Efficient XML and Weather: Efficient XML Report 1.0

Aaron Braeckel

Briefing to WXXM/AIXM Conference
04 May 2010
National Center for Atmospheric Research
Boulder, CO

XML Benefits:

- Extensible
- Human-readable
- Self-describing
- Hardware, software, platform-independent
- Expressive data model (trees, graphs, etc.)
- Validatable
- Namespaced
- …

XML Issues:

One of the difficult weather problems is data volume

Processing efficiency
- Encoding time
- Decoding time
- Message throughput
- Battery life (handheld devices, etc)

Compactness
- Storage
- Archival
- Bandwidth/transmission
  - Has a (non-trivial) recurring cost in some cases
  - Next-generation aircraft
  - Handheld devices

3

Processing and Compactness:

a

XML Solution Classes:

Data-agnostic Compression (GZIP, ZIP, TAR, etc)
- Worse processing efficiency
- Does not address several fundamental XML issues
- Documents must be decompressed before being operated upon (e.g., XSLT)

Hardware/Appliances

XML-Wrapped Binary
- Use XML for "metadata" or other portions, binary for data
- Binary portions are "opaque", no XML compatibility

Binary XML Formats
- Non-textual encodings of the XML data model
- Human readability lost

5

Background and Motivation

Prior Work:

DoD Comparison (binary to XML) [1]
- 10x, 17x, 22x, and 121x larger with XML

W3C XML Binary Characterization Working Group
- Collected requirements

W3C Efficient XML Interchange Working Group
- Performed measurements
- Made recommendations
- Produced a measurement framework

MIT LL Sun's Fast Infoset and AgileDelta's EfficientXML Weather Comparison
- EfficientXML provided better compactness (83% vs 75%)
- Sun's Fast Infoset provided better processing (86ms vs 207ms)

NCAR Preliminary Efficient XML Weather Assessment
- Compared Fast Infoset vs normal XML representations
- Fast Infoset 75% of XML size
- Fast Infoset 33% of parsing time

6

[1] Efficient XML – Taking Net-Centric Operations to the Edge. John Schneider

W3C Binary Characterization:

- Collected use cases
- Collected desirable properties of efficient XML approaches
- Listed a minimum set of properties for a recommended solution

"Binary XML is needed" [2]

W3C Efficient XML Interchange:

- Collected use cases
- Created a measurement framework
- Collected a varied set of XML files for many different use cases
- Measured several different solutions
- Recommended a solution
- Produced a specification (EXI)

"the results indicate it is possible to achieve substantial gains over XML… in a wide variety of use cases" [3]

[2] http://www.w3.org/TR/exi-measurements/#conclusions
[3] http://www.w3.org/TR/xbc-characterization/#N107D4

Background and Motivation

W3C Binary Characterization Working Group:

Desirable Properties:
- Directly Readable & Writable
- Transport Independence
- Compactness
- Human Language Neutral
- Platform Neutrality
- Integratable into XML Stack
- Royalty Free
- Fragmentable
- Streamable
- Roundtrip Support
- Generality
- Schema Extensions and Deviations
- Format Version Identifier
- Content Type Management
- Self-Contained

Background and Motivation

Potential Binary XML Solutions:

Data Format	Standards Bodies	W3C EXI Characteristics	Notes
Fast Infoset	ITU-T ISO	Was not considered to satisfy: •Compactness •Generality	
EXI	W3C*	Meets all characteristics	W3C Candidate Recommendation
BiM	ISO (MPEG WG)	Not Measured	
BXML	OGC*	Not Measured	OGC Best Practice, not a standard
WBXML	Open Mobile Alliance W3C*	Not Measured	Proposed W3C Standard (inactive?)

Other formats/approaches (such as XMill) not considered for lack of standardization support

9

Efficient XML Report 1.0
Compactness Analysis

Analysis:

Compare:

- Legacy binary formats
- ADDS Dataserver XML
    - EXI
    - Fast Infoset
    - GZIP
- WXXM 1.1.1 XML
    - EXI
    - Fast Infoset
    - GZIP

Environment:
Based on W3C EXI Test Framework
•JAPEX (micro-benchmarking library)

10

Efficient XML Report 1.0
Compactness Analysis

Data:

Products (single record **and** 24 hours of records):
- Aircraft reports (31057 records for a 24 hour period)
- AIR/SIGMET reports (317 records for a 24 hour period)
- METAR reports (160761 records for a 24 hour period)
- TAF reports (24783 records for a 24 hour period)

Data was retrieved from the ADDS Dataserver then converted into WXXM, legacy binary, and other downstream equivalents

WXXM Conversion:
- Not perfect…
- Introduced some uncertainty

Efficient XML and Weather

Efficient XML Report 1.0
Compactness Analysis

ADDS Dataserver Example:

```xml
<time_taken_ms>13</time_taken_ms>
<data num_results="1">
  <AIRSIGMET>
    <raw_text>WSUS33 KKCI 100555 SIGW CONVECTIVE SIGMET...NONE OUTLOOK
      VALID 100755-101155 TS ARE NOT EXPD TO REQUIRE WST ISSUANCES.</raw_text>
    <valid_time_from>2010-03-10T05:55:00Z</valid_time_from>
    <valid_time_to>2010-03-10T07:55:00Z</valid_time_to>
    <hazard type="CONVECTIVE" severity="NONE" />
    <airsigmet_type>SIGMET</airsigmet_type>
    <area num_points="36">
      <point>
        <longitude>-127.03</longitude>
        <latitude>48.41</latitude>
      </point>
      <point>
        <longitude>-128.01</longitude>
        <latitude>48.33</latitude>
      </point>
      <point>
        <longitude>-127.49</longitude>
        <latitude>47.48</latitude>
      </point>
      <point>
        <longitude>-127.01</longitude>
        <latitude>46.0</latitude>
      </point>
      <!-- 33 more points... -->
    </area>
  </AIRSIGMET>
</data>
</response>
```
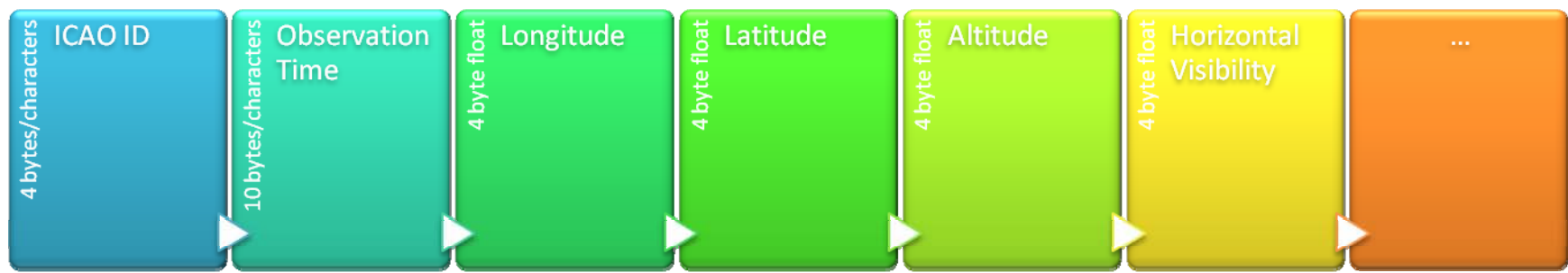
12

Efficient XML Report 1.0
Compactness Analysis

WXXM Example:

```xml
<ns3:AreaReport xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:type="ns3:SIGMETType">
    <ns1:boundedBy xsi:nil="true"/>
    <ns3:rawText>WSUS33 KKCI 100555 SIGW  CONVECTIVE SIGMET...NONE  OUTLOOK VALID...</ns3:rawText>
    <ns3:airspaceAreaForecast>
        <ns4:Forecast>
            <ns1:boundedBy xsi:nil="true"/>
            <ns8:samplingTime ns2:type="simple">
                <ns1:TimePeriod>
                    <ns1:beginPosition>2010-03-10T05:55:00Z</ns1:beginPosition>
                    <ns1:endPosition>2010-03-10T07:55:00Z</ns1:endPosition>
                </ns1:TimePeriod>
            </ns8:samplingTime>
            <ns8:featureOfInterest ns2:type="simple">
                <ns3:Airspace>
                    <ns1:boundedBy xsi:nil="true"/>
                    <ns1:location ns2:type="simple">
                        <ns1:Polygon srsName="urn:ogc:def:crs:EPSG::4326">
                          <ns1:exterior>
                            <ns1:LinearRing>
                              <ns1:posList>-122.7699966430664 49.0 ...</ns1:posList>
                            </ns1:LinearRing>
                          </ns1:exterior>
                        </ns1:Polygon>
                    </ns1:location>
                </ns3:Airspace>
            </ns8:featureOfInterest>
            <ns8:result xsi:type="ns3:AirspaceWxType">
                <ns1:boundedBy xsi:nil="true"/>
                <ns3:convection>
                    <ns3:Convection>
                        <ns1:boundedBy xsi:nil="true"/>
                        <ns4:intensity>MODERATE</ns4:intensity>
```
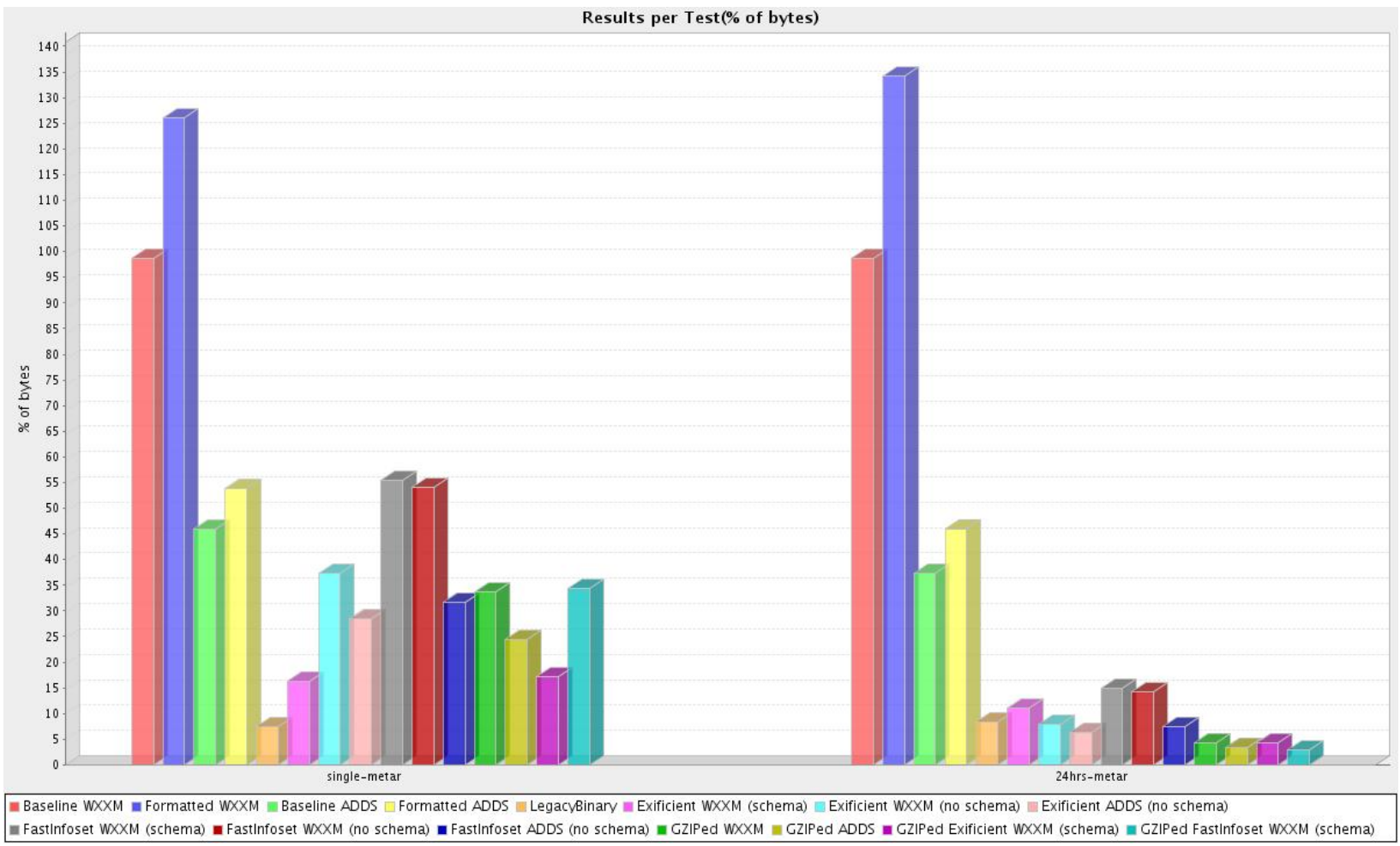
13

Efficient XML Report 1.0
Compactness Analysis

Legacy Binary:

ICAO ID	Observation Time	Longitude	Latitude	Altitude	Horizontal Visibility	...
4 bytes/characters	10 bytes/characters	4 byte float	4 byte float	4 byte float	4 byte float	

14

Efficient XML Report 1.0
Compactness Analysis

Result Sample (METARs):

15

Efficient XML Report 1.0
Compactness Analysis

Results (All Products – 24 hours of records):

Data Format	Average Compaction (relative to baseline WXXM)
Formatted WXXM	1.36
Baseline ADDS	0.38
Exificient WXXM (with schema)	0.11
Exificient WXXM (without schema)	0.08
Exificient ADDS (without schema)	0.06
Sun's Fast Infoset WXXM (with schema)	0.15
Sun's Fast Infoset WXXM (without schema)	0.14
Sun's Fast Infoset ADDS (without schema)	0.08
GZIP WXXM	0.04
GZIP ADDS	0.03
GZIP Exificient WXXM (with schema)	0.04
GZIP Sun's Fast Infoset WXXM (without schema)	0.03
Legacy Binary (METARs and TAFs only)	0.15

16

Future Work

Efficient XML Report 2.0+:

Processing Analysis (transactions per second)
- Encode/decode

Network Roundtrips (transactions per second)
- Two machines
- Fat/skinny pipe test runs
- Combined compaction/processing analysis

Expanded Compaction Analysis?
- Commercial libraries
- With/without schema
- BiM?

Solutions

Recommendations:

EXI format
- Promising industry standard for efficient XML transmission*
- Very favorable compactness
- Open source and commercial libraries
  - Exificient (still evolving)
  - AgileDelta's Efficient XML
  - …

  Advantages:
  - XML flexibility
  - XML tooling support (when developed)
  - Increased efficiency
  - Lossless conversion to/from XML

Use event-based (SAX or StAX) parsing techniques
- DOM has a non-trivial overhead in some cases (such as memory usage)

Further combined processing/compactness analysis required

Efficient XML and Weather

References:

Efficient XML Report 1.0 -
https://wiki.ucar.edu/display/NNEWD/Public+Documentation

W3C Binary Characterization Working Group –
http://www.w3.org/XML/Binary/

W3C Efficient XML Interchange Working Group –
http://www.w3.org/XML/EXI/