XML Introduction

Gerry Coleman
WI Dept of Justice

coleman@doj.state.wi.us
XML and Bovine Geometry

Cattle-ized Biometric Identification of Fertility Features (CBIFF)
Image Quality
Facial Recognition

120 Pixels
Minutiae Templates

- Island
- Ridge Ending
- Dot
- Horizontal Reference

\[ \theta \]
What is XML?

- eXtensible Markup Language
- Generic format for tagging data
- World Wide Web Consortium (W3C) specification
- Tutorial: [www.w3Schools.com/xml](http://www.w3Schools.com/xml)
- A text format
- Related:
  - Stylesheet Transformations
  - Schema
Transmission format:
Now is the <b>time</b> for all good men to come to.

Presentation format:
Now is the <b>time</b> for all good men to come to.
XML

Like HTML, the tags are symmetrical . . .

<open> DATA </open>

. . . but the tag names are user defined.
Tags can be nested…

<verbs>
  <open>DATA</open>
  <close>DATA</close>
</verbs>

…well-formed XML
Officially…

- The design goals for XML are:
  - XML shall be straightforwardly usable over the Internet.
  - XML shall support a wide variety of applications.
  - XML shall be compatible with SGML.
  - It shall be easy to write programs which process XML documents.
  - The number of optional features in XML is to be kept to the absolute minimum, ideally zero.
  - XML documents should be human-legible and reasonably clear.
  - The XML design should be prepared quickly.
  - The design of XML shall be formal and concise.
  - XML documents shall be easy to create.
  - Terseness in XML markup is of minimal importance.
For the ANSI/NIST fingerprint standard, XML would replace:

- **Field tag names**
  - 2.001 would become `<Length>`

- **Separators**
  - `FS, GS, RS, US` items would be delimited by tag names and angle brackets. `< >`
Why XML?

• XML is “built in”:
  – Databases
  – Web Servers
  – Browsers
  – Application development software

• Applications don’t have to handle:
  – Parsing
  – Validation
  – Translation to presentation format
Why XML?

• It is the currency of web services and SOAP communication.
• XML-appliances are available to optimize security and performance.
• A growing infrastructure of tools (DOM, SAX, XPath, XML Query).
• XML is open, interoperable across vendor platforms.
• XML has archival “readability”
Why not XML (yet)?

• XML is a **text**-formatting convention, and not optimized today for handling binary objects.

• *WebServices* is the principal mechanism for carrying XML payloads.
  - *The SOAP envelope, security and guaranteed delivery have yet to mature*
Your programmers
can do anything. . .

. . . but you lose the power of the tools to parse and validate data.
Proposals for Binary Data

- Embedding Base-64 coded data
- Referencing (URI)
- SOAP with Attachments (SwA)
- WebService Attachments (WS-attachments)
- Multipurpose Internet Mail Extensions (MIME)
- Direct Internet Message Encapsulation (DIME)
- XML-binary Optimized Packaging (XOP)
- Message Transmission Optimization Mechanism (MTOM)
Joint Task Force on Rapsheet Standardization

- FBI, SEARCH, State Criminal History Repositories, CJIS Advisory Policy Board, NLETS
- Created a rapsheet transmission standard using tagged data, like the ANSI/NIST fingerprint transmission standard
- Created an XML transmission standard conforming to the Global Justice XML Data model
JTF Findings

• Rapsheet presentation formats across 50 states were all different.
• XML and XSL can bring order to diversity.
• NCIC and IAFIS data sets and codes were not compatible.
• Code values are meaningless to the civilian users of criminal history records.
<PalmImageTypeCode>
45
</PalmImageTypeCode>

<PalmImageType>
Right Carpal Roll
</PalmImageType>
JTF Recommendations

• The FBI should develop a Type-2 Record specification using GJXDM-conformant XML.

• An XML version of the entire structure of the fingerprint transmission package should be developed.
Example

<FingerprintTransmissionPackage>
  <RecordType01TransactionRecord>
    <Length>305</Length>
    <Version>0900</Version>
    <Content>…</Content>
  </RecordType01TransactionRecord>
  <RecordType02DescriptiveText>
    <Length>813</Length>
    <ImageDesignationCharacter>00</ImageDesignationCharacter>
  </RecordType02DescriptiveText>
  <RecordType14VariableResolutionTenPrintRecord>
    <Length>813</Length>
    <ImageDesignationCharacter>01</ImageDesignationCharacter>
    <CompressionAlgorithm>WSQ20</CompressionAlgorithm>
    <ImageBase64>8vnsI3ULdnsU=</ImageBase64>
  </RecordType14VariableResolutionTenPrintRecord>
</FingerprintTransmissionPackage>
edi meets xml

• Information site:
  • http://www.doj.state.wi.us/les