

AMPLEXOR

XML

→ Paul Hermans

- what is?
- where does it come from?
- meta-language
- how does it look like?
- main concepts
- where is it used?
- working with XML

[article](#) [discussion](#) [edit this page](#) [history](#)

*Your **continued donations** keep Wikipedia running!*

XML

From Wikipedia, the free encyclopedia

(Redirected from Xml)

The **Extensible Markup Language (XML)** is a general-purpose [markup language](#).^[1] Its primary purpose is to facilitate the sharing of data across different information systems, particularly via the [Internet](#).^[2]

It is a simplified subset of the [Standard Generalized Markup Language](#) (SGML), and is designed to be relatively human-legible. By adding semantic constraints, application languages can be implemented in XML. These include [XHTML](#),^[3] [RSS](#), [MathML](#), [GraphML](#), [Scalable Vector Graphics](#), [MusicXML](#), and thousands of others. Moreover, XML is sometimes used as the [specification language](#) for such application languages.

XML is recommended by the [World Wide Web Consortium](#). It is a fee-free [open standard](#). The W3C recommendation specifies both the [lexical grammar](#), and the requirements for [parsing](#).

Contents [hide]

- 1 Well-formed and valid XML documents
- 2 Well-formed documents: XML syntax
 - 2.1 Entity references
 - 2.1.1 Numeric character references
 - 2.2 Well-formed documents
 - 2.3 Automatic verification

Extensible Markup Language

File extension:	.xml
MIME type:	application/xml, text/xml (deprecated)
Uniform Type Identifier:	public.xml
Developed by:	World Wide Web Consortium
Type of format:	Markup language
Extended from:	SGML
Extended to:	XHTML, RSS, Atom, ...
Standard(s):	1.0 (Fourth Edition) ↗ 1.1 (Second Edition) ↗

→ **HTML** (HyperText Markup Language)

- Fixed tagset
- Too much formatting oriented
- Not enough formatting possibilities

→ **SGML** (Standard Generalized Markup Language)

- Much too complicated
- Too expensive

→ thus: **XML** (eXtensible Markup Language)

→ Evolution:

- Coming from the **document** world
- taking over the **data** world

→ Meta-language

- ▭ XHTML
- ▭ MathML
- ▭ SVG
- ▭ XSL
- ▭ XSLT
- ▭ XSD
- ▭ SOAP
- ▭ WSDL
- ▭ UDDI
- ▭ DITA
- ▭ Docbook
- ▭ TEI
- ▭ ...

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet href="mystyle.css" type="text/css"?>
<Order xmlns="urn:oasis:names:tc:ubl:Order:1.0:0.70"
  xmlns:po="urn:oasis:names:tc:ubl:Order:1.0:0.70"
  xmlns:cat="urn:oasis:names:tc:ubl:CommonAggregateTypes:1.0:0.70"
  xmlns:cct="urn:oasis:names:tc:ubl:CoreComponentTypes:1.0:0.70"
  xmlns:ccts="urn:oasis:names:tc:ubl:CoreComponentParameters:1.0:0.70"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:xs="http://www.w3.org/2001/XMLSchema"
  xsi:schemaLocation="urn:oasis:names:tc:ubl:Order:1.0:0.70
    UBL_Library_0p70_Order.xsd">
  <!-- Reviewed by Paul Hermans -->
  <cat:ID>20031234-1</cat:ID>
  <cat:IssueDate>2003-02-02</cat:IssueDate>
  <cat:LineExtensionTotalAmount currencyID="USD">498.5</cat:LineExtensionTotalAmount>
  <cat:BuyerParty>
    <cat:ID/>
    <cat:PartyName>
      <cat:Name>Bills Microdevices</cat:Name>
    </cat:PartyName>
    <cat:Address>
      <cat:ID/>
      <cat:Street>413 Spring St</cat:Street>
      <cat:CityName>Elgin</cat:CityName>
      <cat:PostalZone>60123</cat:PostalZone>
      <cat:CountrySub-Entity>IL</cat:CountrySub-Entity>
    </cat:Address>
    <cat:BuyerContact>
      <cat:ID/>
      <cat:Name>George Tirebiter</cat:Name>
    </cat:BuyerContact>
  </cat:BuyerParty>
  <cat:SellerParty>
    <cat:ID>JOS</cat:ID>
    <cat:PartyName>
      <cat:Name>Joes Office Supply</cat:Name>
    </cat:PartyName>
    <cat:Address>
```

What's in an XML file?

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet href="mystyle.css" type="text/css"?>
<Order xmlns="urn:oasis:names:tc:ubl:Order:1.0:0.70"
  xmlns:po="urn:oasis:names:tc:ubl:Order:1.0:0.70"
  xmlns:cat="urn:oasis:names:tc:ubl:CommonAggregateTypes:1.0:0.70"
  xmlns:cct="urn:oasis:names:tc:ubl:CoreComponentTypes:1.0:0.70"
  xmlns:ccts="urn:oasis:names:tc:ubl:CoreComponentParameters:1.0:0.70"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:xs="http://www.w3.org/2001/XMLSchema"
  xsi:schemaLocation="urn:oasis:names:tc:ubl:Order:1.0:0.70
    UBL_Library_0p70_Order.xsd">
  <!-- Reviewed by Paul Hermans -->
  <cat:ID>20031234-1</cat:ID>
  <cat:IssueDate>2003-02-02</cat:IssueDate>
  <cat:LineExtensionTotalAmount currencyID="USD">498.5</cat:LineExtensionTotalAmount>
  <cat:BuyerParty>
    <cat:ID/>
    <cat:PartyName>
      <cat:Name>Bills Microdevices</cat:Name>
    </cat:PartyName>
    <cat:Address>
      <cat:ID/>
      <cat:Street>413 Spring St</cat:
      <cat:CityName>Elgin</cat:CityNe
      <cat:PostalZone>60123</cat:Post
      <cat:CountrySub-Entity>IL</cat:
    </cat:Address>
    <cat:BuyerContact>
      <cat:ID/>
      <cat:Name>George Tirebiter</cat
    </cat:BuyerContact>
  </cat:BuyerParty>
  <cat:SellerParty>
    <cat:ID>JOS</cat:ID>
    <cat:PartyName>
      <cat:Name>Joes Office Supply</cat:Name>
    </cat:PartyName>
    <cat:Address>
```

Elements
Attributes
Comments
Processing Instructions

→ Element well-formedness rules

1. one or more elements
2. one is the root
3. elements nest properly within each other
4. match element's starttag - endtag

→ Attribute well-formedness rules

1. unique attribute specification
2. no "<" in value

Hierarchical data model

XML	
xml-stylesheet	href="mystyle.css" type="text/css"
Order	
xmlns	urn:oasis:names:tc:ubl:Order:1.0:0.70
xmlns:po	urn:oasis:names:tc:ubl:Order:1.0:0.70
xmlns:cat	urn:oasis:names:tc:ubl:CommonAggregateTypes:1.0:0.70
xmlns:cct	urn:oasis:names:tc:ubl:CoreComponentTypes:1.0:0.70
xmlns:ccts	urn:oasis:names:tc:ubl:CoreComponentParameters:1.0:0.70
xmlns:xsi	http://www.w3.org/2001/XMLSchema-instance
xmlns:xs	http://www.w3.org/2001/XMLSchema
xsi:schemaLoca...	urn:oasis:names:tc:ubl:Order:1.0:0.70 UBL_Library_0p70_Order.xsd
Comment	Reviewed by Paul Hermans
cat:ID	20031234-1
cat:IssueDate	2003-02-02
cat:LineExtensionTotalAmount	
currencyID	USD
Text	498.5
cat:BuyerParty	
cat:ID	
cat:PartyName	
cat:Name	Bills Microdevices
cat:Address	
cat:ID	
cat:Street	413 Spring St
cat:CityName	Elgin
cat:PostalZone	60123
cat:CountrySub...	IL
cat:BuyerContact	
cat:SellerParty	
cat:DeliveryTerms	
cat:OrderLine	
cat:BuyersID	1
cat:SellersID	
cat:LineExtensionAmount	currencyID=USD
cat:Quantity	unitCode=
cat:Item	
cat:DeliveryRequirement	
cat:OrderLine	
cat:BuyersID	2
cat:LineExtensionAmount	currencyID=USD
cat:Quantity	unitCode=PKG
cat:Item	
cat:DeliveryRequirement	
cat:OrderLine	

XML is about documents/data

AMPLEXOR

Document format

Portals
Intranet
E-shops
(Web, WAP,
TV, ...)

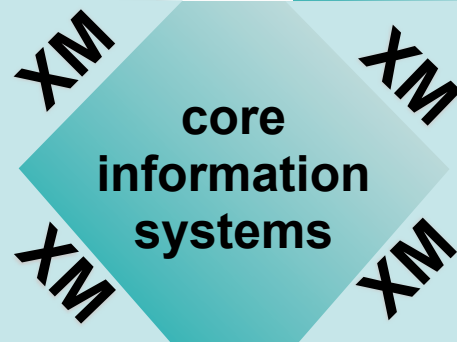
Document
components
Digital assets
Multimedia
Workflow

*managing and
delivering documents*

***“XML as the unifying
information standard”***

Web site content
management

electronic
document
management



B2B
e-commerce

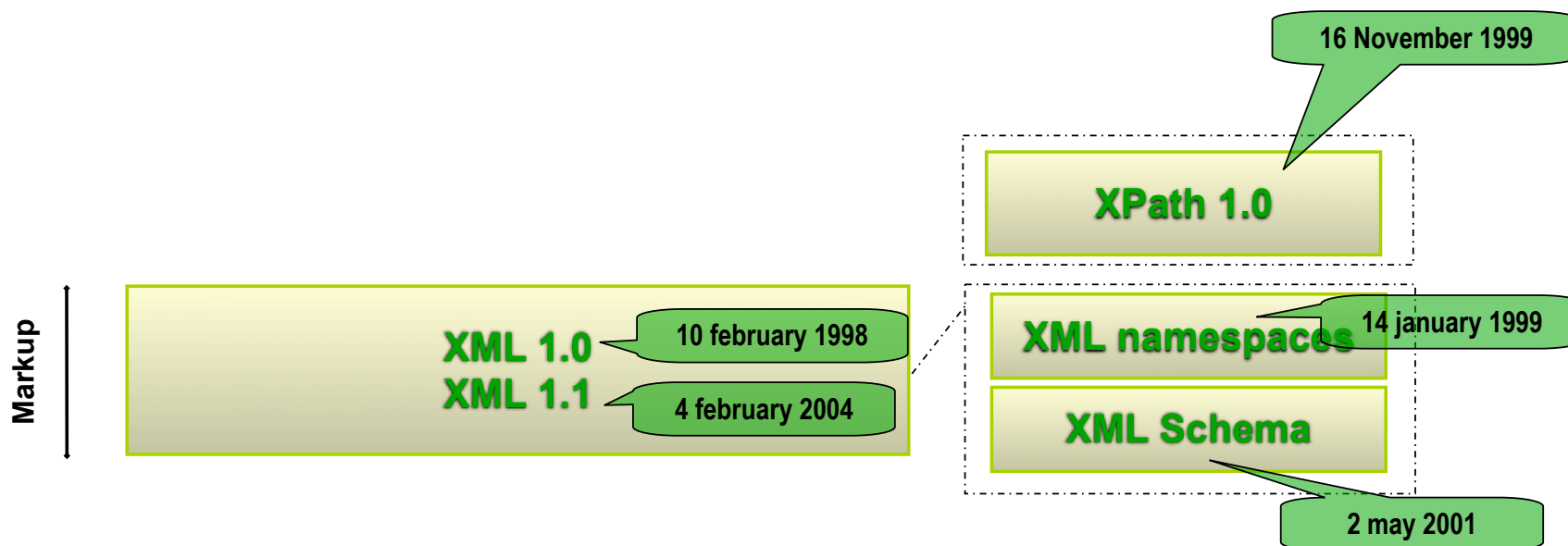
enterprise
application
integration

Data format

E-procurement
Supply chain
management
Commerce
exchanges

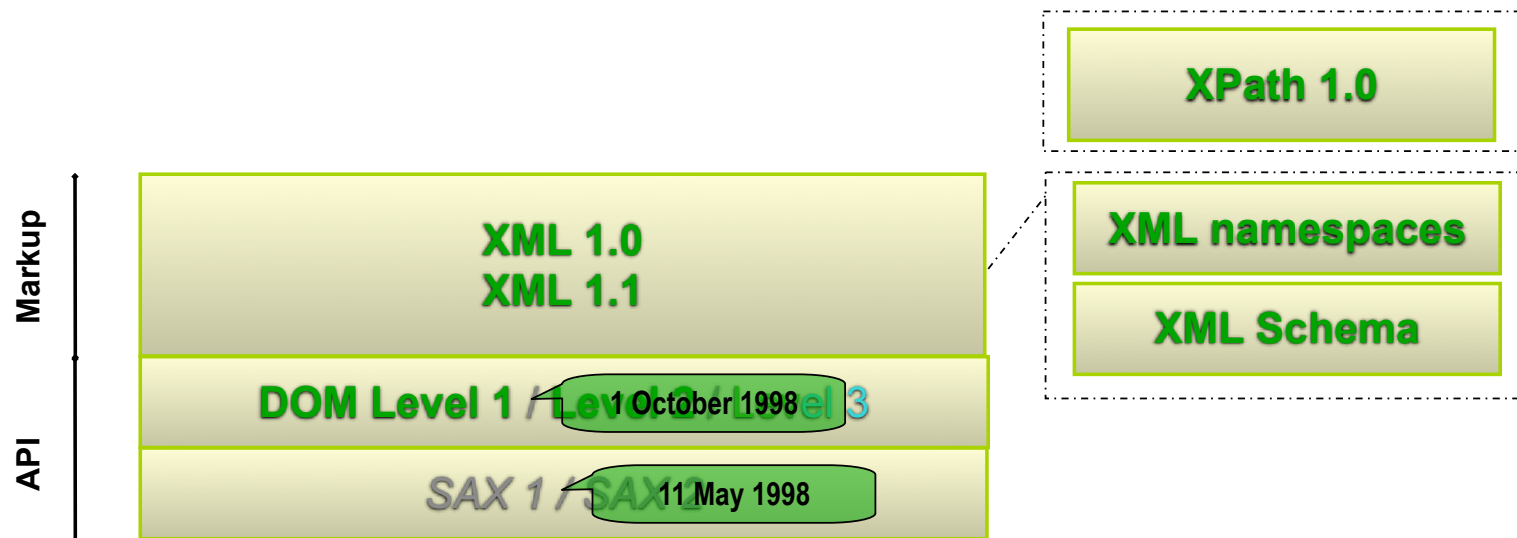
Data
exchange
Distributed
applications
Web services

*exchanging and
processing data*

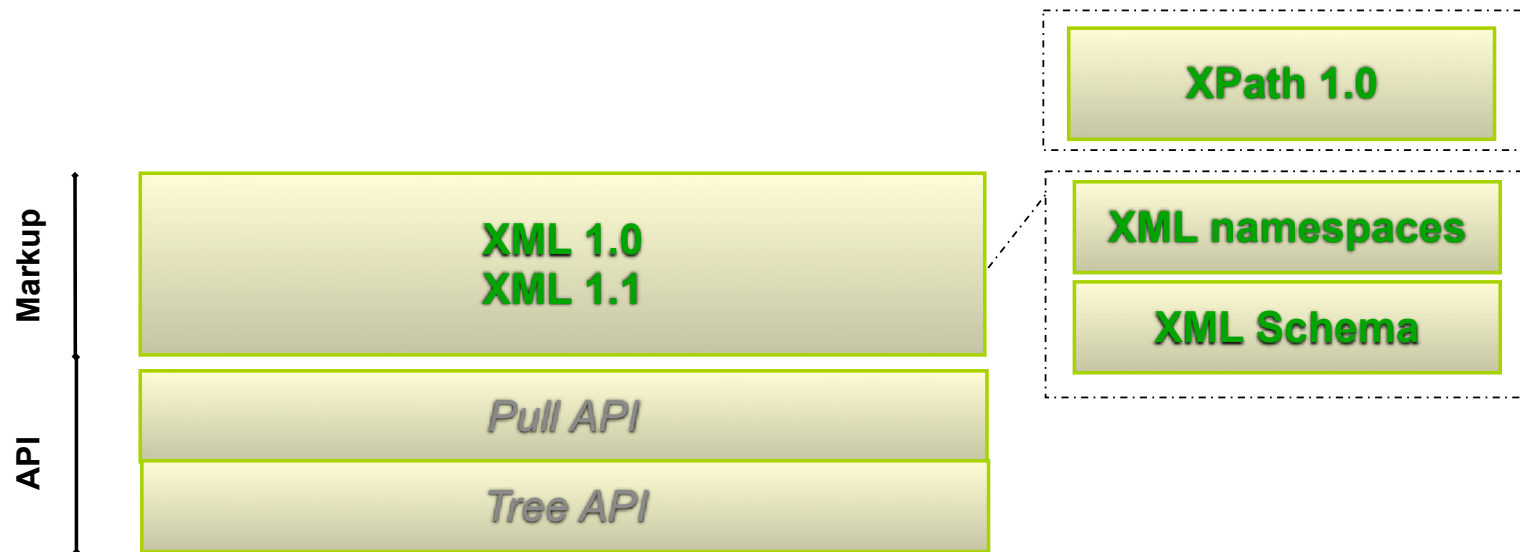


W3C: [Note](#) > [Work. Draft](#) > [Cand. Recom.](#) > [Prop. Recom.](#) > **Recommendation** / *Not W3C*

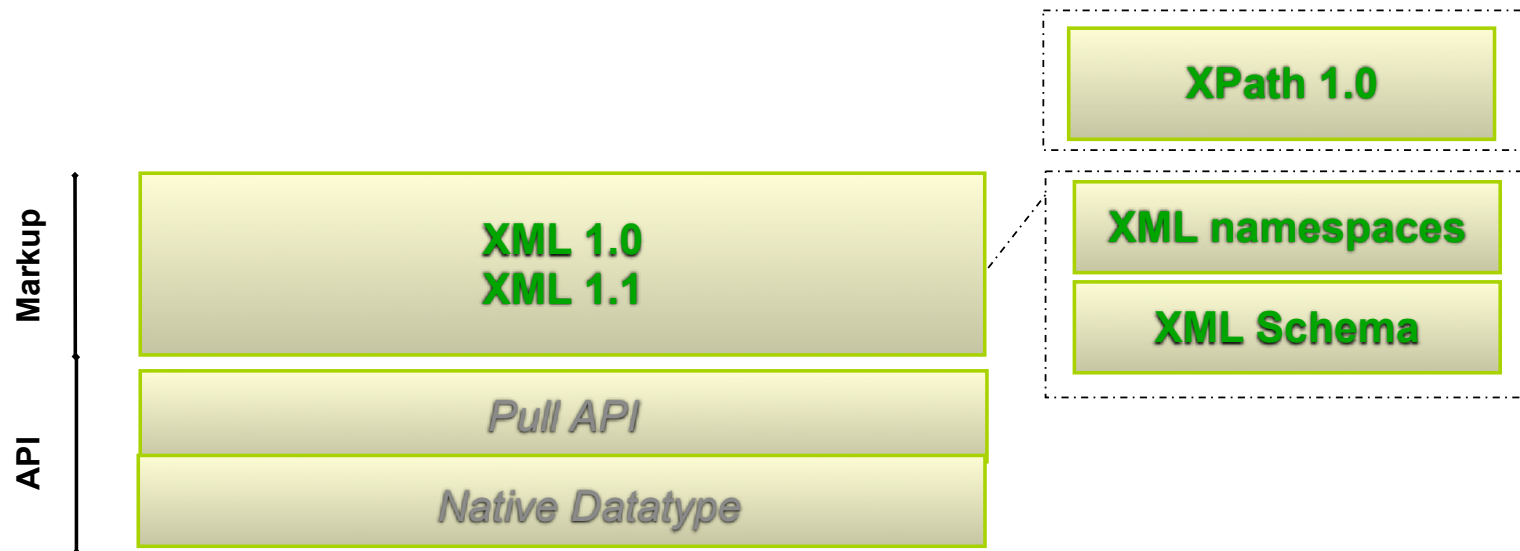
Working with XML in your programming language (past)



W3C: [Note](#) > [Work. Draft](#) > [Cand. Recom.](#) > [Prop. Recom.](#) > **Recommendation** / *Not W3C*

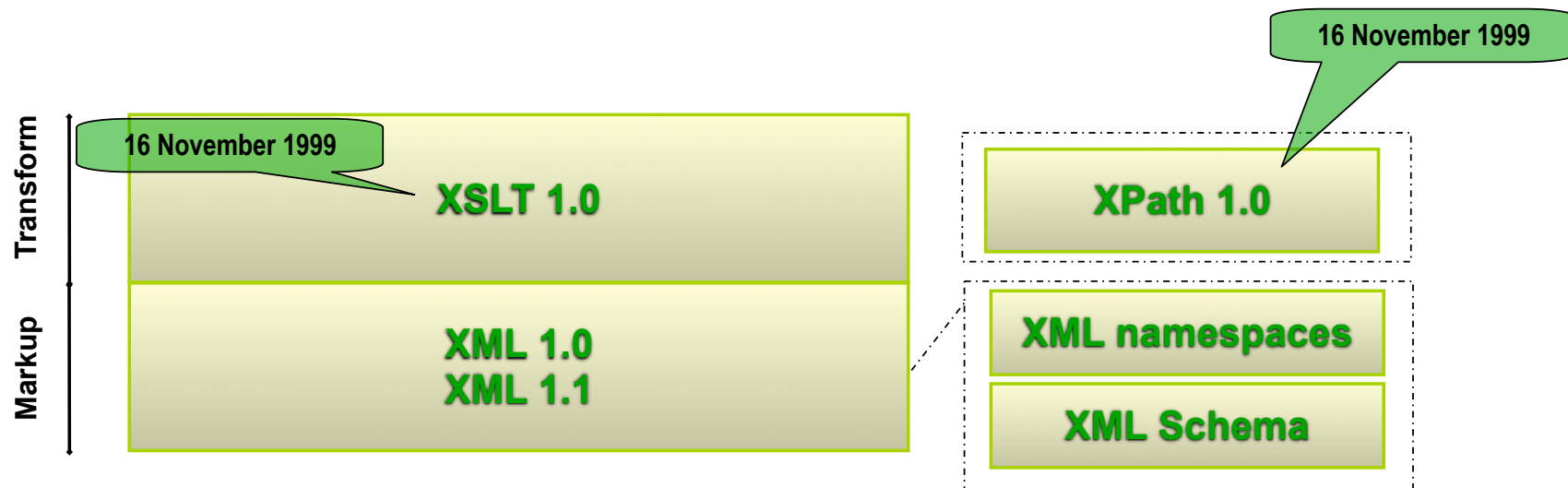


W3C: [Note](#) > [Work. Draft](#) > [Cand. Recom.](#) > [Prop. Recom.](#) > **[Recommendation](#)** / *Not W3C*



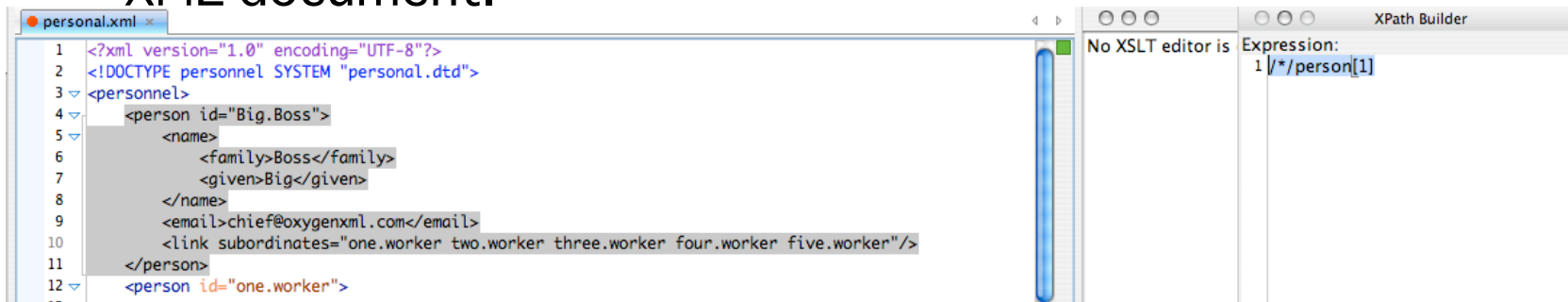
W3C: [Note](#) > [Work. Draft](#) > [Cand. Recom.](#) > [Prop. Recom.](#) > **Recommendation** / *Not W3C*

Native XML approaches (past and present)



W3C: [Note](#) > [Work. Draft](#) > [Cand. Recom.](#) > [Prop. Recom.](#) > [Recommendation](#) / *Not W3C*

- ➔ XPath is an expression language for addressing portions of an XML document, or for computing values (strings, numbers, or boolean values) based on the content of an XML document.

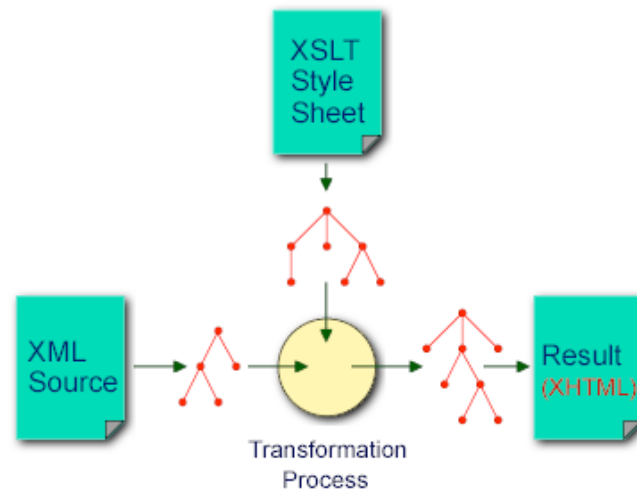


The screenshot shows two windows. The left window, titled 'personal.xml', displays XML code with line numbers 1 through 12. The code is as follows:

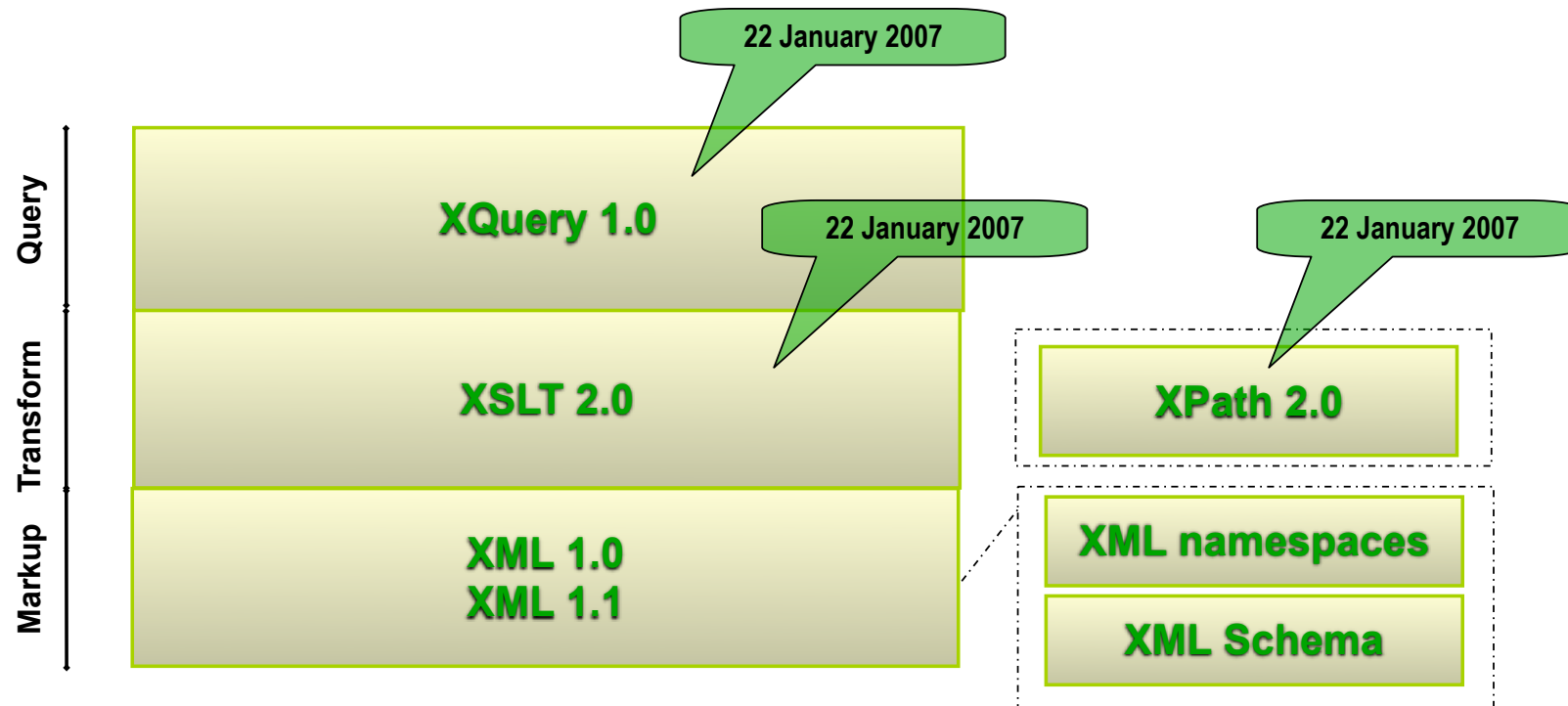
```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <!DOCTYPE personnel SYSTEM "personal.dtd">
3 <personnel>
4   <person id="Big.Boss">
5     <name>
6       <family>Boss</family>
7       <given>Big</given>
8     </name>
9     <email>chief@oxygenxml.com</email>
10    <link subordinates="one.worker two.worker three.worker four.worker five.worker"/>
11  </person>
12  <person id="one.worker">
```

The right window, titled 'XPath Builder', shows the text 'No XSLT editor is' and an 'Expression:' field containing the XPath expression `1 //*/person[1]`.

- XML-based language used for the transformation of XML documents



Native XML approaches (present and future)



**FOR
LET
WHERE
RETURN**

**XPath
Expression**

```
<customers>{  
  for $cust in //customer[division="sales"]  
  let $name := $cust / name / @surname  
  let $value := sum($cust/get-orders()/value)  
  where $value > 1000.00  
  return  
    <customer>  
      <name>{$name}</name>  
      <address>{$cust/address}</address>  
    </customer>  
}</customers>
```

**Element
Constructor**

**XPath
Expression**

```
for $c in //customer, $p in //product
where $c/order/product = $p/code
return <order customer="{ $c/name}" product="{ $p/code}"/>
```

```
SELECT c.name, p.code
FROM customer c, product p, order o
WHERE o.customer = c.number AND o.product = p.code
```

- No updates
- No full text retrieval
- No high-level grouping capability
- no date and number formatting
- control over serialization

- Saxon
- DataDirect XQuery
- Relational databases
 - Oracle
 - IBM DB2
 - Microsoft
- Native XML databases
 - eXist
 - X-Hive
 - MarkLogic
 - TigerLogic
 - Berkeley DBXML
 - Tamino

Conclusion