



# **W3C Efficient XML Interchange (EXI)**

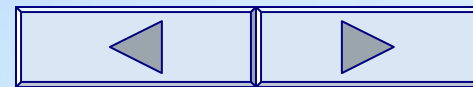
*Ed Day*  
*Objective Systems, Inc.*  
*<http://www.obj-sys.com>*

# What is EXI?

From the Draft Standard

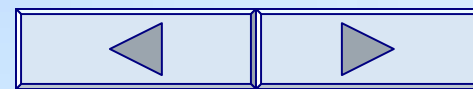
(<http://www.w3.org/TR/2007/WD-exi-20070716/>):

“Efficient XML Interchange (EXI) is a very compact representation for the XML Information Set that is intended to simultaneously optimize performance and the utilization of computational resources.”



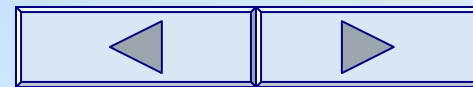
# History

- Started with W3C XML Binary Characterization (XBC) Working Group
- Tasked to gather use cases and requirements to determine the feasibility of producing an alternate encoding for XML
- Produced Use Cases, Properties, Measurement Methodologies, and Characterization documents



# Key Use Cases

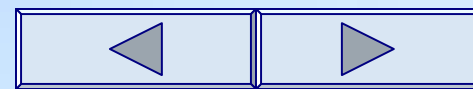
- Mobile : low bandwidth, low power devices (note: Moore's law does not apply to battery life).
- Scientific data : processing large amounts of data in limited time frames
- Financial : similar to scientific – required high throughput of time sensitive data
- Others: X3D graphics, Broadcast Metadata, Military..



# History

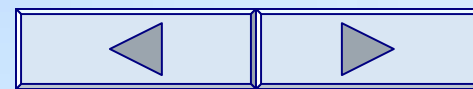
- Efficient XML Interchange (EXI) Working Group formed after TAG review of XBC results.
- Mission:  
“..develop a format that allows efficient interchange of the XML Information Set, based on the conclusions of the XML Binary Characterizations Working Group” \*

\* from the EXI charter



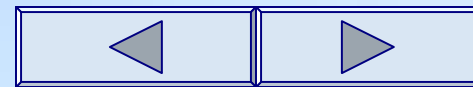
# Documents Available Now

- Draft format specification  
(published this past July):  
<http://www.w3.org/TR/2007/WD-exi-20070716/>
- Measurements Note – comprehensive report on  
all tests that were done:  
<http://www.w3.org/TR/2007/WD-exi-measurements-20070725/>



# Documents Coming Soon

- Updated format specification draft
- Primer
- Best Practices
- Impact on existing XML Technologies



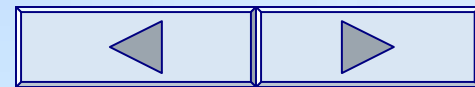
# Open Source Project

- OpenEXI project formed by Sun:

<https://exi.dev.java.net>

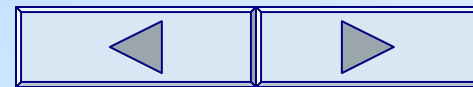
(project is not yet public due to legal issues,  
but expected to be soon..)

Java SAX/StAX API based on EXI.



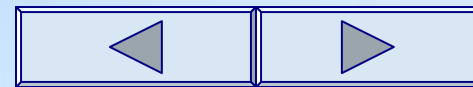
# Brief Overview of the Format

- “Schemaless” mode uses learned knowledge.
- “Schema-informed” mode can leverage information from a schema.
- Special compression mode biases stream for better compression.



# Schemaless Mode

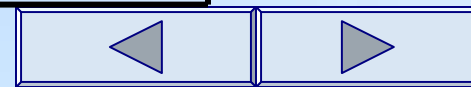
- Built-in XML Grammars describe XML Infoset.
- Knowledge is learned as move through instance.
- Event codes and string tables are key structural concepts.



# Event Codes

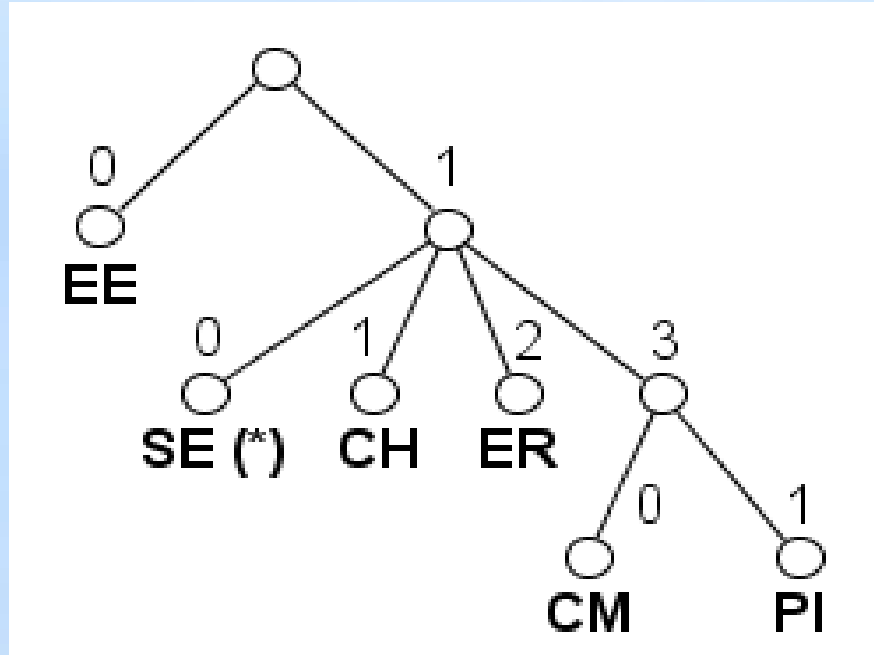
- Use Huffman-like trees
- More likely events use fewer bits

Event Type	Content	Grammar Notation
Start Document		SD
Start Element	qname	SE(*),SE(qname)
Characters	value	CH
Attribute	qname, value	AT(*),AT(qname)
Namespace	prefix, uri	NS
...		



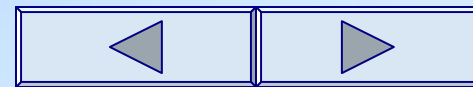
# Example Tree

- Element Content



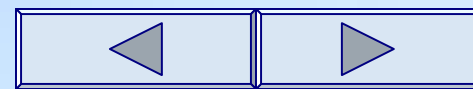
# String Tables

- All content in schemaless mode stored as strings
- Each content item is remembered when it is first seen.
- If seen again, is replaced with a compact identifier (index into a string table).
- Partitions for different items: namespace URI's, prefixes, local names, and values.



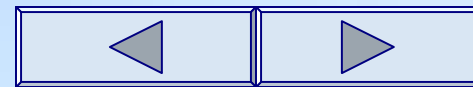
# Schema Informed Mode

- Specific grammars based on schemas
- Start Element (SE) and Attribute (AT) events can be predicted with more certainty.
- Can provide enhanced compactness through knowledge of what is expected
- Also, data type of content is used instead of strings.
- Can handle extensions and deviations from schema.



# Compression Mode

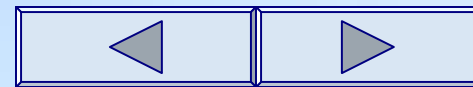
- Start by using a byte-aligned instead of a bit-packed encoding
- Events are sorted into channels
- Standard “deflate” algorithm can then be applied to biased stream



# FAQ

Q: Why is gzip not good enough?

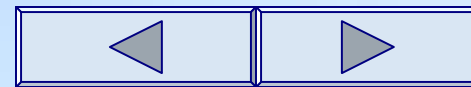
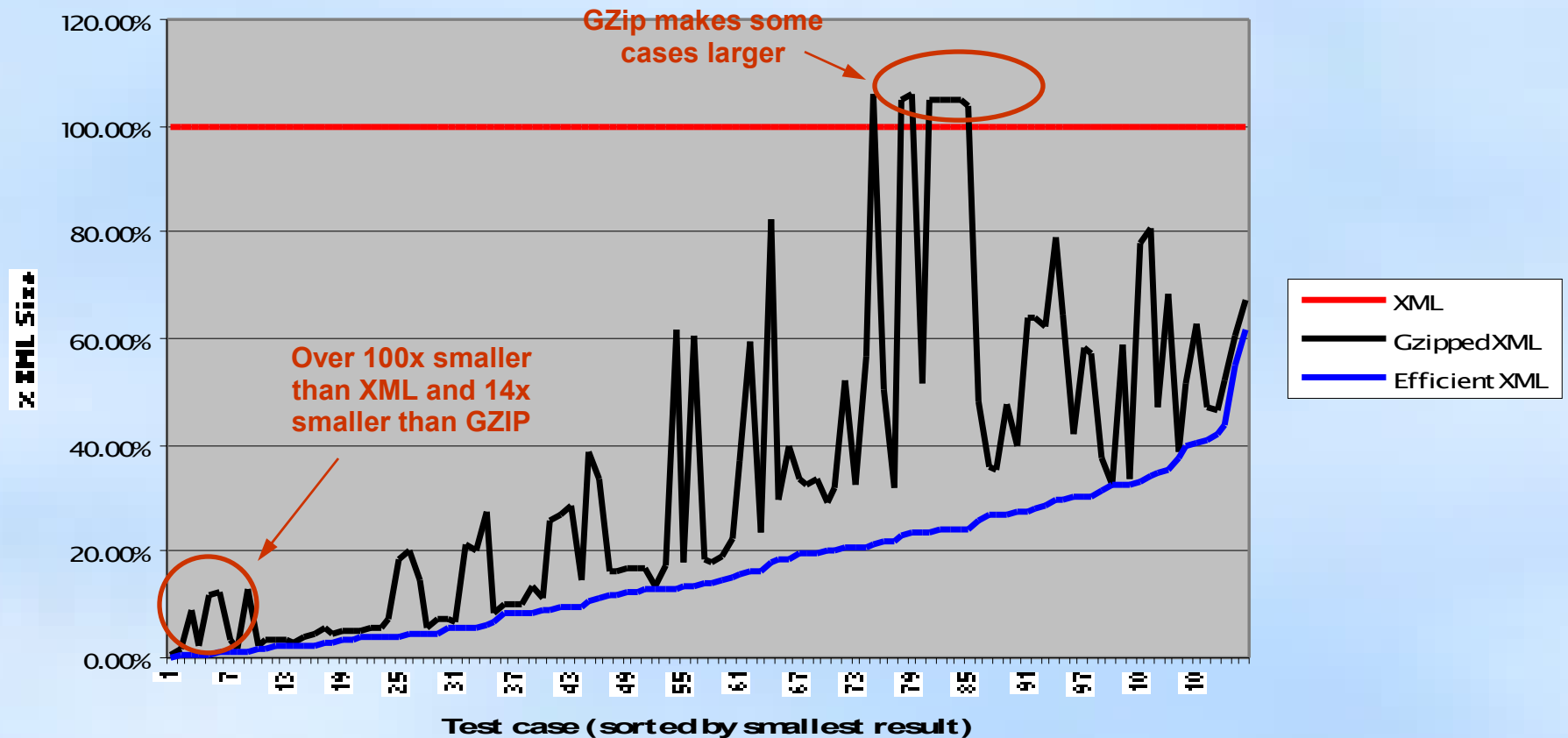
A: In general, computational resources required are too great for many use cases. Also requires working with XML text.



# Compactness Summary

(produced by AgileDelta Efficient XML)

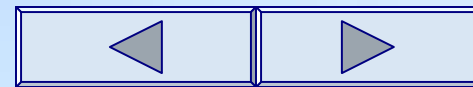
## Compactness : With Schemas and Compression



## FAQ (cont'd)

Q: What about Fast Infoset? (ISO standard)

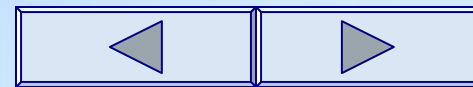
A: Was considered as a candidate for the format. In the end, it was determined through testing that the base candidate format we selected outperformed FI in many areas.



# FAQ (cont'd)

Q: Why do you call it XML? (it is not XML)

A: It is an exchange format that is designed specifically for XML data. It is no way intended to replace XML. It is a transmission format (similar to gzip).



# FAQ

Q: Why are you trying to destroy the web?

A: We are not. We are trying to expand the web by getting XML into places where it could not be used before.

