



# Position on the Binary Interchange of XML Infosets

Oliver Goldman

Larry Masinter  
*Adobe Systems*



## Summary Position

### **At most one alternate Infoset encoding**

- XML has demonstrated the value of a single encoding
- Potential value in trading text encoding for other properties

### **The alternate encoding must be viable for all applications**

- Web services, documents, others

### **The process must give due consideration to all requirements**

- The result should augment, not detract from, XML



## Background and Experience: XML

	Data	Content	Templates	Web Services	Metadata
<b>Acrobat</b>	in/out	out		client	out
<b>Document Server</b>	in/out	in	in	server	in/out
<b>Form Server</b>	in/out		in	server	
<b>Graphics Server</b>	in	in	in	server	in/out
<b>FrameMaker</b>		in/out	in/out		in/out
<b>GoLive</b>		in/out	in/out		in/out
<b>Illustrator</b>		in/out	in/out		in/out
<b>InDesign</b>		in/out			in/out



## Documents vs. Web Services

### **Several recent proposals focus on web services**

- E.g. encodings designed solely for small data sets
- Tends to proliferate encodings

### **Some properties are valuable in both cases**

- Web service payloads sometimes are documents
- Chunking and incremental update are related



# Background and Experience: PDF

## Portable Document Format:

- Over 10 years old (predates XML)
- Scales from small to large (1 GB+) documents
- Supports embedded images, video, data
- Supports encryption, digital signatures



# XML for Documents

## Desirable document format properties:

- ✓ **Structure**
- ❑ **Random Access:** better-than-linear performance when accessing elements within the document
- ❑ **Compactness:** document size not excessive with regard to the amount of information present
- ❑ **Non-destructive Incremental Update:** update time proportional to new, not existing, data; existing data untouched



## XML for Documents (cont.)

### **Desirable document format properties:**

- Cannot be achieved via simple compression
- At odds with textual encodings
- At odds with the original goals of XML
- Essential for a certain class of documents
- A binary encoding does not necessarily achieve these properties



# Binary Data

**Binary data is increasingly important due to images, video**

- Native binary encodings
- Base64 encoding is expensive in time and space
- Binary encoding does not solve this problem

**Possible alternatives:**

- 1. Extend Infoset to make it binary-data aware**
  - A change to XML
- 2. Use a packaging mechanism**
  - Attachments inaccessible to XQuery, etc.
- 3. Define binary encoding on Post Schema Validation Infoset**
  - Also problematic...





# Relationship to Infoset and Schema

## Using the PSVI:

- Encoding can be more compact.
  - Requires schema definition and valid documents. In practice, many documents used without either.
  - Difficult to combine schemas and namespaces, yet both are in widespread use.
- **New encoding must work with Infoset, without Schema**
- Might have optional features which leverage PSVI



## Conclusion

### **At most one alternate Infoset encoding**

- XML has demonstrated the value of a single encoding
- Potential value in trading text encoding for other properties

### **The alternate encoding must be viable for all applications**

- Web services, documents, others

### **The process must give due consideration to all requirements**

- The result should augment, not detract from, XML



**Adobe**

Tools for the New Work™