



# XML SWOT

"Opening the debate"  
Jouko Salonen  
Republica Oy

13/06/2002

XML Finland 28.5.2002  
XML SWOT, Jouko Salonen  
Republica Oy

## Motivation: to explicate the XML-related arguments in the Finnish citizen-centered e-government discourse

- Sources:

- Julkisten verkkopalvelujen kehittäminen ja XML (JUNA)  
<http://www.intermin.fi/suom/juna/julkaisut/>
- GAO, Electronic Government, Challenges to Effective Adoption of the Extensible Markup Language (XML), April 2002 (United States General Accounting Office)
- Federal Tag Standards for Extensible Markup Language (LMI)
- DON XML Development Guidelines
- Jon Bosak: Ritual XML Reminders
- <http://www.irs.gov>
- Owen Ambur; eXtensible markup language (XML): Greek, Esperanto, Panacea or Snake Oil?  
<http://users.erols.com/ambur/xml.htm>
- Susan Blackmore; The Meme Machine, Oxford University Press, 1999

# Content

- XMLReminders (slides 4-7)
- XML SWOT (slide 8)
- Strength and Opportunity (slides 9,10)
- Weaknesses and Threats (slide 11)
- Distributed control of standardization (slides 12-14)
- XML performance: What is good performance? (slides 15-22)
- Slides: <http://www.republica.fi/xhtml/news.html>

# Reminder: XML is not\* ...

- XML is not a language but a metalanguage
  - It is designed to support the definition of an unlimited number of languages for specific industries and applications
- XML is not semantics, tags have no predefined meaning
  - XML by itself conveys only content and structure, not presentation, behavior or meaning
  - The meaning of XML-languages must be defined outside XML-standard itself
  - Application behaviour can be connected to XML after structure definition (late binding)

# Reminder: XML is\* ...

- XML builds on 30 years of research and 16 years of standardization (ISO 8879, 1986)
- XML is a framework for developing an unlimited number of special-purpose data languages
- XML allows people sharing a common data exchange problem to develop an open solution to that problem; in principle that lets anyone with a similar problem use the same solution
- without interference of third parties, without dependence of large software vendors

# XML is about collaboration\*

- XML provides a standard framework for making agreements about communication
  - Industry DTDs
  - Industry Schemas
  - Industry namespaces
- But it does not make those agreements by itself!

# XML trade-off\*

- XML trades off
  - performance
  - Centralized control
  - Uniformity
- in order to get
  - distributed control
  - asynchronicity
  - persistence (!)
  - a structure obvious both to humans and machines
  - a very low cost of entry



# XML SWOT

<http://www.intermin.fi/suom/juna/julkaisut/>

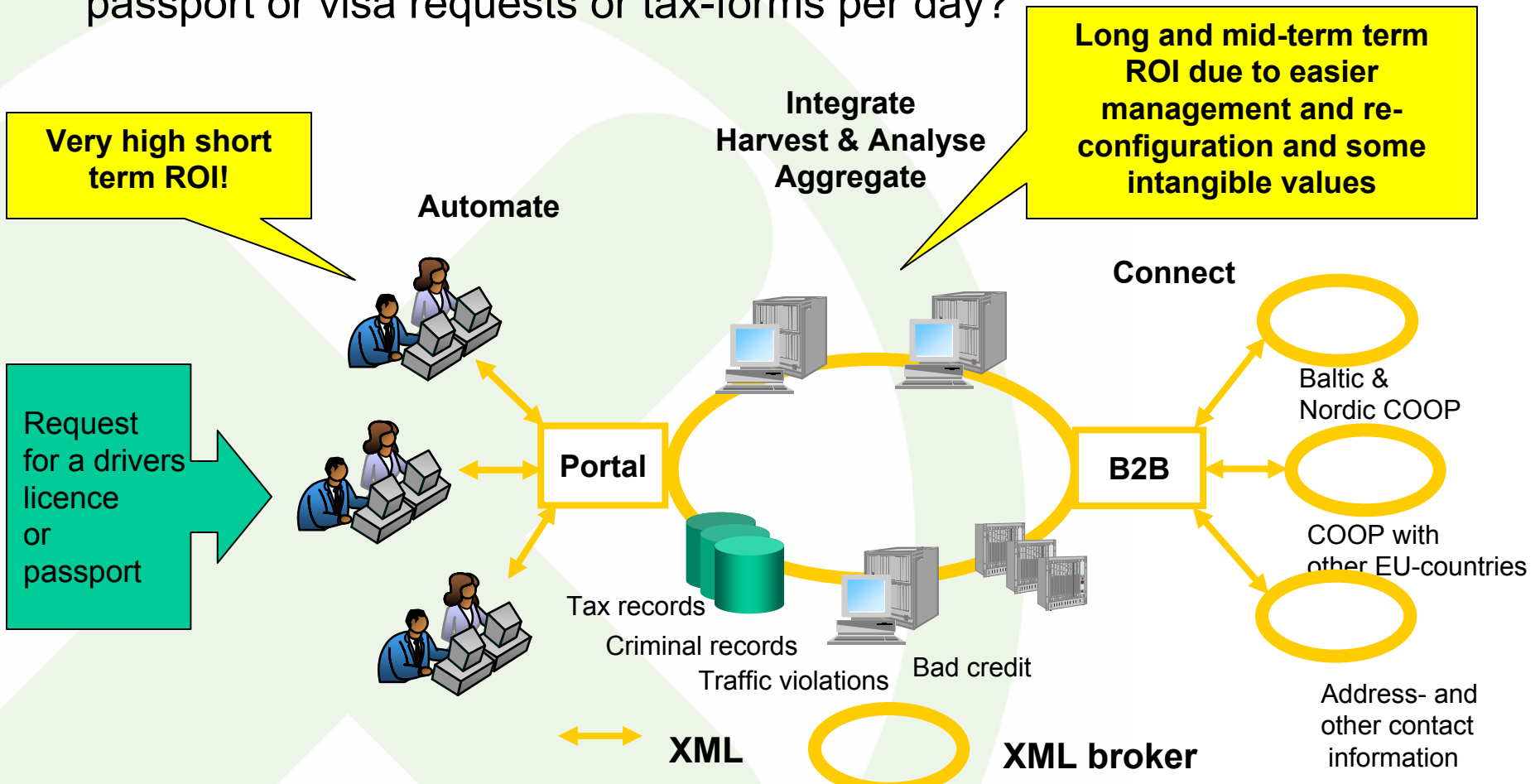
- **Strengths**
  - open standard, platform neutral, self-describing, promotes integration and co-operation, enables knowledge management
- **Weaknesses**
  - dispersed uncontrolled vocabulary-work, uncertainty concerning the consortia-driven standard development, uncertainty due to the immaturity of sub-standards, products and tools not ready, solutions not reliable due to the lack of know-how and experience, poor performance due to the text-based, self-describing tags
- **Opportunity**
  - common ground for information exchange, enables the usage of old legacy data in multi channel services, releases from the vendor lock-in, enables information aggregation and composite services
- **Threats**
  - no agreement of commonly accepted vocabularies, solutions do not support cooperation and re-usability, applications do not really support XML-standard, standardization work becomes fragmented

# Strenght and Opportunity

- Promotes integration and co-operation
- Common ground for information exchange
- Enables the usage of old legacy data in multi channel services
- Releases from the vendor lock-in
- Enables information aggregation and composite services
- Remember XML trade-off! ([Slide 7](#))
- Quotes: Julkisten verkkopalvelujen kehittäminen ja XML (JUNA)  
<http://www.intermin.fi/suom/juna/julkaisut/>

# Strenght & Opportunity

What would it mean if one could process 10X more drivers licence, passport or visa requests or tax-forms per day?



# Weaknesses and Threats

- Dispersed uncontrolled vocabulary-work
- Uncertainty concerning the consortia-driven standard development
- Uncertainty due to the immaturity of sub-standards
- Standardization work becomes fragmented
- No agreement of commonly accepted vocabularies
- Solutions do not support cooperation and re-usability
- Products and tools not ready
- Applications do not really support XML-standard
- Poor performance
- Security issues (NOT IN JUNA-REPORT!)

# Weaknesses and Threats 1.

- Dispersed uncontrolled vocabulary-work
- Uncertainty concerning the consortia-driven standard development
- Uncertainty due to the immaturity of sub-standards
- Standardization work becomes fragmented
- No agreement of commonly accepted vocabularies

# 489 times "Date": No return to centralized uniformity

- General Service Administration (USA) and Logistics Management Institute reviewed different XML tag dictionaries to determine if the federal government can adopt a single standard for establishing XML tags
  - Commerce One xCBL
  - ebXML
  - Oasis
  - RosettaNet
- Together more than 8400 separate name entities
- 489 different expressions for "DATE"
- Remember: More than 400.000 developers are using XMLSPY, a leading XML Schema development tool

Category	Number of entries
Name	357
Date	489
Time	160
Address	41
Agency	12
Contact	206
Telephone	56
Amount	172
Quantity	186

# Fragmented XML Schema work - IRS Form 1120

## 149 Schemas!

```
<xsd:schema targetNamespace="http://www.irs.gov/Schemas/IRSCreditsRelatedToRentalRealEstateActivitiesStatementSchema"
xmlns:xsd="http://www.w3.org/2001/XMLSchema"
xmlns:elems="http://www.irs.gov/Schemas/IRSXMLGeneralElements"
xmlns:types="http://www.irs.gov/Schemas/IRSXMLSchemaTypes"
xmlns="http://www.irs.gov/Schemas/IRSCreditsRelatedToRentalRealEstateActivitiesStatementSchema"
elementFormDefault="qualified" attributeFormDefault="unqualified" version="1.0">
  <xsd:annotation>
    <xsd:documentation xml:lang="en">
      </xsd:documentation>
    </xsd:annotation>

  <!-- Imports for schema types and complex elements -->
  <xsd:import namespace="http://www.irs.gov/Schemas/IRSXMLSchemaTypes" schemaLocation="IRSXMLSchemaTypes.xsd"/>
  <xsd:import namespace="http://www.irs.gov/Schemas/IRSXMLGeneralElements" schemaLocation="IRSXMLGeneralElements.xsd"/>

  <!-- Definition of form type -->
  <xsd:complexType name="CreditsRelatedToRentalRealEstateActivitiesStatementType">
    <xsd:sequence>
      <xsd:element name="Cost" type="ItemType" maxOccurs="unbounded"/>
    </xsd:sequence>
  </xsd:complexType>

  <xsd:complexType name="ItemType">
    <xsd:sequence>
      <xsd:element name="Type"/>
      <xsd:element name="Amount"/>
    </xsd:sequence>
  </xsd:complexType>
</xsd:schema>
```

•Remember: More than 400.000 developers are using XMLSPY, a leading XML Schema development tool

# The truth is: heteroglossia will prevail

- No universal set of TAG-design rules exists
- The corporate consortias do not coordinate with design rules
- However, ebXML chose DeliveryScheduledDate (see right column!) because of ISO11179-5\*
- ISO11179-5
  - Syntactic rules (like "there must be at least one object class and one property term")
  - Semantic rules (like "Oxford English Dictionary spellings must be used")
- ScheduledDeliveryDate (RosettaNet)
- EarliestDeliveryDate (CommerceOne)
- ScheduledShipdate(Oasis)
- DeliveryScheduledDate (ebXML)
- DateDeliveryScheduled

\*ISO11179-5: Information Technology Specification and Standardization of Data Elements, Part 5

Naming and Identification Principles for Data Elements

13/06/2002

XML Finland 28.5.2002

XML SWOT, Jouko Salonen

Republica Oy

# Guidelines and recommendations: fragmented vocabulary-work

- The ebXML suite of technical specifications and related joint UN/EDIFACT work represent the only neutral, horizontal-based, accredited XML business standards development effort and appears, thus, to hold the most promise
- ISO11179 offers a sound basis
- tag-name limited to 4-6 words, a total of 30 letters
- No acronyms or abbreviations
- Initial capital letters
- English dictionary spellings?
- Establish a consistent process, use vertical guidelines
- Establish registry /repository/ data element dictionaries
- Use "translator's approach" instead of "dictator's"
- Do not forget, that this is an evolutionary, memetic process, where we have the role of being a part of the evolutionary choice-making-machine.
- Within an evolution, there is no "wait outside and see" mode. The train is moving, you are there, your choices are: steer yourself or let others do that!

# Application oriented Weaknesses and Threats

- Solutions do not support cooperation and re-usability
- Products and tools not ready
- Applications do not really support XML-standard
- Poor performance
- Security issues (NOT IN JUNA-REPORT!)

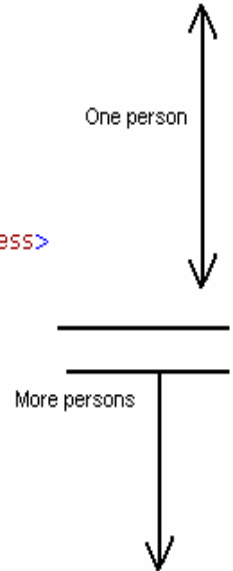
# Application oriented Weaknesses and Threats

- Products and tools not ready
- Applications do not really support XML-standard
- Check "DON XML Development Guidelines" 12/2001 [slide 12](#)
- W3C, Oasis offer validating tools for checking the standard conformance

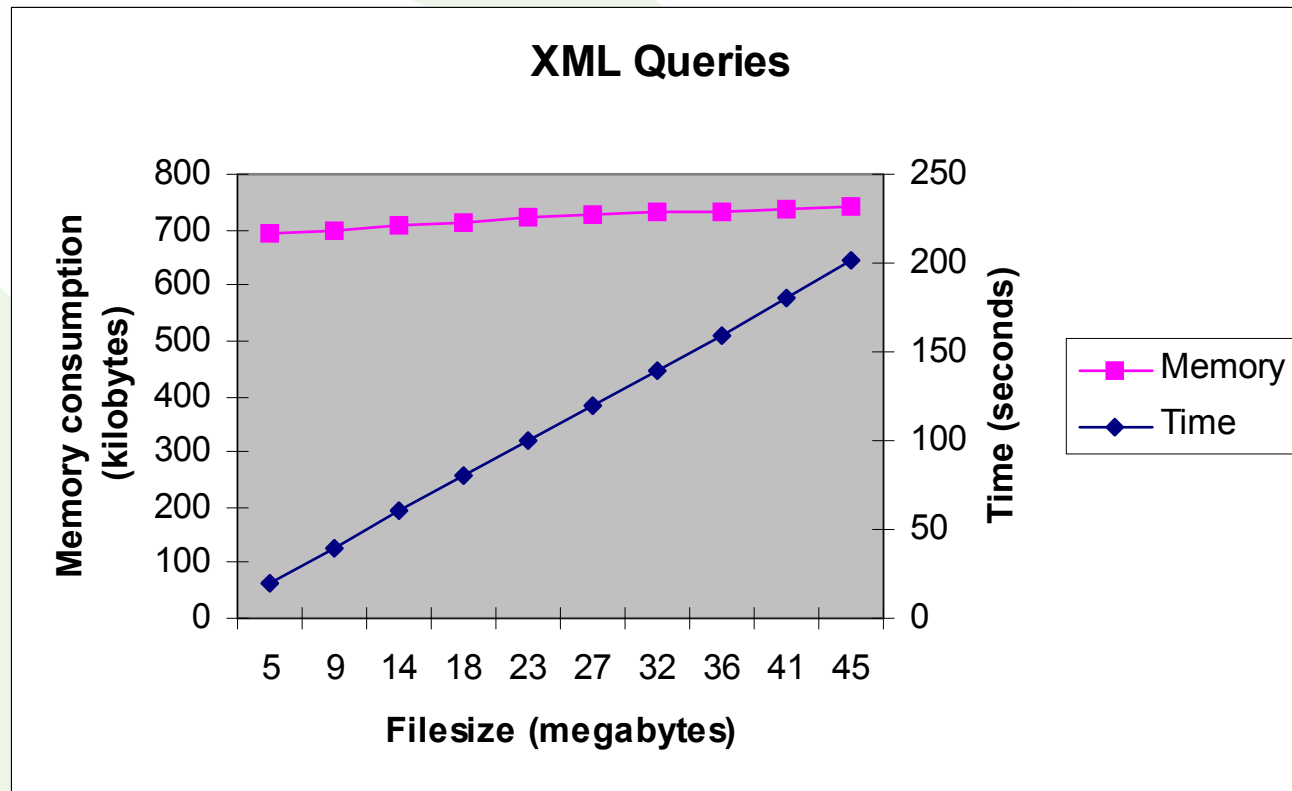
# Performance issue

- **XML Query (Person database)**
  - Searching persons with certain occupation, age and gender
  - Flat, structured, list-type input material enables linear query execution time
  - XML database (e.g. Software AG Tamino) enables logarithmic query execution time

```
<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
- <person:data xmlns:person="http://www.bureau.gov/humaninfo/std.xhtml">
- <person:info>
  <person:first-name>Julie</person:first-name>
  <person:third-name>Karen</person:third-name>
  <person:surname>King</person:surname>
  <person:gender>Female</person:gender>
  <person:date-of-birth>3-12-1957</person:date-of-birth>
  <person:occupation>Accountant</person:occupation>
  <person:address>Beacon Street 171 B 251</person:address>
  <person:zipcode>41682</person:zipcode>
  <person:city>Boise</person:city>
</person:info>
- <person:info>
  <person:first-name>Vincent</person:first-name>
  <person:second-name>Kenneth</person:second-name>
  <person:surname>Scott</person:surname>
  <person:gender>Male</person:gender>
  <person:date-of-birth>7-24-1957</person:date-of-birth>
  <person:occupation>Mechanical Engineer</person:occupation>
```



# XML query performance

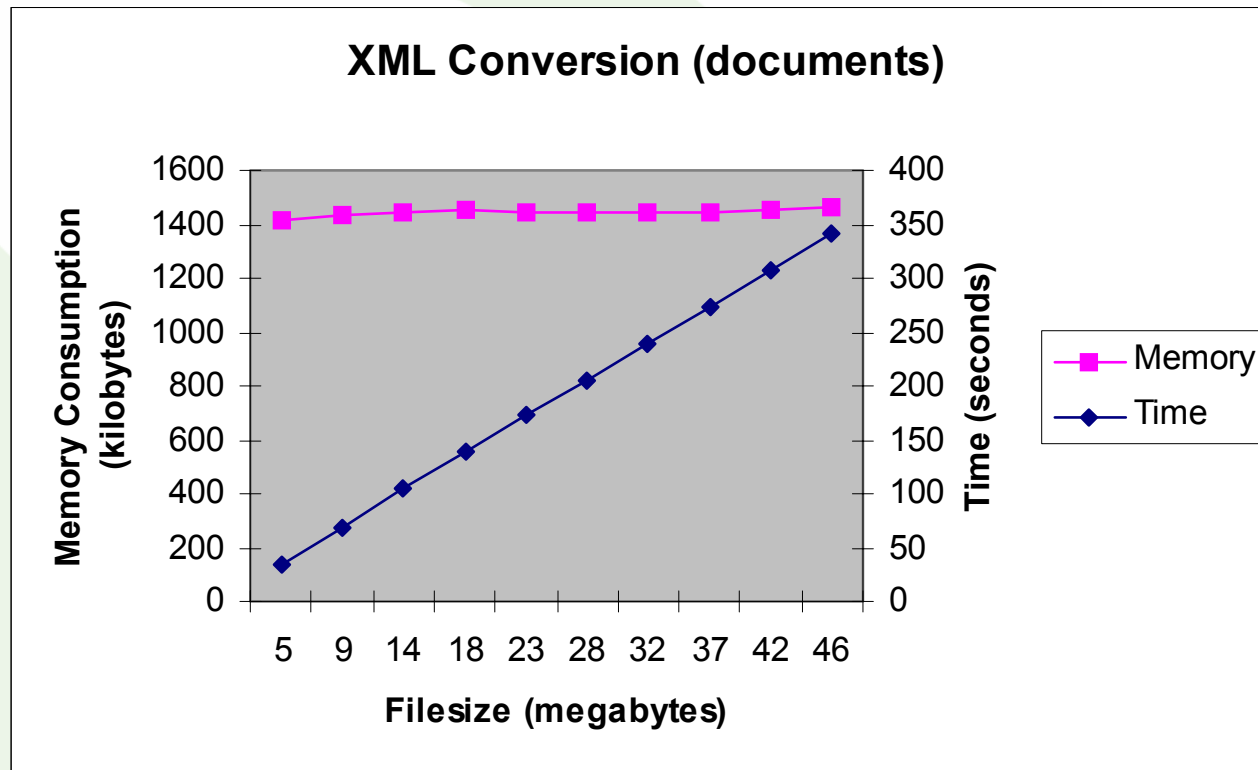


# Performance issue: conversions

- Highly structured, document-like input material
- Conversion XML/HTML
- Streaming mode, memory consumption stays on a constant level
- Time curve is linear: When input size is doubled, conversion time is doubled

```
- <para paratype="normal">  
- <list listtype="unordered">  
- <item>  
- <para paratype="normal">  
  <code codetype="xmlattribute"  
    range="inline">processcomments</code>  
  determines whether a MAP processor  
  should forward XML comments  
  presented in the input XML document  
  to the application. With value  
  <code codetype="xmlattributevalue"  
    range="inline">>true</code>  
  , the comments are forwarded; with
```

# Performance issue: conversions

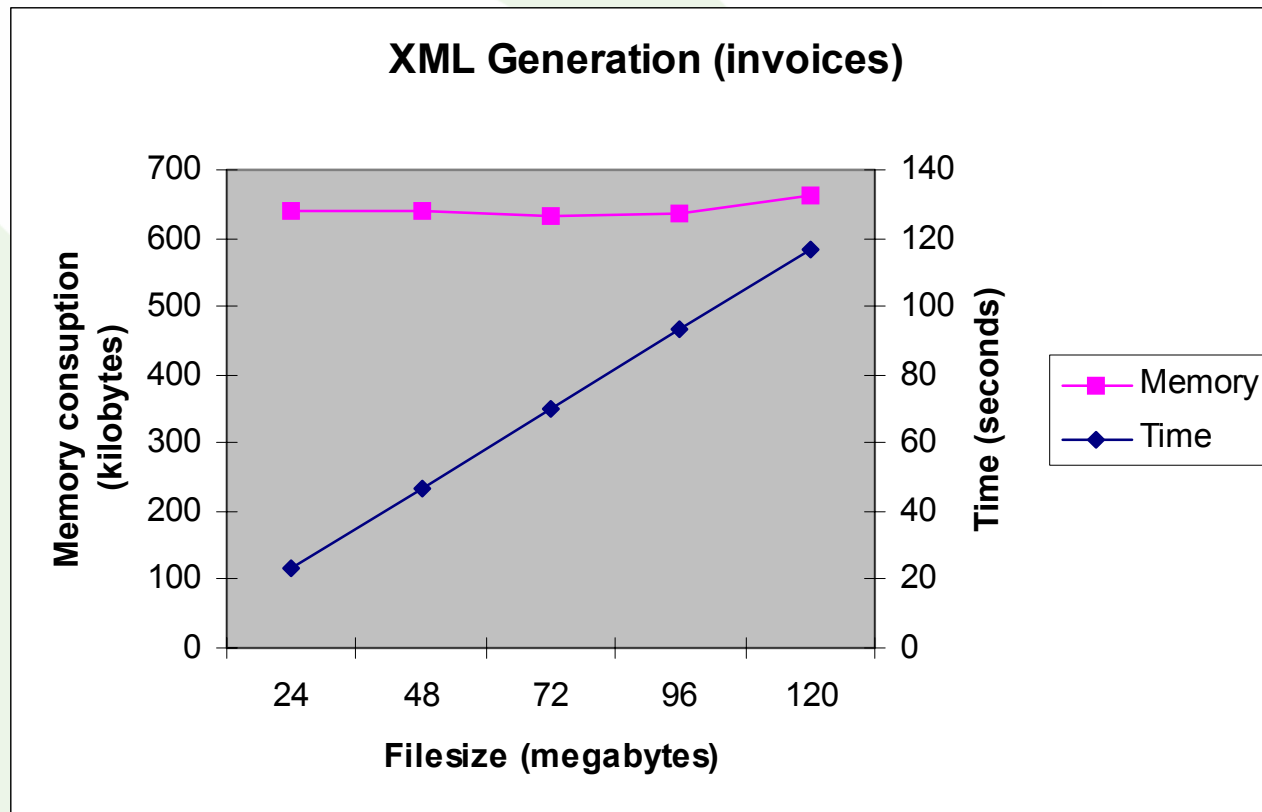


# Performance issue: XML Generation and Serialization (invoices)

- XML Generation and Serialization (invoices)
- Generating XML invoices using random values
- Average structured XML material
- Serializing invoices to hard drive

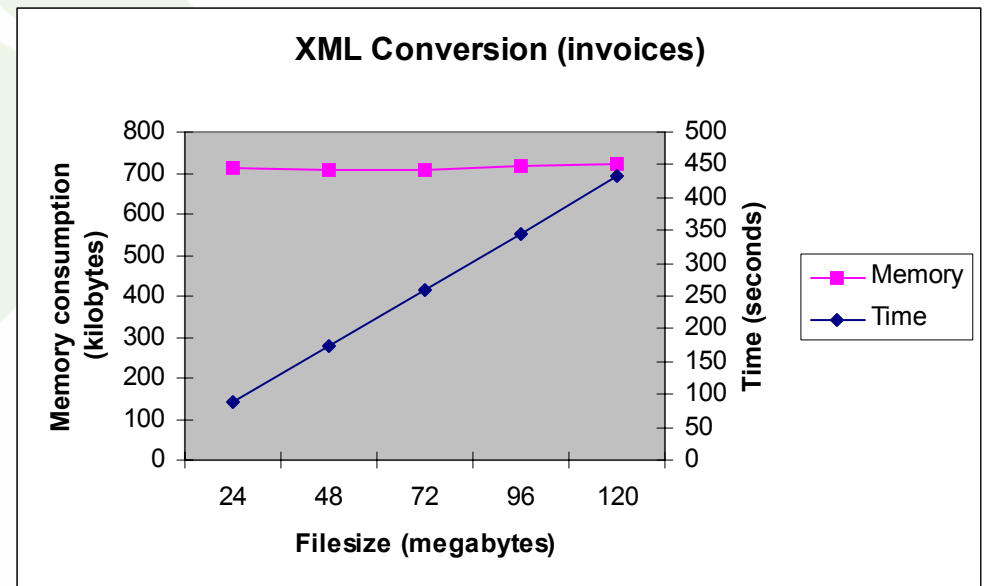
```
- <address>
  <street>Beechwood Drive</street>
  <zipcode>5318</zipcode>
  <city>Altoona</city>
  <state>Delaware</state>
  <country>Colombia</country>
</address>
</customer>
<accountNumber>89-106-641-5374</accountNumber>
<reference>4121-PW-5643</reference>
</header>
<items>
- <item itemid="WWY">
  <name>Winter White Yak</name>
  <price cu="EUR">45.56</price>
  <amount>34</amount>
  <vatprcnt>22.1</vatprcnt>
```

# Performance issue: XML Generation and Serialization (invoices)




## Performance issue: Complex XML Lay-Out conversion into human readable form

- **Complex XML Lay-Out Conversion (invoices)**
- Converting the invoices generated at previous case into human-readable form



# Performance issue: Complex XML Lay-Out conversion into human readable form

	Snow Leopard	41	2345.20	117403.05
	Austrian Ladys Desk	36	37.00	1626.37
<b>Summary</b>				
			<b>Total (EUR)</b>	3302938.82
64 4 22 3 07 24			<b>Expiration</b>	14th April 2002
<b>Invoice</b>				
<b>Reference</b>	1081/71/2140			
<b>Account Number</b>	5150394286			

# Guidelines and recommendations: performance issue

- XML performance is really a practical trade-off issue. Self-describing natural language like tag-format brings you more than you really lose in terms of performance
- As the above examples show, in terms of performance, XML is fully scalable data format
- You get:
  - no vendor lock in
  - open standard component architecture with self-describing, open interfaces
  - re-usable solution that you can circulate
- Jon Bosak (Sun Microsystems): "XML allows people sharing a common data exchange problem, to work out an **open solution** to that problem. Without interference of third parties, without dependence on large software vendors."

# Are there XML-standard related security issues?

- Security is not a separate issue!
- "Element crypting"?
- DRM - "Digital Rights Management"?