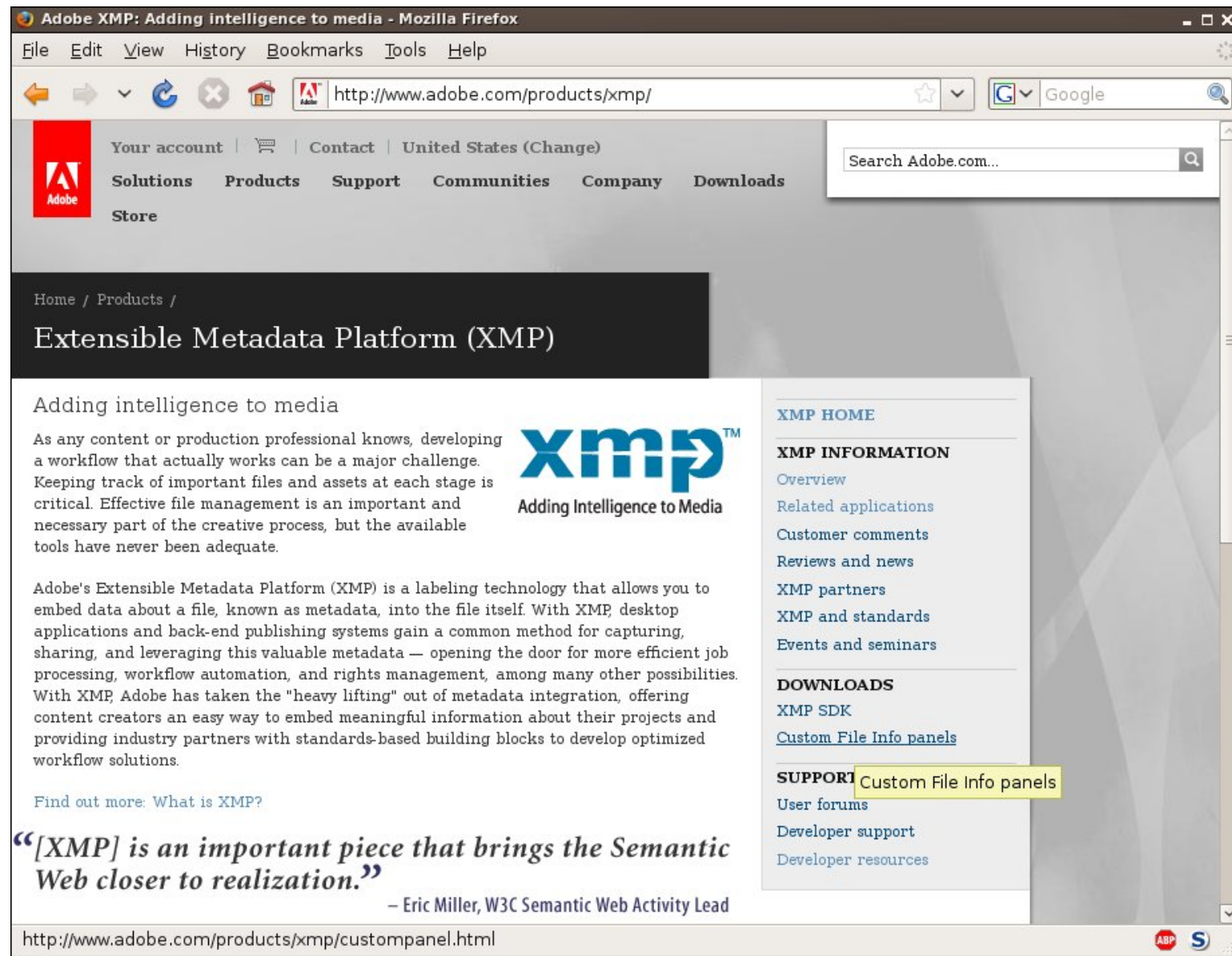
- Chaque sous-élément de Description est une propriété qui contient une valeur textuelle ou XML.

Format Dublin Core

- Donner une structure standardisée à des méta-données
- En particulier aux propriétés de RDF.
- Pas d'arbre par défaut, juste un ensemble de 15 éléments avec un sens bien défini
 - title
 - creator
 - subject
 - description
 - publisher
 - contributor
 - date
 - type
 - format
 - identifier
 - source
 - language
 - relation
 - coverage
 - rights

Méta-données dans un PDF

Adobe utilise un format de méta-données insérées dans les PDF : le format XMP.



The screenshot shows a Mozilla Firefox browser window displaying the Adobe XMP website. The address bar shows the URL <http://www.adobe.com/products/xmp/>. The page title is "Adobe XMP: Adding intelligence to media - Mozilla Firefox". The browser's menu bar includes File, Edit, View, History, Bookmarks, Tools, and Help. The address bar also shows a search engine dropdown set to Google. The website header includes the Adobe logo, navigation links for "Your account", "Contact", "United States (Change)", "Solutions", "Products", "Support", "Communities", "Company", and "Downloads", and a search bar labeled "Search Adobe.com...". The main content area features a breadcrumb trail "Home / Products /" followed by the heading "Extensible Metadata Platform (XMP)". Below this, the text reads "Adding intelligence to media" and "As any content or production professional knows, developing a workflow that actually works can be a major challenge. Keeping track of important files and assets at each stage is critical. Effective file management is an important and necessary part of the creative process, but the available tools have never been adequate." To the right of this text is the XMP logo with the tagline "Adding Intelligence to Media". Below the text, it states "Adobe's Extensible Metadata Platform (XMP) is a labeling technology that allows you to embed data about a file, known as metadata, into the file itself. With XMP, desktop applications and back-end publishing systems gain a common method for capturing, sharing, and leveraging this valuable metadata — opening the door for more efficient job processing, workflow automation, and rights management, among many other possibilities. With XMP, Adobe has taken the 'heavy lifting' out of metadata integration, offering content creators an easy way to embed meaningful information about their projects and providing industry partners with standards-based building blocks to develop optimized workflow solutions." A link "Find out more: What is XMP?" is provided. A quote is displayed: "[XMP] is an important piece that brings the Semantic Web closer to realization." attributed to "– Eric Miller, W3C Semantic Web Activity Lead". A right-hand sidebar contains navigation links under "XMP HOME", "XMP INFORMATION", "DOWNLOADS", and "SUPPORT". The "SUPPORT" link is highlighted with a yellow box, and the sub-link "Custom File Info panels" is also highlighted. The browser's status bar at the bottom shows the URL <http://www.adobe.com/products/xmp/custompanel.html> and logos for Adobe and the Semantic Web.

Du contenu XML dans un PDF

```
<?xpacket begin="" id="W5M0MpCehiHzreSzNTczkc9d"?>
<x:xmpmeta xmlns:x="adobe:ns:meta/"
  x:xmptk="Adobe XMP Core 4.0-c316 44.253921, Sun Oct 01 2006 17:08:23">
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    <rdf:Description rdf:about="" xmlns:xap="http://ns.adobe.com/xap/1.0/">
      <xap:ModifyDate>2009-07-10T11:03:46+02:00</xap:ModifyDate>
      <xap:CreateDate>2009-07-10T11:03:46+02:00</xap:CreateDate>
      <xap:MetadataDate>2009-07-10T11:03:46+02:00</xap:MetadataDate>
      <xap:CreatorTool>Vector Data Adobe Photoshop 7.0</xap:CreatorTool>
    </rdf:Description>
    <rdf:Description rdf:about="" xmlns:dc="http://purl.org/dc/elements/1.1/">
      <dc:format>application/pdf</dc:format>
    </rdf:Description>
    <rdf:Description rdf:about=""
      xmlns:xapMM="http://ns.adobe.com/xap/1.0/mm/">
      <xapMM:DocumentID>uuid:d11d6b8c-3219-44f0-b663-460ea508aaf7</xapMM:DocumentID>
      <xapMM:InstanceID>uuid:fd9e6ee6-1963-436f-b908-8ddc64f92924</xapMM:InstanceID>
    </rdf:Description>
    <rdf:Description rdf:about="" xmlns:pdf="http://ns.adobe.com/pdf/1.3/">
      <pdf:Producer>Adobe Photoshop for Windows</pdf:Producer>
    </rdf:Description>
  </rdf:RDF>
</x:xmpmeta>
```

```
<?xpacket begin="" id="W5M0MpCehiHzreSzNTczkc9d"?>
<x:xmpmeta xmlns:x="adobe:ns:meta/" x:xmptk="3.1-701">
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
    <rdf:Description rdf:about=""
      xmlns:pdf="http://ns.adobe.com/pdf/1.3/">
      <pdf:Producer>Acrobat Distiller 7.0 (Windows)</pdf:Producer>
    </rdf:Description>
    <rdf:Description rdf:about="" xmlns:xap="http://ns.adobe.com/xap/1.0/">
      <xap:CreateDate>2009-06-16T09:53:18+02:00</xap:CreateDate>
      <xap:CreatorTool>PScript5.dll Version 5.2.2</xap:CreatorTool>
      <xap:ModifyDate>2009-06-16T09:53:18+02:00</xap:ModifyDate>
    </rdf:Description>
    <rdf:Description rdf:about=""
      xmlns:dc="http://purl.org/dc/elements/1.1/">
      <dc:format>application/pdf</dc:format>
      <dc:title>
        <rdf:Alt>
          <rdf:li xml:lang="x-default">Microsoft PowerPoint -
            090615 DP PTTN Lancement Nord Cortentin</rdf:li>
        </rdf:Alt>
      </dc:title>
      <dc:creator>
        <rdf:Seq>
          <rdf:li>hfarnham</rdf:li>
        </rdf:Seq>
      </dc:creator>
    </rdf:Description>
    <rdf:Description rdf:about="" xmlns:xapMM="http://ns.adobe.com/xap/1.0/mm/">
      <xapMM:DocumentID>uuid:15594a7b-a0a6-43d3-9b98-481e4015c8e2</xapMM:DocumentID>
      <xapMM:InstanceID>uuid:51799985-3abe-4c00-81e6-0167593cd477</xapMM:InstanceID>
    </rdf:Description>
  </rdf:RDF>
</x:xmpmeta>
```

Méta-données ailleurs ?

- Dans du RSS : Exemple au TD-2/TP-2
- Dans du XHTML :
 - Pas de standard.
 - Balise meta dans l'entête
 - Inclusion avec RDF dans l'entête : casse la DTD.
- Pour donner un sens à tout fichier de données non-standard.