Introduction to XML–

How it Works and What it Offers Us

Introduction

- NAACCR & XML
- Basics of the XML format
- What an XML structure offers

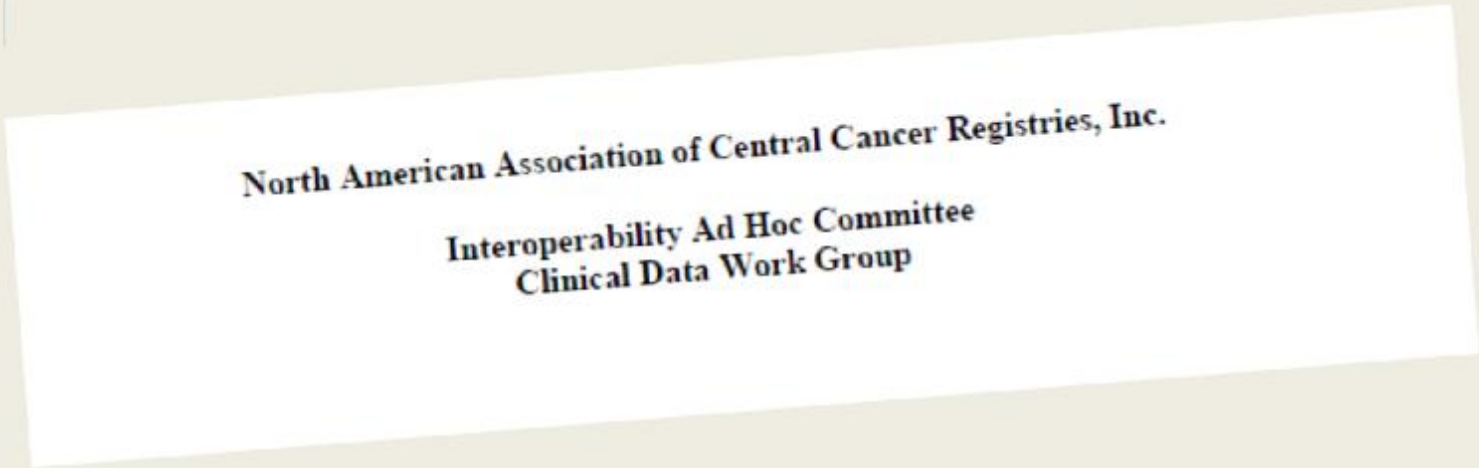
NAACCR & XML

The Flat File – file sizes (full format)

Type A – Full Case Abstract

1994	V3	5350 byte record
2000	V 9	5966 byte record
2004	V 11	6694 byte record
2009	V12	22,824 byte record

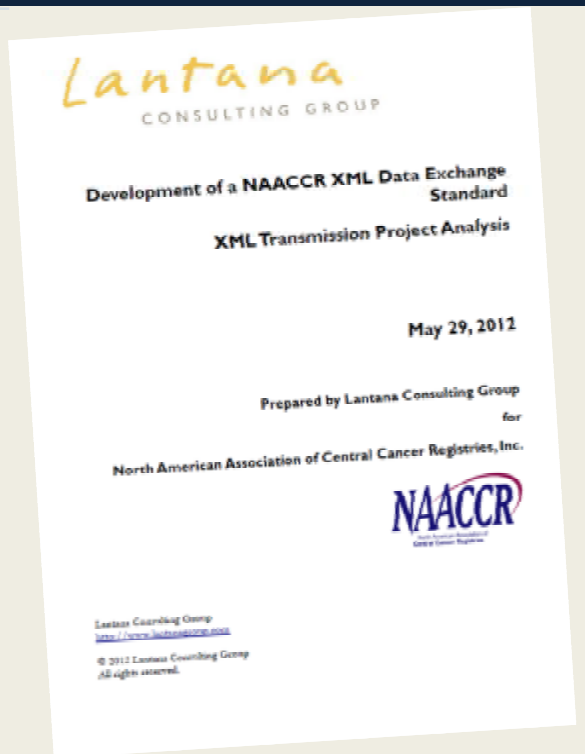
NAACCR & XML



North American Association of Central Cancer Registries, Inc.

Interoperability Ad Hoc Committee
Clinical Data Work Group

NAACCR & XML



Lantana
CONSULTING GROUP

Development of a NAACCR XML Data Exchange Standard

XML Transmission Project Analysis

May 29, 2012

Prepared by Lantana Consulting Group

for

North American Association of Central Cancer Registries, Inc.

NAACCR

Lantana Consulting Group
http://www.lantanagroup.com/

© 2012 Lantana Consulting Group
All rights reserved.

NAACCR XML Implementation Initiative: NAACCR XML Data Transmission Standard for the Cancer Registry Community

Version: September 6, 2011

1 Preface and Introduction

In 2006, NAACCR initiated a project to explore alternative formats for the transmission of the cancer abstract report as presently defined in NAACCR Standards for Cancer Registries, Volume II, Data Standards and Data Dictionary. A work group (WG) was formed, originally the Cancer Abstract Transmission WG under the Information Technology (IT) Committee; now the Clinical Data WG under the Interoperability Ad Hoc Committee. The objective and charge to this WG are noted below.

Objective: Explore alternate mechanisms/messages to transmit the cancer abstract (Volume II).

Charge: This WG will explore alternate mechanisms or messages to transmit and receive the cancer abstract. The investigation will examine not only the steps to transmit and receive data within the cancer registry community, but also consider transmissions to organizations typically outside that community. The current format to transmit this information within the cancer registry community is a column-delimited flat file described in NAACCR Standards for Cancer Registries, Volume II, Data Standards and Data Dictionary.

The Clinical Data WG conducted a pilot implementation using HL7 Clinical Document Architecture (CDA) to transmit cancer data from a hospital cancer registry to a central cancer registry. In December 2009 the WG issued a report to the Board with the recommendation that to explore the use of a NAACCR XML (Extensible Markup Language) structure.

2 Rationale for Change

The WG identified some advantages and disadvantages of moving from the column-delimited format to an alternate format.

2.1 Advantages:

- An alternate message format might produce a smaller transmission file. For example, the format of the NAACCR, NPCR, and SEER Calls for Data contain extensive empty fields.
- Currently, data transmissions over the web are in compressed formats. The compressed formats would not be needed with an alternate format. Web browsers can easily read XML files. An alternate format would be more Web-friendly.
- An alternate message format would be more consistent with software industry standards. In addition to software industry standards, there are related health information technology standards.

DRAFT

NAACCR & XML

2013 NAACCR research plan objectives:

- Develop a new XML-based data standard positioned for easy extension to include emerging future data sources such as claims data, discharge data, and EHR data.
- Accelerate the adoption and implementation of the new data standard by cancer registries.
- Integrate the new standard with emerging research data systems and develop a specific example of rapid case ascertainment for clinical trial patient recruitment with the Kentucky Cancer Registry.

NAACCR & XML

September 2011 – NAACCR Work Group:

6.1 A set of XML format data transmission standards and an implementation guideline.

6.2 A set of web--based tools to register, maintain, and manage NAACCR XML schemas/templates, and NAACCR--specific vocabulary for data transmission in XML format. These tools will be a library of data items, templates, consistency and conformance checks with constraints, and validation rules.

6.3 Combine (components 1 and 2) to form a technology to assist vendors' software with the function in their applications that directly transmits data. The technology will replace the flat--file format, and will be Web--based for senders and receivers.

6.4 Sets of sample data used for the bidirectional translation tests below.

6.5 Bidirectional translation software (flat--file ↔ XML).

6.6 An outline of the elements that must be in the documentation and training manuals for cancer registrars, cancer registry IT staff, and vendors to successfully transition and use XML.

NAACCR & XML

September 2011 – NAACCR Work Group:

6.1 A set of XML format data transmission standards and an implementation guideline.

6.2 A set of web--based tools to register, maintain, and manage NAACCR XML schemas/templates, and NAACCR--specific vocabulary for data transmission in XML format. These tools will be a library of data items, templates, consistency and conformance checks with constraints, and validation rules.

6.3 Combine (components 1 and 2) to form a technology to assist vendors' software with the function in their applications that directly transmits data. The technology will replace the flat--file format, and will be Web--based for senders and receivers.

6.4 Sets of sample data used for the bidirectional translation tests below.

6.5 Bidirectional translation software (flat--file ↔ XML).

6.6 An outline of the elements that must be in the documentation and training manuals for cancer registrars, cancer registry IT staff, and vendors to successfully transition and use XML.

Basics of the XML format

Basics of the XML format

Definition of **MARKUP LANGUAGE**. : a system (as HTML or SGML) for marking or tagging a document that indicates its logical structure (as paragraphs) and gives instructions for its layout on the page especially for electronic transmission and display.

merriam-webster

Basics of the XML format

Starting in late 60’s, popular Markup systems:

- GML (Ibm’s Generalized Markup Language
- SGML (ISO standardized version)
- Tex, LaTeX, Scribe (early document advances)
- HTML, CSS, Docbook …….

Basics of the XML format

Starting in late 60's, popular Markup systems:

- GML (Ibm's Generalized Markup Language
- SGML (ISO standardized version)
- Tex, LaTeX, Scribe (early document advances)
- HTML, CSS, Docbook …….. and **XML**

Basics of the XML format

XML eXtensible Markup Language

Jim Clark- lead of group producing v 1.0 XML spec, circa 1998

The World Wide Web Consortium (W3C)
http://www.w3.org/XML/

Basics of the XML format

List of XML markup languages

From Wikipedia, the free encyclopedia

It has been suggested that *List of XML schemas* be merged into

This is a list of **XML markup languages**.

Contents :A · B · C · D · E · F · G · H · I · J · K · L · M · N · O · P · Q · R · S · T · U · V · W · Y · Z

Lists 170 XML 'languages'

A [edit]

- AdsML Markup language used for interchange of data between advertising systems.
- Agricultural Ontology Service
- AIML Markup language used for creating artificial intelligence chatterbots.
- Attention Profiling Mark-up Language (APML): format for capturing a person's interests and dislikes
- Atom (standard): The Atom *Syndication Format* is a language used for web feeds
- Automated Test Markup Language (ATML): defines a standard exchange medium for sharing information between components of automatic test systems.
- Attention.xml[1]
- aecXML: a mark-up language which uses Industry Foundation Classes to create a vendor-neutral means to access data generated by Building Information Modeling.
- Auto-lead Data Format: an open XML-based standard specifically for communicating consumer purchase requests to automotive dealerships.

B [edit]

- BeerXML: a free XML based data description standard for the exchange of brewing data [1]

Basics of the XML format

List of XML markup languages

From Wikipedia, the free encyclopedia

It has been suggested that *List of XML schemas* be merged into this article. (Discuss) *Proposed since December 2012.*

This is a list of **XML markup languages**.

Contents :A · B · C · D · E · F · G · H · I · J · K · L · M · N · O · P · Q · R · S · T · U · V · W · X · Y · Z

A [edit]

- AdsML Markup language used for interchange of data between advertising systems.
- Agricultural Ontology Service
- AIML Markup language used for creating artificial intelligence chatterbots.
- Attention Profiling Mark-up Language (APML): format for capturing a person's interests and dislikes
- Atom (standard): The Atom *Syndication Format* is a language used for web feeds
- Automated Test Markup Language (ATML): defines a standard exchange medium for sharing inform… …ween components of automatic test systems.
- Attention.xml[1]
- aecXML: a mark-up language which uses Industry Foundation Classes to create a vendor-ne… …eans to access data generated by Building Information Modeling.
- Auto-lead Data Format: an open XML-based standard specifically for communicating co… …r purchase requests to automotive dealerships.

B [edit]

- BeerXML: a free XML based data description standard for the exchange of brewing data [1]

BeerXML (we'll check that later)!

Basics of the XML format

XML Syntax

Basics of the XML format

- building blocks are called elements
  - Elements can contain data, other elements or both
  - Elements can have attributes, which are named and single-valued
- uses tags to define elements within a document
- is human-readable
- contains structured data
- structure comes within data and/or in external schema files

Basics of the XML format

Flat File format vs XML format

AZ804700147289 35.928486 -88.01972090292 20313203

Basics of the XML format

Flat File format vs XML format

AZ804700147289 35.928486 -88.01972090292 20313203

-- VS --

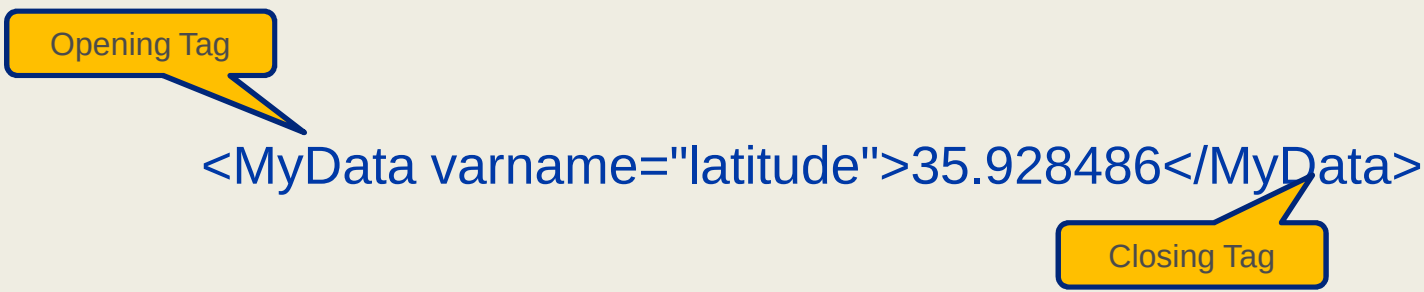
<MyData varname="latitude">35.928486</MyData>
<MyData varname="longitude">-88.019720</MyData>

Basics of the XML format

XML Syntax – sample Element:

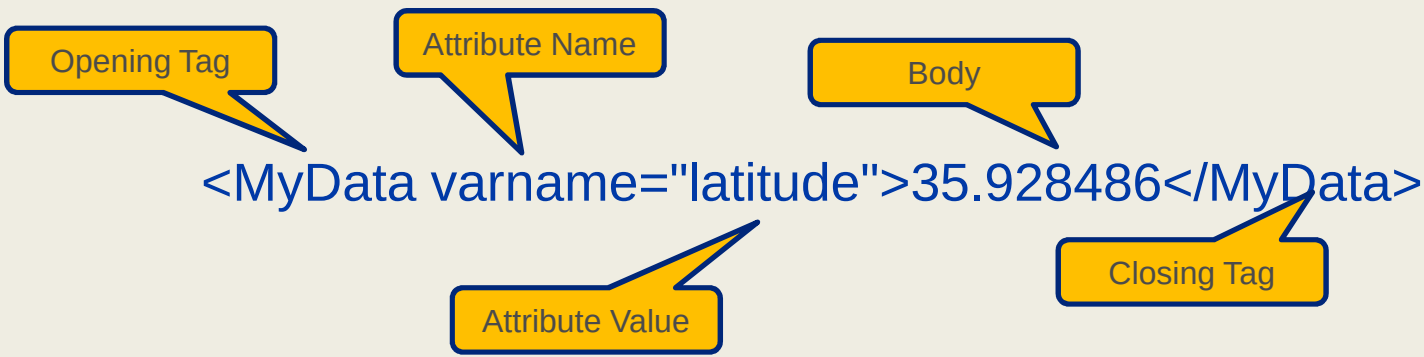
<MyData varname="latitude">35.928486</MyData>

Basics of the XML format



Basics of the XML format

Basics of the XML format



Basics of the XML format

XML – Nesting elements

When an element appears within another element, it is said that the inner element is "nested".

Basics of the XML format

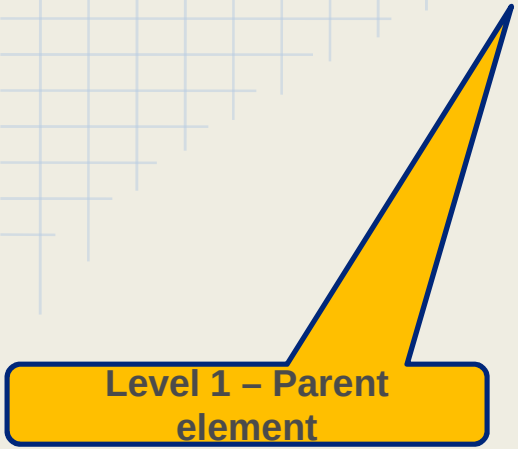
A classically generic example:

Company Ordering System

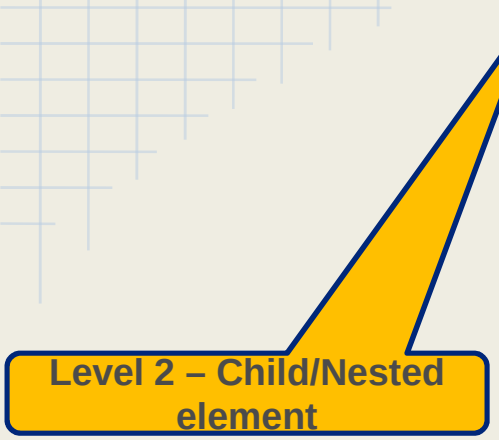
4 Levels of Nesting

```
<CustomerOrders>
  <Customers>
    <CustomerID>ALFKI</CustomerID>
    <Orders>
      <OrderID>10643</OrderID>
      <OrderDate>1997-08-25</OrderDate>
    </Orders>
    <Orders>
      <OrderID>10692</OrderID>
      <OrderDate>1997-10-03</OrderDate>
    </Orders>
    <CompanyName>Alfreds Futterkiste</CompanyName>
  </Customers>
  <Customers>
    <CustomerID>ANATR</CustomerID>
    <Orders>
      <OrderID>10308</OrderID>
      <OrderDate>1996-09-18</OrderDate>
    </Orders>
    <CompanyName>Ana Trujillo Emparedados y helados</CompanyName>
  </Customers>
</CustomerOrders>
```

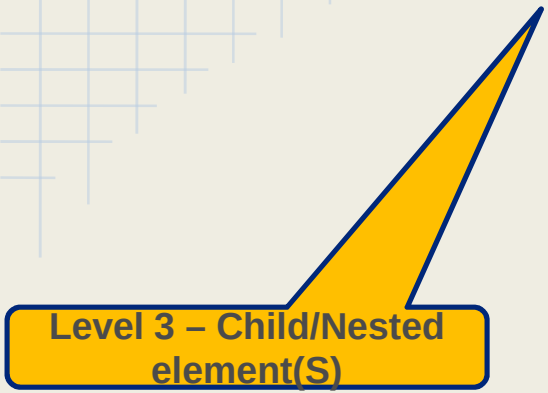
```
<CustomerOrders>
  <Customers>
    <CustomerID>ALFKI</CustomerID>
    <Orders>
      <OrderID>10643</OrderID>
      <OrderDate>1997-08-25</OrderDate>
    </Orders>
    <Orders>
      <OrderID>10692</OrderID>
      <OrderDate>1997-10-03</OrderDate>
    </Orders>
    <CompanyName>Alfreds Futterkiste</CompanyName>
  </Customers>
  <Customers>
    <CustomerID>ANATR</CustomerID>
    <Orders>
      <OrderID>10308</OrderID>
      <OrderDate>1996-09-18</OrderDate>
    </Orders>
    <CompanyName>Ana Trujillo Emparedados y helados</CompanyName>
  </Customers>
</CustomerOrders>
```



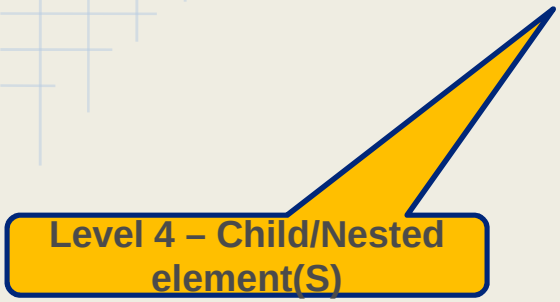

```
<CustomerOrders>
  <Customers>
    <CustomerID>ALFKI</CustomerID>
    <Orders>
      <OrderID>10643</OrderID>
      <OrderDate>1997-08-25</OrderDate>
    </Orders>
    <Orders>
      <OrderID>10692</OrderID>
      <OrderDate>1997-10-03</OrderDate>
    </Orders>
    <CompanyName>Alfreds Futterkiste</CompanyName>
  </Customers>
  <Customers>
    <CustomerID>ANATR</CustomerID>
    <Orders>
      <OrderID>10308</OrderID>
      <OrderDate>1996-09-18</OrderDate>
    </Orders>
    <CompanyName>Ana Trujillo Emparedados y helados</CompanyName>
  </Customers>
</CustomerOrders>
```




```
<CustomerOrders>
  <Customers>
    <CustomerID>ALFKI</CustomerID>
    <Orders>
      <OrderID>10643</OrderID>
      <OrderDate>1997-08-25</OrderDate>
    </Orders>
    <Orders>
      <OrderID>10692</OrderID>
      <OrderDate>1997-10-03</OrderDate>
    </Orders>
    <CompanyName>Alfreds Futterkiste</CompanyName>
  </Customers>
  <Customers>
    <CustomerID>ANATR</CustomerID>
    <Orders>
      <OrderID>10308</OrderID>
      <OrderDate>1996-09-18</OrderDate>
    </Orders>
    <CompanyName>Ana Trujillo Emparedados y helados</CompanyName>
  </Customers>
</CustomerOrders>
```




```
<CustomerOrders>
  <Customers>
    <CustomerID>ALFKI</CustomerID>
    <Orders>
      <OrderID>10643</OrderID>
      <OrderDate>1997-08-25</OrderDate>
    </Orders>
    <Orders>
      <OrderID>10692</OrderID>
      <OrderDate>1997-10-03</OrderDate>
    </Orders>
    <CompanyName>Alfreds Futterkiste</CompanyName>
  </Customers>
  <Customers>
    <CustomerID>ANATR</CustomerID>
    <Orders>
      <OrderID>10308</OrderID>
      <OrderDate>1996-09-18</OrderDate>
    </Orders>
    <CompanyName>Ana Trujillo Emparedados y
helados</CompanyName>
  </Customers>
</CustomerOrders>
```




```
<CustomerOrders>
  <Customers>
    <CustomerID>ALFKI</CustomerID>
    <Orders>
      <OrderID>10643</OrderID>
      <OrderDate>1997-08-25</OrderDate>
    </Orders>
    <Orders>
      <OrderID>10692</OrderID>
      <OrderDate>1997-10-03</OrderDate>
    </Orders>
    <CompanyName>Alfreds Futterkiste</CompanyName>
  </Customers>
  <Customers>
    <CustomerID>ANATR</CustomerID>
    <Orders>
      <OrderID>10308</OrderID>
      <OrderDate>1996-09-18</OrderDate>
    </Orders>
    <CompanyName>Ana Trujillo Emparedados y
helados</CompanyName>
  </Customers>
</CustomerOrders>
```

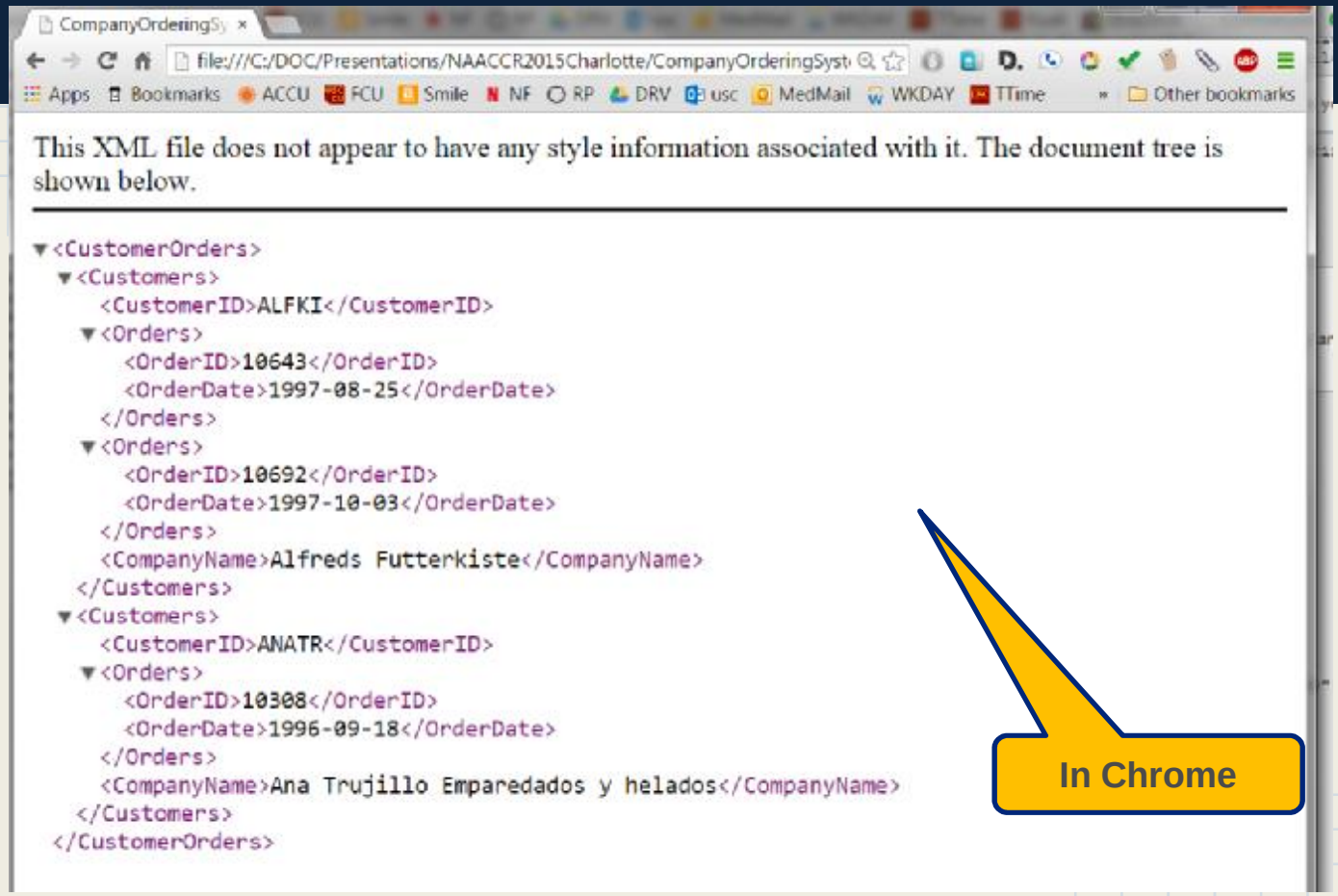
Basics of the XML format

XML – Nesting elements

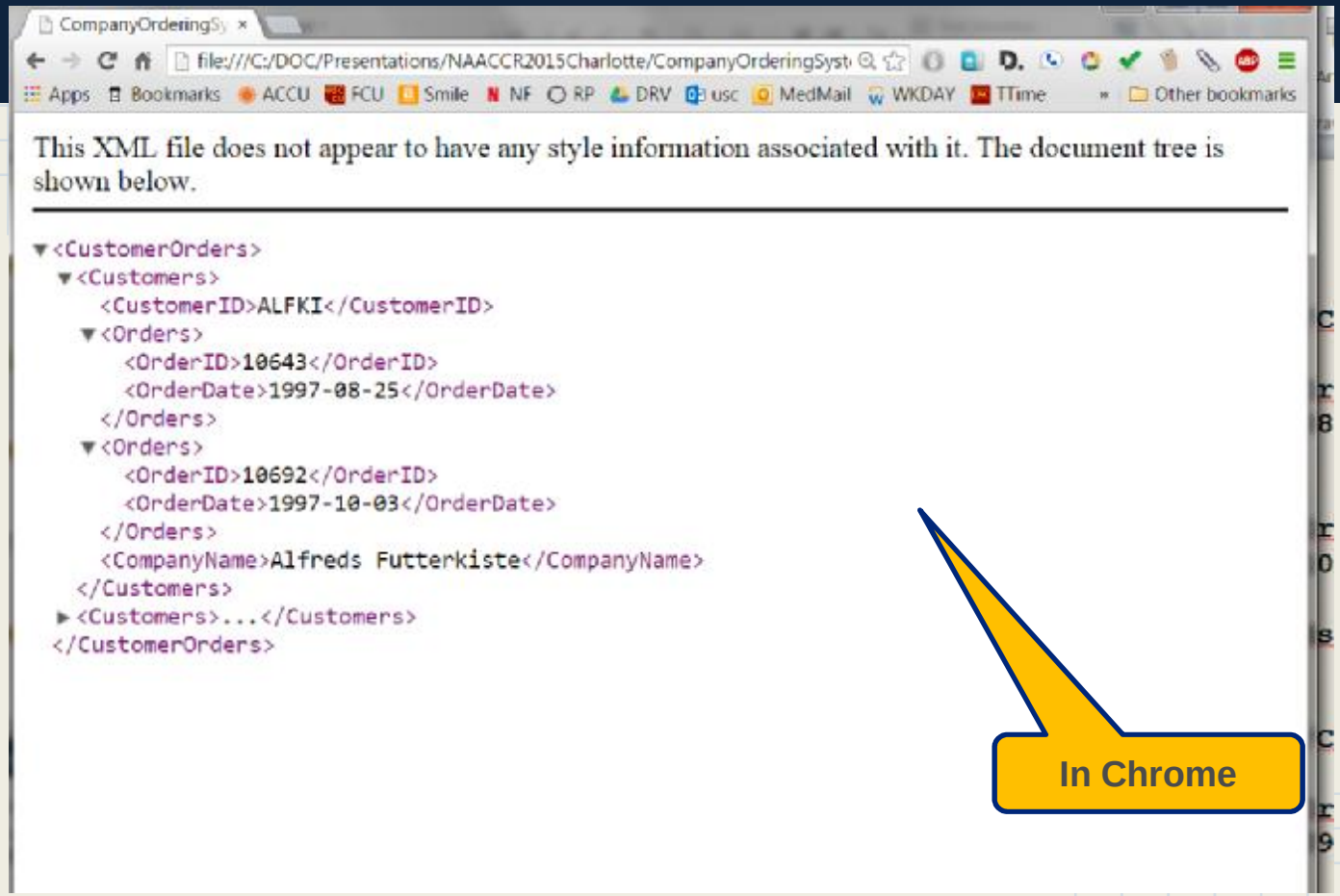
The above Nesting is equally valid like this – no line breaks or spacing necessary:

<CustomerOrders><Customers><CustomerID>ALFKI</CustomerID><Orders><OrderID>10643</OrderID><CustomerID>ALFKI</CustomerID><OrderDate>1997-08-25</OrderDate></Orders><Orders><OrderID>10692</OrderID><CustomerID>ALFKI</CustomerID><OrderDate>1997-10-03</OrderDate></Orders><CompanyName>Alfreds Futterkiste</CompanyName></Customers><Customers><CustomerID>ANATR</CustomerID><Orders><OrderID>10308</OrderID><CustomerID>ANATR</CustomerID><OrderDate>1996-09-18</OrderDate></Orders><CompanyName>Ana Trujillo Emparedados y helados</CompanyName></Customers></CustomerOrders>

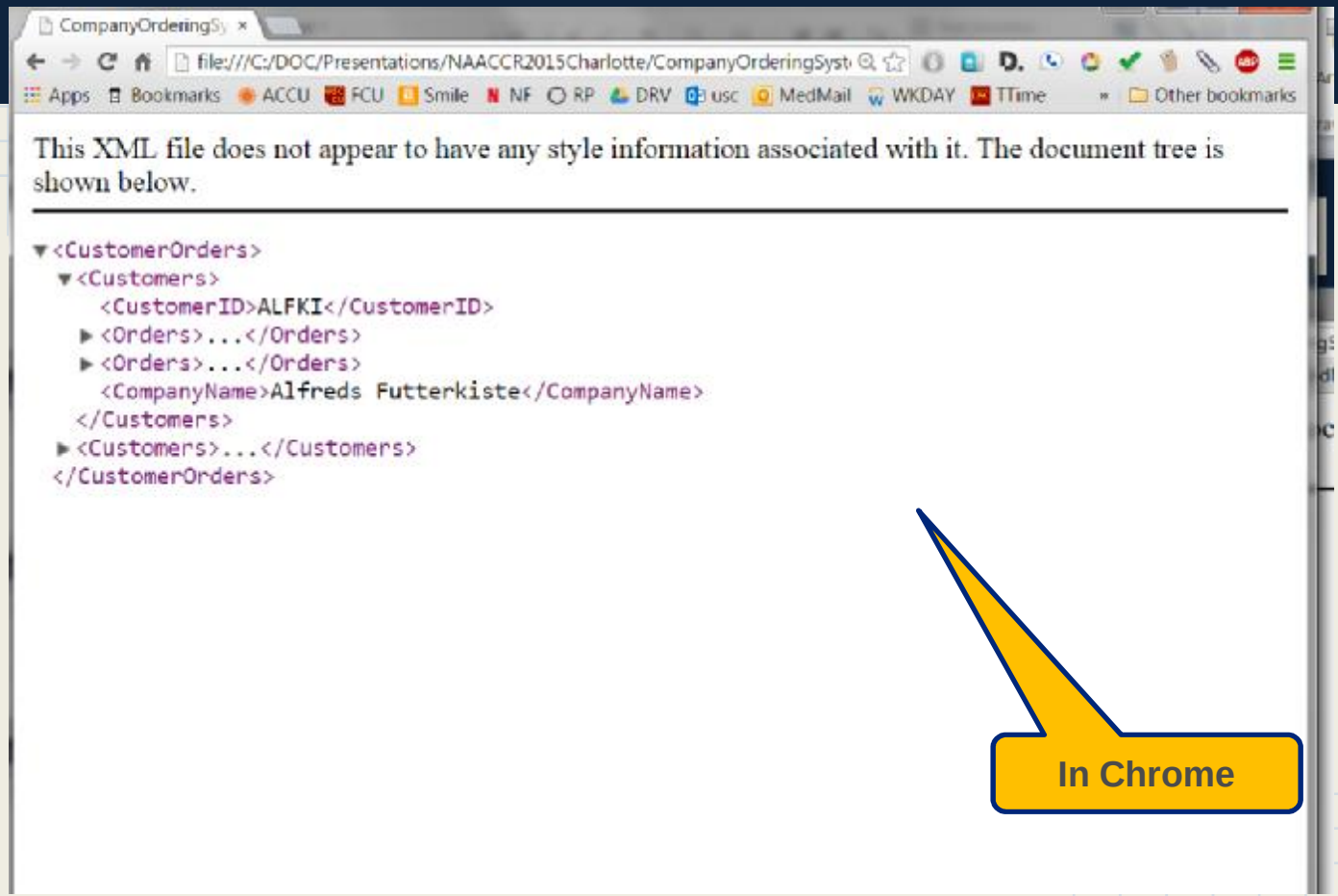
Basics of the XML format



Basics of the XML format



Basics of the XML format



What an XML structure offers

What an XML structure offers

- Benefits of the XML framework

What an XML structure offers

XML : provides an open platform to create and EXTEND data specifications

XML : is a W3C standard

XML : superior to text files of large flat records

XML : does not require the transmission of unused items

XML : permits data files that are easier to interpret for both humans and computers

XML : can contain any possible data type

XML : permits definition and validation OPTIONS, ie using Schemas (XSD's)

XML : has several high level tools for selecting and reformatting records, without intervention of a programmer (ie XPath)

What an XML structure offers

Going forward:

- Totally rethink our record structure:
  Nesting - permits multiple malignancies to be represented without repetition of patient data (more on this next!)

- Build our XML framework to integrate with other data sources

XML
Charlotte
NAACCR 2015

Thanks

- To all the members of our XML Task Force

Rich Pinder
USC School of Medicine
323) 442-1640
rpinder@usc.edu