

Meta-data (or maybe metadata)

Lawrence Snyder
University of Washington, Seattle

OED Entry For Byte

- Suppose you want a digital version of the Oxford English Dictionary

On Paper

byte (balt). *Computers*. [Arbitrary, prob. influenced by bit sb.⁴ and bite sb.] A group of eight consecutive bits operated on as a unit in a computer. **1964** *Blaauw & Brooks* in *IBM Systems Jnl.* III. 122 An 8-bit unit of information is fundamental to most of the formats [of the System/360]. A consecutive group of n such units constitutes a field of length n . Fixed-length fields of length one, two, four, and eight are termed bytes, halfwords, words, and double words respectively. **1964** *IBM Jnl. Res. & Developm.* VIII. 97/1 When a byte of data appears from an I/O device, the CPU is seized, dumped, used and restored. **1967** *P. A. Stark Digital Computer Programming* xix. 351 The normal operations in fixed point are done on four bytes at a time. **1968** *Dataweek* 24 Jan. 1/1 Tape reading and writing is at from 34,160 to 192,000 bytes per second.

- Type it in in ASCII

The Result ...

On Paper

byte (balt). *Computers*. [Arbitrary, prob. influenced by bit sb.⁴ and bite sb.] A group of eight consecutive bits operated on as a unit in a computer. **1964 Blaauw & Brooks** in *IBM Systems Jrnl.* III.

As ASCII

byte (baIt). *Computers*. [Arbitrary, prob. influenced by bit sb.⁴ and bite sb.] A group of eight consecutive bits operated on as a unit in a computer. 1964 Blaauw & Brooks in *IBM Systems Jrnl.* III.

- Not all of the information in the dictionary is captured

Fix the Formatting w/HTML

On Paper

byte (balt). *Computers*. [Arbitrary, prob. influenced by bit sb.⁴ and bite sb.] A group of eight consecutive bits operated on as a unit in a computer. **1964 Blaauw & Brooks** in *IBM Systems Jrnl.* III.

As ASCII + HTML tags

```
<b>byte</b> (baIt). Computers. [Arbitrary, prob. influenced  
By bit sb.<sup>4</sup> and bite sb.] A group of eight  
consecutive bits operated on as a unit in a computer.<b>  
1964 <i>Blaauw & Brooks</i></b> in <i>IBM Systems Jrnl. </i>  
III.
```

- It might look like the dictionary, but it's not very helpful: "What's the definition of *computer*"

Metadata – Specify with Tags

■ Metadata is information about information

byte (balt). *Computers*. [Arbitrary, prob. influenced by bit sb.⁴ and bite sb.] A group of eight consecutive bits operated on as a unit in a computer. **1964 Blaauw & Brooks** in *IBM Systems Jrnl.* III. 122 An 8-bit unit of information is fundamental to most of the formats [of the System/360]. A consecutive group of *n* such units constitutes a field of length *n*. Fixed-length fields of length one, two, four, and eight are termed bytes, halfwords, words, and double words respectively. **1964 IBM Jrnl. Res. & Developm.** VIII. 97/1 When a byte of data appears from an I/O device, the CPU is seized, dumped, used and restored. **1967 P. A. Stark** *Digital Computer Programming* xix. 351 The normal operations in fixed point are done on four bytes at a time. **1968 Dataweek** 24 Jan. 1/1 Tape reading and writing is at from 34,160 to 192,000 bytes per second.

```
<e><hg><hw>byte</hw> <pr><ph>baIt</ph></pr></hg>. <la>Computers</la>. <etym>
Arbitrary, prob. influenced by <xr><x>bit</x></xr> <ps>n.<hm>4</hm> </ps>and
<xr><x>bite</x> <ps>n.</ps> </xr></etym> <s4>A group of eight consecutive bits
operated on as a unit in a computer.</s4> <q><q><q><q>1964 </q><a>Blaauw</a>
& <a>Brooks</a> <bib>in</bib> <w>IBM Systems Jrnl.</w> <lc>III. 122</lc>
<qt>An 8-bit unit of information is fundamental to most of the formats <ed>of
the System/360</ed>.&es.A consecutive group of <i>n</i> such units constitutes
a field of length <i>n</i>.&es.Fixed-length fields of length one, two, four,
and eight are termed bytes, halfwords, words, and double words respectively.
</qt></q><q><q><q>1964</q> <w>IBM Jrnl. Res. & Developm.</w> <lc>VIII.
97/1</lc> <qt>When a byte of data appears from an I/O device, the CPU is
seized, dumped, used and restored.</qt></q> <q><q><q> 1967</q> <a>P. A. Stark
</a> <w>Digital Computer Programming</w> <lc>xix. 351</lc> <qt>The normal
operations in fixed point are done on four bytes at a time.</qt><q><q><q><q>
1968</q> <w>Dataweek</w> <lc>24 Jan. 1/1</lc> <qt>Tape reading and writing is
at from 34,160 to 192,000 bytes per second.</qt></q></qp></e>
```

Metadata Describes Data

- Metadata is data about data ... a description of what the data is
 - Knowing what the data is, as in the OED, allows us to process it better for users
 - Here's an example: Search OED for def of “binary”
 - Without metadata, get 8,311 hits ... of which one is the definition
 - With metadata, get each definition in order ... how?
<e><hg><hw>binary</hw> ... </hg> ... <e>

Metadata Describes Data

- Metadata is data about data ... a description of what the data is
 - Knowing what the data is, as in the OED, allows us to process it better for users
 - Here's an example: Search OED for def of “binary”
 - Without metadata, get 8,311 hits ... which one is the definition?
 - With metadata, get each definition in order ... how?
- <e><hg><hw>binary</hw> ... </hg> ... <e>

The Principle: We can program computers to better help us if we say what the content is

Metadata Separation

- Metadata describes what the data is, but because the tags can be distinguished from the content, it separates itself from the content – that's smart

Separate the content and its tags entirely from the processing – produce a data-only file

The Advantage of Separating

- By separating the content from the processing it is possible to maximize expertise
 - The content expert (you) puts the data together
 - The processing expert (some programmers) write the processing code based on the tags



http://blog.flexcommunity.net/lab/ppv3dGallery_slicedCube/

Photos Rolling Over As Cubes



Just Do It!

http://blog.flexcommunity.net/lab/ppv3dGallery_slicedCube

Enter The World of XML

- The Extensible Markup Language (XML) the tool for defining metadata; YOU think up the tags ... it is a self-defining language!
 - The usual rules for tags apply
 - Enclose in < and > and use lowercase ONLY
 - Start tag `<mynewtag>` and End tag `</mynewtag>`
 - Tags must always be matched or self-terminated
 - Tags can have attributes (think those up, too) of form
`attributename="valueInQuotes"`
 - Use `.xml` as the file extension
 - Always start with “standard text” (shown later)

Example of XML

- Suppose I want to record information about this class; using XML, I might write:

```
<class dept="cse">  
  <catalog qsr="true" credits="5">  
    <num>120</num>  
    <lec len="50" num="3">M, W, F</lec>  
    <lab len="50" num="2"> Tu,Th </lab>  
    <descrip>  
      Must-know computing knowledge for the  
      21st century</descrip>  
  </catalog>  
  <teach>L. Snyder</teach>  
</class>
```

I invented the tags; they make sense to me, and I could write software to process such descriptions

Learning XML

- Since we think up the tags ourselves, it's the easiest language in the world to learn, right?
- Right.
- It's trivial?!
- Not quite ... there is a little technique, and we'll do that now

- Tags can serve in three roles ...

Ways To Use Tags

- **Identity** – tag it so you know what it is

```
<name>George Washington</name>
```

- **Affinity** – all properties of a thing should be collected together

```
<personal>
```

```
  <name>George Washington</name>
```

```
  <height>6' 2"</height>
```

```
  <teeth>Wooden</teeth>
```

```
  <home>Mount Vernon</home>
```

```
</personal>
```

Ways To Use Tags (continued)

- **Collection** – enclose a group of items of the same type in a collective tag

```
<presidents>  
  <prez num="1"><personal><name>George ...  
  <prez num="2"><personal><name>John ...  
  <prez num="3"><personal><name>Thomas ...  
  ...  
  <prez num="44"><personal><name>Barack ...  
</presidents>
```

- These uses become intuitive quickly

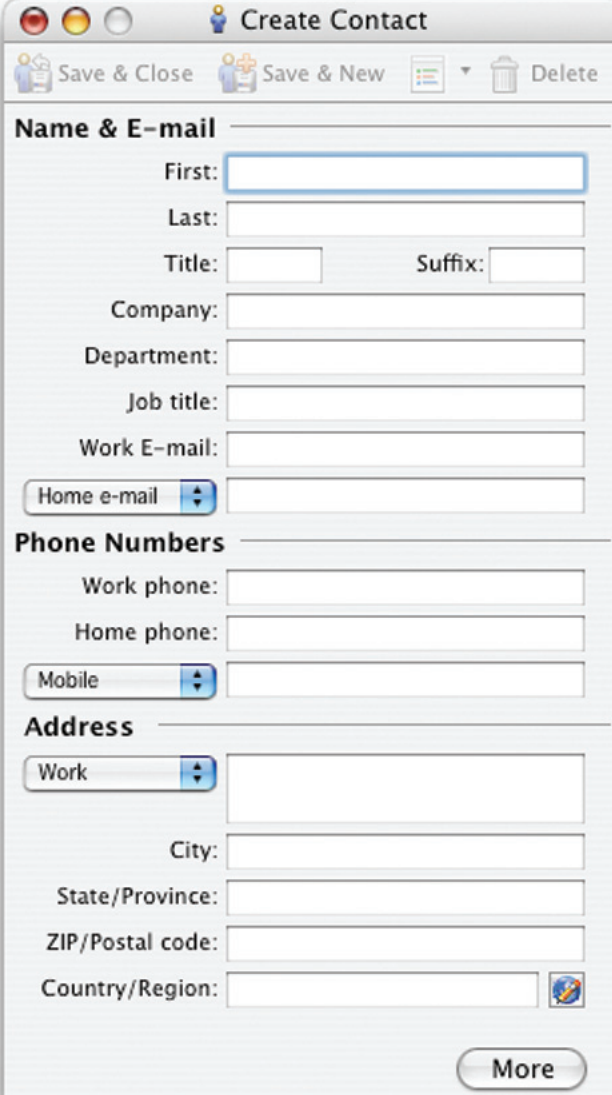
Ex:

Classify
tag types:
identity
affinity
collection

```
- <travels>
  - <visit>
    <sight>Washington State</sight>
    - <action flag="wash.gif">
      The State of Washington is a fun place to visit. We toured Spokane,
      Grand Coulee Dam, Seattle's Space Needle and Mt. Rainier, which
      wasn't rainy at all, but beautiful in the sun!
    </action>
  </visit>
  - <visit>
    <sight>Oregon</sight>
    - <action flag="oregon.jpg">
      South of Washington is Oregon. It is at the end of the old Oregon
      Trail. It is an unusual place. First, the University of Oregon's team is
      called the Ducks. Also, Mt. Bachelor is near the Sisters; with so
      many women around, why is it still a bachelor?
    </action>
  </visit>
  - <visit>
    <sight>California</sight>
    - <action flag="california.png">
      California seems to be a republic, but not a banana republic. More
      like an orange republic. We visited San Francisco, San Quentin, the
      Monterey Bay Aquarium, LA and Hollywood. We didn't see any
      stars, but we were not there in the dark either.
    </action>
  </visit>
</travels>
```


Next few minutes ...

- Here's what we will do now
 - Everyone will come forward so we can take a class picture for use later in the lecture
 - You return to your seats and work out XML tags for an address book – write on paper
 - When that's done you'll vote on your fave movie
 - Return to the lecture



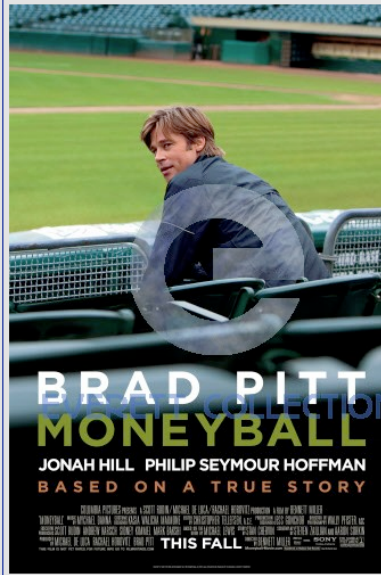
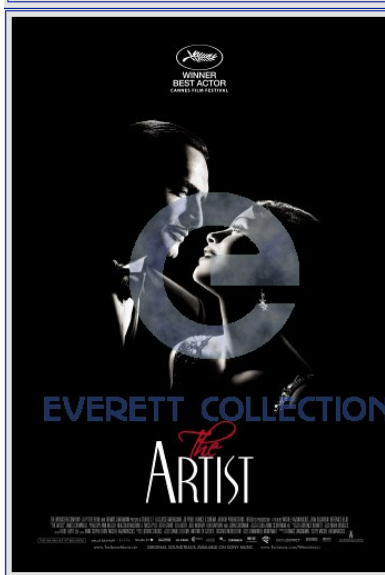
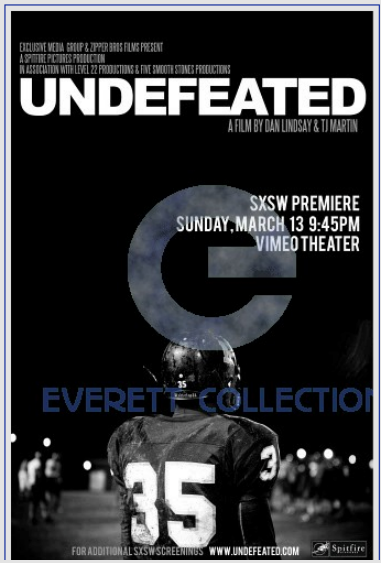
The image shows a screenshot of a web browser window titled "Create Contact". The window has a standard Mac OS X title bar with red, yellow, and green window control buttons. Below the title bar, there are three buttons: "Save & Close", "Save & New", and "Delete". The form is organized into several sections:

- Name & E-mail:** Includes fields for "First:", "Last:", "Title:", "Suffix:", "Company:", "Department:", "Job title:", "Work E-mail:", and "Home e-mail" (with a dropdown arrow).
- Phone Numbers:** Includes fields for "Work phone:", "Home phone:", and "Mobile" (with a dropdown arrow).
- Address:** Includes a "Work" dropdown arrow, a large text input field, "City:", "State/Province:", "ZIP/Postal code:", and "Country/Region:" (with a small globe icon).

At the bottom right of the form, there is a "More" button.

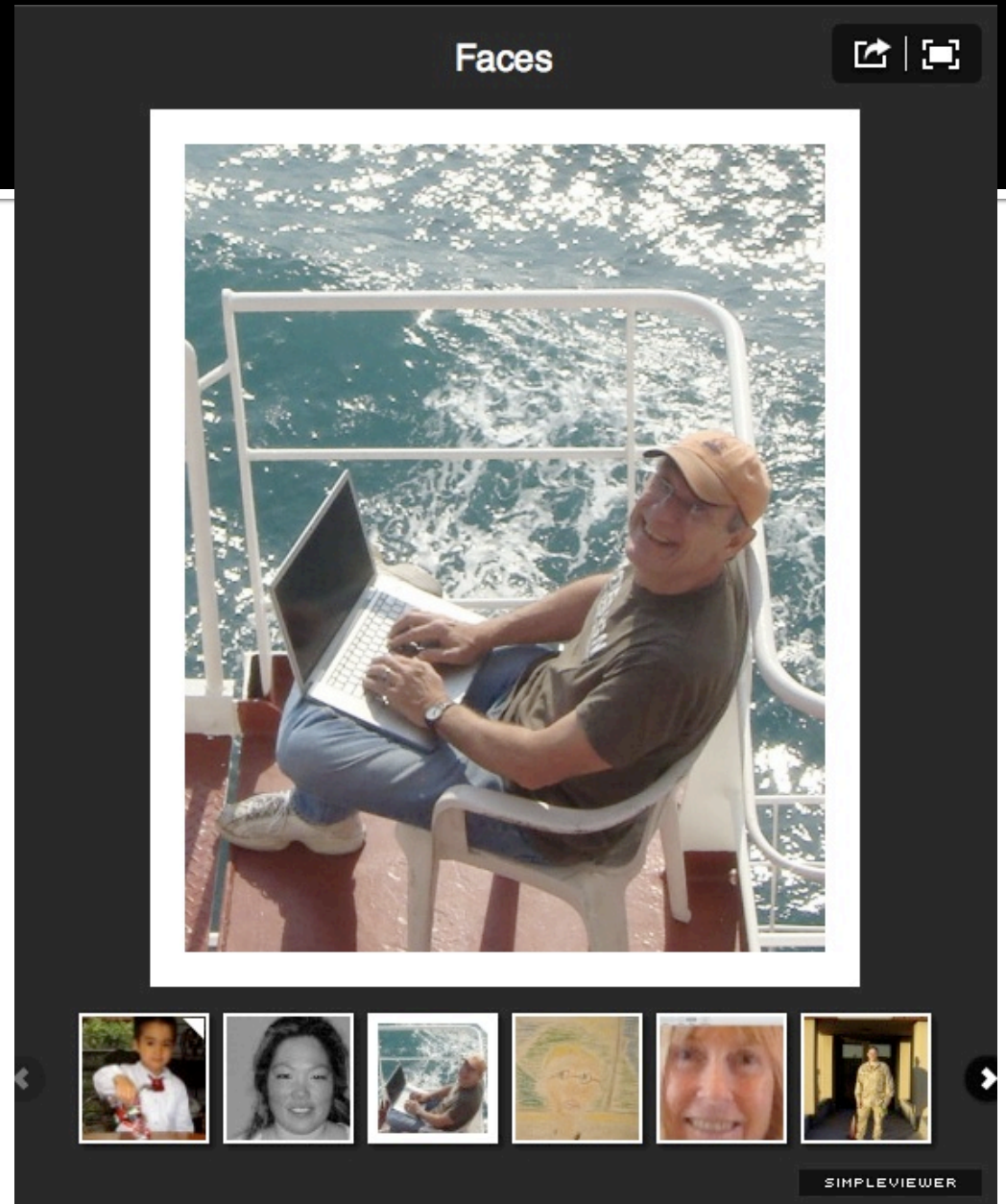


Movies ... Give Your Top 3



Check it Out!

- Here is a simple picture viewer I got off the net...
- Two parts
 - Setup loads pix
 - View



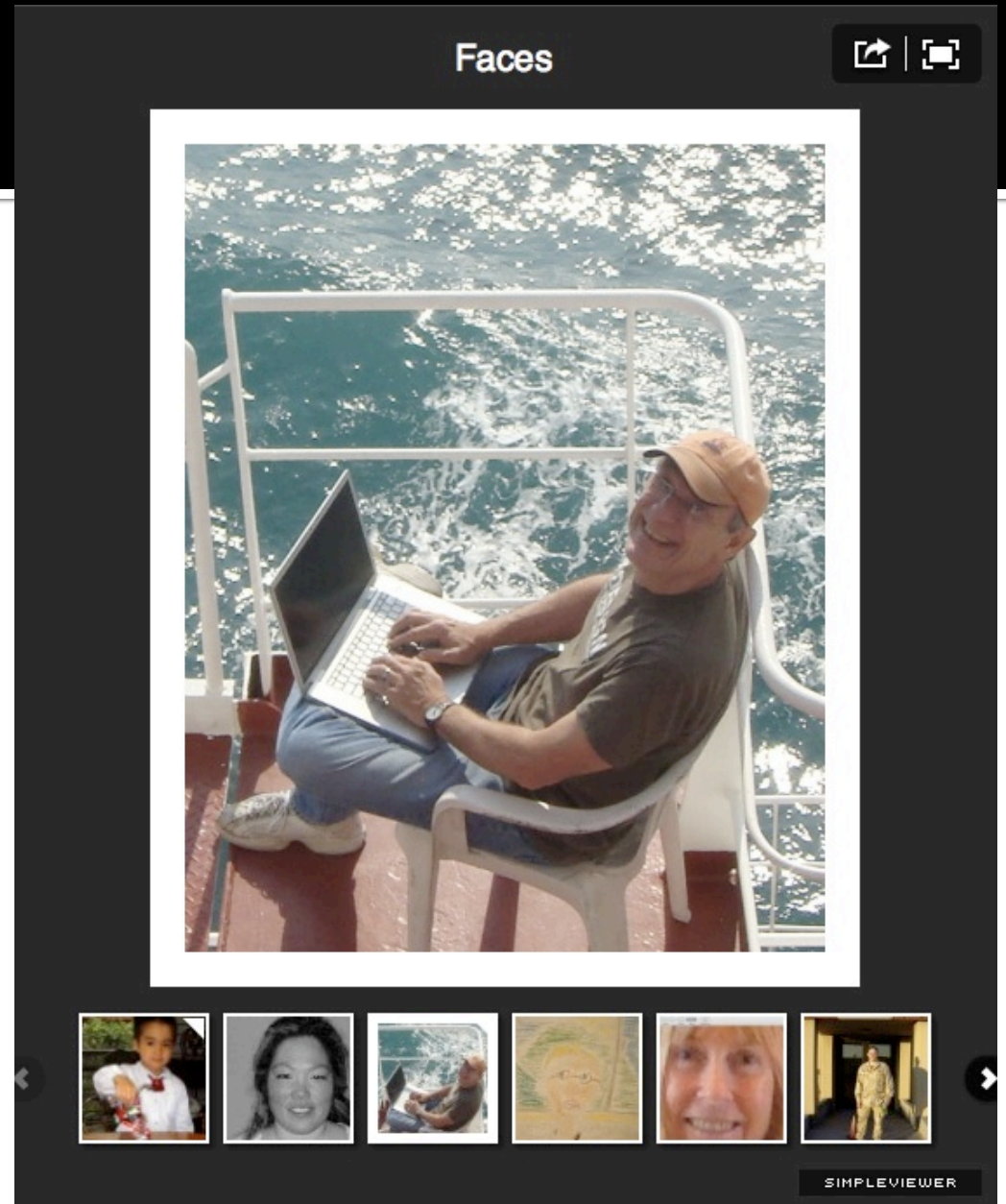
- <http://www.simpleviewer.net/simpleviewer/>

Check it Out!

- Here is a simple picture viewer I got off the net...
- Two parts
 - Setup loads pix
 - View

<file:///Users/lawrencesnyder/Desktop/SimpleViewer1%20copy/index.html>






Just Do IT



- <http://www.simpleviewer.net/simpleviewer/>

How To Use The Software Works

- The Simpleviewer software can be understood

Name	
 gallery.xml	The XML metadata for each item
 images	Folder of picture files
 index.html	The Web page of the display
 svcore	Folder of software files
 thumbs	Folder of tiny picture files

gallery.xml

- Check out the structure, tags
...
- Components
 - XML identifier
 - Collection tag
 - Affinity tag
 - Identity tag

```
gallery.xml (no symbol selected)
<?xml version="1.0" encoding="UTF-8"?>
<simpleviewergallery
    useFlickr="false"
    resizeMode="true"
    cropToFit="false"
    maxImageWidth="1024"
    maxImageHeight="768"
    imageQuality="80"
    thumbWidth="75"
    thumbHeight="75"
    thumbQuality="90"
    title="Faces"
    thumbPosition="BOTTOM"
    thumbRows="1"
    thumbColumns="6"
    showOpenButton="true"
    showFullscreenButton="true"
>
<image imageURL="images/sungeun.gif"
    thumbURL="thumbs/sungeun.gif"
    linkURL="images/sungeun.gif"
    linkTarget="_blank">
    <caption><![CDATA[SungEun]]></caption>
</image>
...
</simpleviewergallery>
```

An Item To Display: <image>

```
<image imageURL="images/sungeun.gif"  
  thumbURL="thumbs/sungeun.gif"  
  linkURL="images/sungeun.gif"  
  linkTarget="_blank">  
  <caption><![CDATA[SungEun]]> </caption>  
</image>
```

```
<image imageURL="images/shipcomputing1.JPG"  
  thumbURL="thumbs/shipcomputing1.JPG"  
  linkURL="images/shipcomputing1.JPG"  
  linkTarget="_blank">  
  <caption><![CDATA[On Board]]> </caption>  
</image>
```

```
<image imageURL="images/portrait.JPG"  
  thumbURL="thumbs/portrait.JPG"  
  linkURL="images/portrait.JPG"  
  linkTarget="_blank">  
  <caption><![CDATA[Karalina's Art]]> </caption>  
</image>
```

Just Do IT

Summary

- Metadata is data about data
- Tags are a common form of metadata
- XML is main technology for metadata spec.
- Three roles for tags to fill ... you're building a tree
- By separating data from processing, expertise can be exploited, flexibility, wide usage
- We used metadata to add an image