Data Formats on the Web

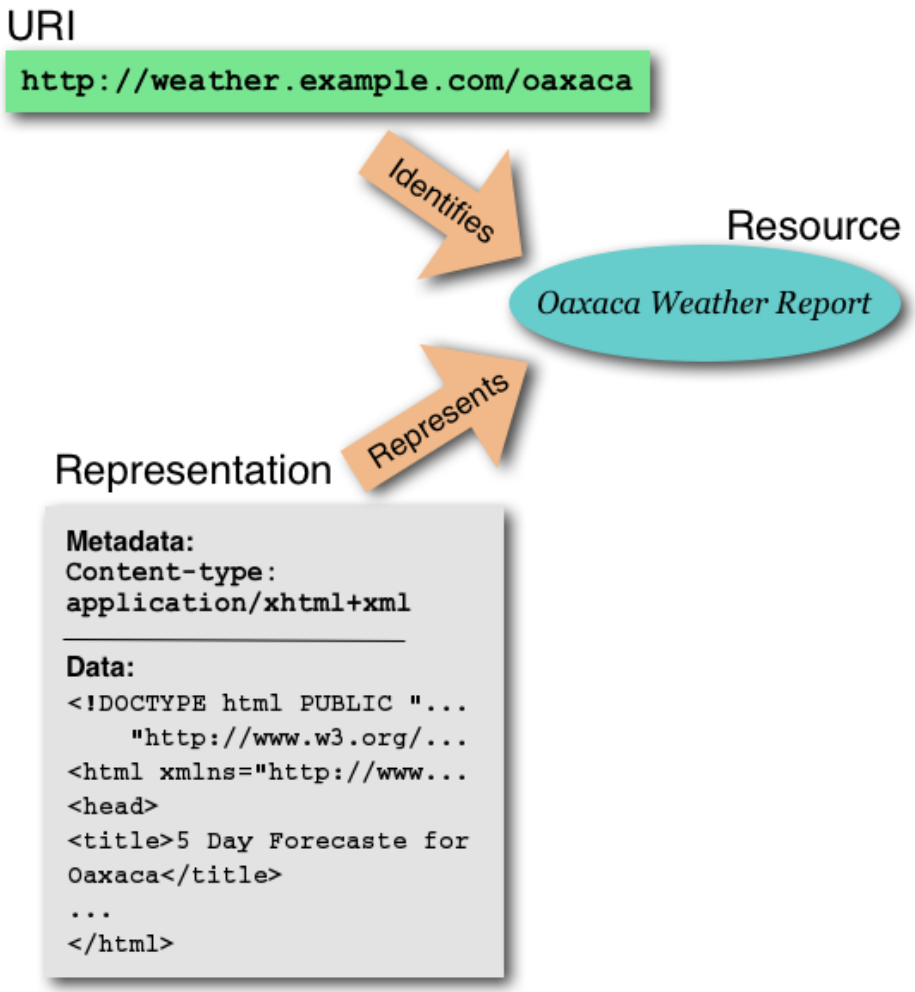
COMP6218

Web Architecture

How to use this Lecture

- In this lecture I will go into a lot of detail about a variety of Web formats
- Do not panic - you are not expected to memorise all these details!
- You are expected to be familiar with the range of data formats, broadly how they work and what they look like “under the hood”, and what they are useful for.
  - E.g. what are the sections inside a PDF file? Is it best to use plain text, PDF or ePUB for formatting your new novel?

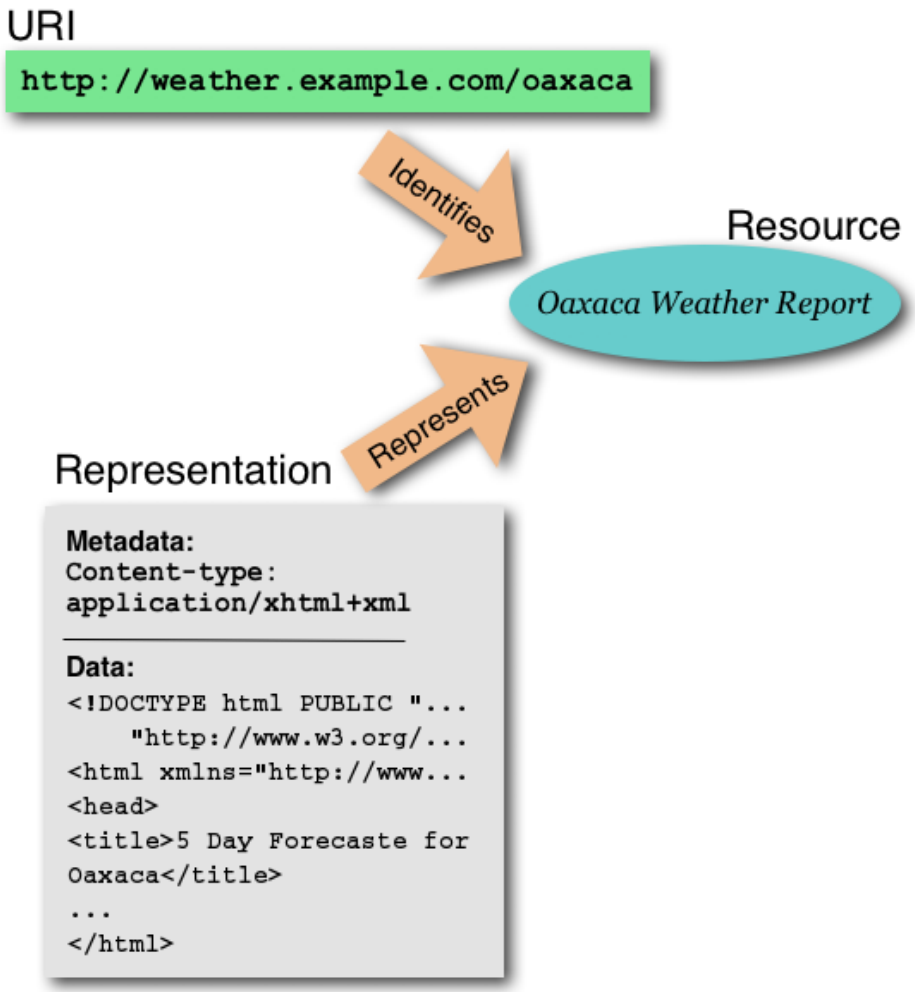
Data Formats for Web Resources



"The ***World Wide Web*** is an information space in which the items of interest, referred to as resources, are identified by global identifiers called Uniform Resource Identifiers (***URI***)."

http://www.w3.org/TR/webarch/

URIs, Resources, and Representations



- URIs *identify* Resources
- Representations *represent* Resources
- When URIs are dereferenced, they return representations (not resources)
- Different representations may be returned for the same URI (e.g., English vs. French version)

http://www.w3.org/TR/webarch/

Content Negotiation

- Content negotiation (RFC 2616 sec 12) is used in HTTP to allow servers to send different representations of the same resource at the same URI
- The user agent tells the server which encoding, language, media type, etc. it prefers, and the server responds with the “best” representation
- Example HTTP request headers:

```
Accept-Language: fr; q=1.0, en; q=0.5
Accept: text/html; q=1.0, text/*; q=0.8,
   image/gif; q=0.6, image/jpeg; q=0.6, image/*;
   q=0.5, */*; q=0.1
```

  This agent prefers French over English, HTML over other document types, and GIF and JPEG over other image formats

Internet Media Types

- A media type is composed of a *type*, a *subtype*, and optional parameters.
  - text/html; charset=UTF-8
  - text is the type, html is the subtype, and charset=UTF-8 is a parameter (the character encoding).
- **Common examples**
  - text/html
  - text/plain
  - application/pdf
  - application/json
  - audio/mpeg
  - video/mp4
  - image/png
  - application/x-www-form-urlencoded
  - application/vnd.openxmlformats-officedocument.wordprocessingml.document

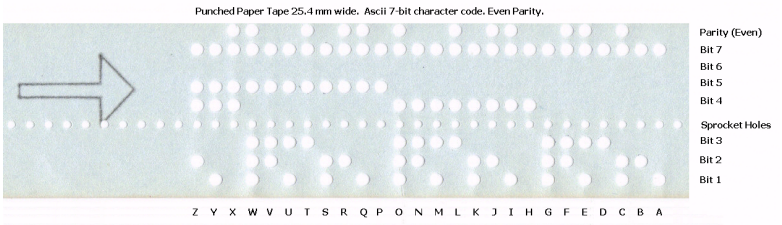
These replace PC filename extensions for the internet (e.g. “text/plain” vs “.txt”) and are translated by the web server.

Common Web Document Formats

- Plain text
  - ASCII vs Unicode
- HTML
  - (+CSS +JavaScript +media)
  - HTML 4
  - XHTML
  - HTML5
- XML
  - RSS
  - MathML
  - SVG
  - Office Open XML
    - docx, pptx, xlsx
  - EPUB
- PDF
  - Based on PostScript
- Data
  - TSV
  - CSV
  - JSON

Plain Text - ASCII

- ASCII maps the 1-byte numbers 0-127 into letters
  - Initially used for paper tape / punched cards
  - 1 byte = 8 bits = 0 - 255, but the 8th bit was used for error correcting (parity bit).



Dec	Hx	Oct	Char	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr
0	0	000	NUL (null)	32	20	040		Space	64	40	100	@	@	96	60	140	`	`
1	1	001	SOH (start of heading)	33	21	041	!	!	65	41	101	A	A	97	61	141	a	a
2	2	002	STX (start of text)	34	22	042	"	"	66	42	102	B	B	98	62	142	b	b
3	3	003	ETX (end of text)	35	23	043	#	#	67	43	103	C	C	99	63	143	c	c
4	4	004	EOT (end of transmission)	36	24	044	$	$	68	44	104	D	D	100	64	144	d	d
5	5	005	ENQ (enquiry)	37	25	045	%	%	69	45	105	E	E	101	65	145	e	e
6	6	006	ACK (acknowledge)	38	26	046	&	&	70	46	106	F	F	102	66	146	f	f
7	7	007	BEL (bell)	39	27	047	'	'	71	47	107	G	G	103	67	147	g	g
8	8	010	BS (backspace)	40	28	050	(	(	72	48	110	H	H	104	68	150	h	h
9	9	011	TAB (horizontal tab)	41	29	051	)	)	73	49	111	I	I	105	69	151	i	i
10	A	012	LF (NL line feed, new line)	42	2A	052	*	*	74	4A	112	J	J	106	6A	152	j	j
11	B	013	VT (vertical tab)	43	2B	053	+	+	75	4B	113	K	K	107	6B	153	k	k
12	C	014	FF (NP form feed, new page)	44	2C	054	,	,	76	4C	114	L	L	108	6C	154	l	l
13	D	015	CR (carriage return)	45	2D	055	-	-	77	4D	115	M	M	109	6D	155	m	m
14	E	016	SO (shift out)	46	2E	056	.	.	78	4E	116	N	N	110	6E	156	n	n
15	F	017	SI (shift in)	47	2F	057	/	/	79	4F	117	O	O	111	6F	157	o	o
16	10	020	DLE (data link escape)	48	30	060	0	0	80	50	120	P	P	112	70	160	p	p
17	11	021	DC1 (device control 1)	49	31	061	1	1	81	51	121	Q	Q	113	71	161	q	q
18	12	022	DC2 (device control 2)	50	32	062	2	2	82	52	122	R	R	114	72	162	r	r
19	13	023	DC3 (device control 3)	51	33	063	3	3	83	53	123	S	S	115	73	163	s	s
20	14	024	DC4 (device control 4)	52	34	064	4	4	84	54	124	T	T	116	74	164	t	t
21	15	025	NAK (negative acknowledge)	53	35	065	5	5	85	55	125	U	U	117	75	165	u	u
22	16	026	SYN (synchronous idle)	54	36	066	6	6	86	56	126	V	V	118	76	166	v	v
23	17	027	ETB (end of trans. block)	55	37	067	7	7	87	57	127	W	W	119	77	167	w	w
24	18	030	CAN (cancel)	56	38	070	8	8	88	58	130	X	X	120	78	170	x	x
25	19	031	EM (end of medium)	57	39	071	9	9	89	59	131	Y	Y	121	79	171	y	y
26	1A	032	SUB (substitute)	58	3A	072	:	:	90	5A	132	Z	Z	122	7A	172	z	z
27	1B	033	ESC (escape)	59	3B	073	;	;	91	5B	133	[	[	123	7B	173	{	{
28	1C	034	FS (file separator)	60	3C	074	<	<	92	5C	134	\	\	124	7C	174			|
29	1D	035	GS (group separator)	61	3D	075	=	=	93	5D	135	]	]	125	7D	175	}	}
30	1E	036	RS (record separator)	62	3E	076	>	>	94	5E	136	^	^	126	7E	176	~	~
31	1F	037	US (unit separator)	63	3F	077	?	?	95	5F	137	_	_	127	7F	177	&#127;	DEL

Source: www.Lookup

Later standards defined the numbers 128 – 255 for different regions of the world e.g. ISO-LATIN-1 for European diacriticals

code	char	code	char	code	char	code	char	code	char	code	char	co
160		161	¡	162	¢	163	³	164	¤	165	¹	1
168	¨	169	©	170	º	171	«	172	¬	173		1
176	°	177	±	178	²	179	³	180	´	181	µ	1
184	¸	185	¹	186	º	187	»	188	¾	189	½	1
192	à	193	-	194	â	195	ã	196	ä	197	å	1
200	è	201	é	202	ê	203	ë	204	ì	205	í	2
208	ð	209	ñ	210	ò	211	ó	212	ô	213	õ	2
216	ø	217	ù	218	ú	219	û	220	ü	221	ý	2
224	à	225	á	226	â	227	ã	228	ä	229	å	2
232	è	233	é	234	ê	235	ë	236	ì	237	í	2
240	ð	241	ñ	242	ò	243	ó	244	ô	245	õ	2
248	ø	249	ù	250	ú	251	û	252	ü	253	ý	2

Plain Text - Unicode

- Uses 4-byte numbers.
- Defines characters from 0 – 1114111
  - European, Asian, Egyptian Hieroglyphics, Emoji
- Allows 1-byte (8-bit) representation: UTF-8
  - UTF-8 is different from ASCII, and responsible for these errors when you try and cut/paste from MS Word into a Web browser or similar
- Backward compatible with ASCII
- UTF-8 is used by 87.9% of Web pages
- Any byte value > 127 indicates a multi-byte letter

IngrÃ©dients

IngrÃ©dients

Farine de blÃ© (85%), matiÃ¨re grasse vÃ©gÃ©tale (palme), sucre, levure, gluten, agent de traitement de la farine : acide ascorbique.PrÃ©sence possible de fruits Ã coque, de lait, d'oeufs, de sÃ©same et de soja.

= 128169 (Unicode decimal) = 1F4A9 (Unicode hex)
= (UTF-8 bit representation) 11110000, 10011111, 10010010, 10101001 = 240, 159, 146, 169

HTML

- Key document format for Web
- Structured for main applications of Web pages
  - Headers, articles, sections, media, divs, links
- Incorporates stylesheets, media, scripts

XML

- More general (eXtensible) document and data language
- No fixed semantics, tag names or elements
- Used to define other languages
  - MathML
  - SVG
  - docx, pptx, xslx
  - EPub

```
<?xml version="1.0" encoding="utf-8" ?>
<books>
    <book>
        <bookid> B001 </bookid>
        <title> Understanding XML </title>
        <Price> $30 </Price>
        <Author>
            <FirstName> Lily </FirstName>
            <LastName> Hicks </LastName>
        </Author>
    </book>
    <book>
        <bookid> B002 </bookid>
        <title> .NET Framework </title>
        <Price> $45 </Price>
        <Author>
            <FirstName> Jasmine </FirstName>
            <LastName> Williams </LastName>
        </Author>
    </book>
</books>
```

MathML

$$f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots$$

- MathML facilitates the use and re-use of mathematical and scientific content on the Web
  - computer algebra systems
  - print typesetting
  - voice synthesis
- MathML can be used to encode both the presentation of mathematical notation for high-quality visual display, and mathematical content, for applications where the semantics is important.
- MathML is an application of XML
  - with adequate style sheet support, it is possible for browsers to natively render maths
  - several vendors offer applets and plug-ins which can render MathML in place in a browser.
  - Translators and equation editors can generate HTML pages with embedded MathML.

- **Why is the W3C working in this area?**
  - Although the mark-up language HTML has a large repertoire of tags, it does not cater for math.
  - With no means of using HTML tags to mark up mathematical expressions, authors have resorted to drastic means. A popular method involves inserting images - literally snap shots of equations taken from other packages and saved in GIF format - into technical documents.

Example MathML

$$x^2 + 4x + 4 = 0$$

```
<apply>
    <eq/>
    <apply>
        <plus/>
        <apply>
            <power/>
            <ci>x</ci>
            <cn>2</cn>
        </apply>
        <apply>
            <times/>
            <cn>4</cn>
            <ci>x</ci>
        </apply>
        <cn>4</cn>
    </apply>
    <cn>0</cn>
</apply>
```

This is Semantic MathML, using tags that are oriented to the mathematical meaning

Example MathML

$$
\begin{aligned}
(a+b)^2 &= c^2 + 4 \cdot (\tfrac{1}{2}ab) \\
a^2 + 2ab + b^2 &= c^2 + 2ab \\
a^2 + b^2 &= c^2
\end{aligned}
$$

- <math style="display: block;">
- <mtable columnalign="right center left">
- <mtr> <mtd> <msup> <mrow> <mo> (</mo> <mi> a </mi> <mo> + </mo> <mi> b </mi> <mo>) </mo> </mrow> <mn> 2 </mn> </msup> </mtd> <mtd> <mo> = </mo> </mtd> <mtd> <msup><mi> c </mi><mn>2</mn></msup> <mo> + </mo> <mn> 4 </mn> <mo> ⁢ </mo> <mo>(</mo> <mfrac> <mn> 1 </mn> <mn> 2 </mn> </mfrac> <mi> a </mi><mi> b </mi> <mo>)</mo> </mtd> </mtr>
- <mtr> <mtd> <msup><mi> a </mi><mn>2</mn></msup> <mo> + </mo> <mn> 2 </mn><mi> a </mi><mi> b </mi> <mo> + </mo> <msup><mi> b </mi><mn>2</mn></msup> </mtd> <mtd> <mo> = </mo> </mtd> <mtd> <msup><mi> c </mi><mn>2</mn></msup> <mo> + </mo> <mn> 2 </mn><mi> a </mi><mi> b </mi> </mtd> </mtr>
- <mtr> <mtd> <msup><mi> a </mi><mn>2</mn></msup> <mo> + </mo> <msup><mi> b </mi><mn>2</mn></msup> </mtd> <mtd> <mo> = </mo> </mtd> <mtd> <msup><mi> c </mi><mn>2</mn></msup> </mtd> </mtr>
- </mtable>
- </math>

This is Presentational MathML, using tags that are oriented to the printed layout

SVG – Scalable Vector Graphics

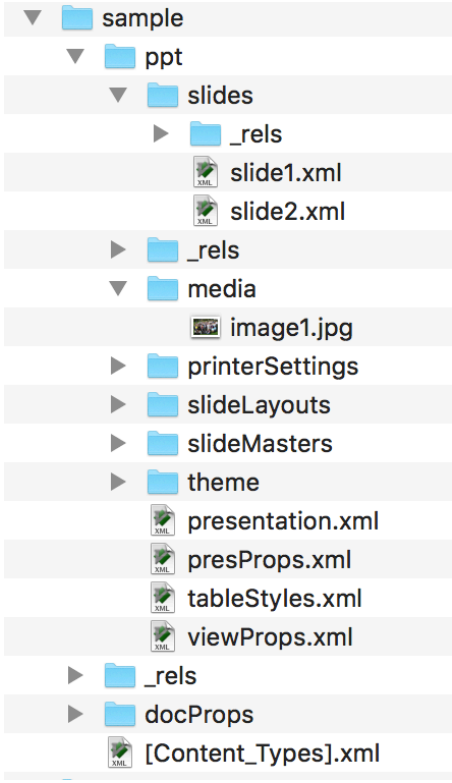
- SVG is a language for describing 2D graphics in XML.
- In SVG, each drawn shape is remembered as an independent object. If its attributes are changed, the browser automatically redraws the shape.

```
<svg  width="100"  height="100">
  <circle cx="50" cy="50" r="40" stroke="green" stroke-width="4" fill="yellow" />
</svg>
```

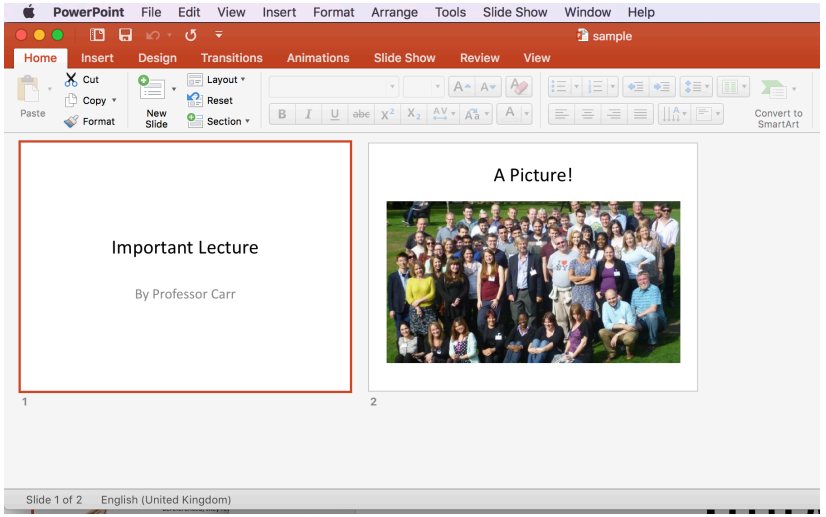
- **Native HTML Drawing (Canvas)**
  - Resolution dependent
  - No support for event handlers
  - Poor text rendering capabilities
  - You can save the resulting image as .png or .jpg
  - Well suited for graphic-intensive games
- **SVG**
  - Resolution independent
  - Support for event handlers
  - Best suited for applications with large rendering areas (Google Maps)
  - Slow rendering if complex (anything that uses the DOM a lot will be slow)
  - Not suited for game applications

Office Open XML

- From Office 2007, Microsoft moved to XML-based formats
- An Office document is a ZIP file of a directory hierarchy of lots of XML files
  - docprops directory contains metadata
  - ppt directory contains all the slide info
  - media directory contains images etc
  - rels directories translate file names into XML attributes
    - imageid1 = file sample/media/image1.jpg

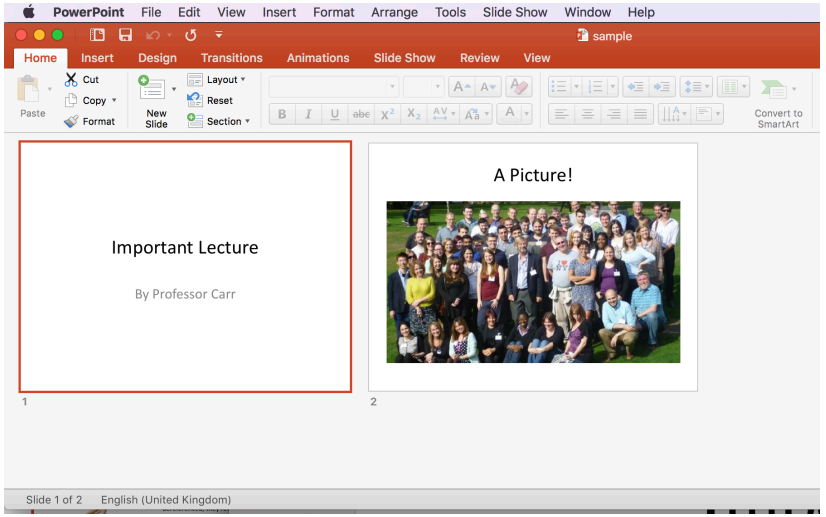


ppt/slide1.xml



- <?xml version="1.0" encoding="UTF-8" standalone="yes"?><p:sld xmlns:a="http://schemas.openxmlformats.org/drawingml/2006/main" xmlns:r="http://schemas.openxmlformats.org/officeDocument/2006/relationships" xmlns:p="http://schemas.openxmlformats.org/presentationml/2006/main"><p:cSld><p:spTree><p:nvGrpSpPr><p:cNvPr id="1" name=""/><p:cNvGrpSpPr/><p:nvPr/></p:nvGrpSpPr><p:grpSpPr><a:xfrm><a:off x="0" y="0"/><a:ext cx="0" cy="0"/><a:chOff x="0" y="0"/><a:chExt cx="0" cy="0"/></a:xfrm></p:grpSpPr><p:sp><p:nvSpPr><p:cNvPr id="2" name="Title 1"/><p:cNvSpPr><a:spLocks noGrp="1"/></p:cNvSpPr><p:nvPr><p:ph type="ctrTitle"/></p:nvPr></p:nvSpPr><p:spPr/><p:txBody><a:bodyPr/><a:lstStyle/><a:p><a:r><a:rPr lang="en-US" dirty="0" smtClean="0"/>**<a:t>Important Lecture</a:t>**</a:r><a:endParaRPr lang="en-US" dirty="0"/></a:p></p:txBody></p:sp><p:sp><p:nvSpPr><p:cNvPr id="3" name="Subtitle 2"/><p:cNvSpPr><a:spLocks noGrp="1"/></p:cNvSpPr><p:nvPr><p:ph type="subTitle" idx="1"/></p:nvPr></p:nvSpPr><p:spPr/><p:txBody><a:bodyPr/><a:lstStyle/><a:p><a:r><a:rPr lang="en-US" dirty="0" smtClean="0"/>**<a:t>By Professor Carr</a:t>**</a:r><a:endParaRPr lang="en-US" dirty="0"/></a:p></p:txBody></p:sp></p:spTree><p:extLst><p:ext uri="{BB962C8B-B14F-4D97-AF65-F5344CB8AC3E}"><p14:creationId xmlns:p14="http://schemas.microsoft.com/office/powerpoint/2010/main" val="1773458668"/></p:ext></p:extLst></p:cSld><p:clrMapOvr><a:masterClrMapping/></p:clrMapOvr></p:sld>

ppt/slide2.xml



- <?xml version="1.0" encoding="UTF-8" standalone="yes"?><p:sld xmlns:a="http://schemas.openxmlformats.org/drawingml/2006/main" xmlns:r="http://schemas.openxmlformats.org/officeDocument/2006/relationships" xmlns:p="http://schemas.openxmlformats.org/presentationml/2006/main"><p:cSld><p:spTree><p:nvGrpSpPr><p:cNvPr id="1" name=""/><p:cNvGrpSpPr/><p:nvPr/></p:nvGrpSpPr><p:grpSpPr><a:xfrm><a:off x="0" y="0"/><a:ext cx="0" cy="0"/><a:chOff x="0" y="0"/><a:chExt cx="0" cy="0"/></a:xfrm></p:grpSpPr><p:sp><p:nvSpPr><p:cNvPr id="2" name="Title 1"/><p:cNvSpPr><a:spLocks noGrp="1"/></p:cNvSpPr><p:nvPr><p:ph type="title"/></p:nvPr></p:nvSpPr><p:spPr/><p:txBody><a:bodyPr/><a:lstStyle/><a:p><a:r><a:rPr lang="en-US" dirty="0" smtClean="0"/>**<a:t>A Picture!</a:t>**</a:r><a:endParaRPr lang="en-US" dirty="0"/></a:p></p:txBody></p:sp><p:pic><p:nvPicPr><p:cNvPr id="4" name="Content Placeholder 3" descr="P1040467.jpg"/><p:cNvPicPr><a:picLocks noGrp="1" noChangeAspect="1"/></p:cNvPicPr><p:nvPr>**<p:ph idx="1"/>**</p:nvPr></p:nvPicPr><p:blipFill><a:blip r:embed="rId2"><a:extLst><a:ext uri="{28A0092B-C50C-407E-A947-70E740481C1C}"><a14:useLocalDpi xmlns:a14="http://schemas.microsoft.com/office/drawing/2010/main" val="0"/></a:ext></a:extLst></a:blip><a:srcRect t="10121" b="10121"/><a:stretch><a:fillRect/></a:stretch></p:blipFill><p:spPr/></p:pic></p:spTree><p:extLst><p:ext uri="{BB962C8B-B14F-4D97-AF65-F5344CB8AC3E}"><p14:creationId xmlns:p14="http://schemas.microsoft.com/office/powerpoint/2010/main" val="2699647514"/></p:ext></p:extLst></p:cSld><p:clrMapOvr><a:masterClrMapping/></p:clrMapOvr></p:sld>

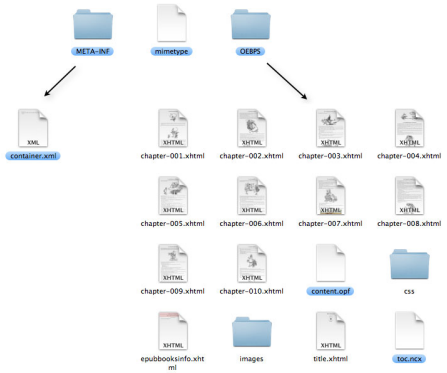
ePUB

- ePUB is an HTML-based e-book file format with the extension .epub that can be downloaded and read on smartphones, tablets, computers, or e-reader devices.
- It is a technical standard published by the International Digital Publishing Forum (IDPF), now jointly with W3C.
- Endorsed buy the Book Industry Study Group as the format of choice for packaging content.
  - ePUB is the most widely supported vendor-independent XML-based (as opposed to PDF) e-book format
  - It is widely used on many software readers such as iBooks on iOS and Google Books on Android,
  - It is NOT used by Amazon Kindle e-readers.

ePUB Features

- Reflowable document (main application) for e-readers
  - Can also support fixed-layout pre-paginated content for certain kinds of highly designed books, comics or advertising.
- Inline images, metadata, and CSS styling
- Page bookmarking
- Passage highlighting and notes
- A library that stores books and can be searched
- Re-sizable fonts, and changeable text and background colors
- Support for a subset of MathML
- Digital rights management (DRM) as an optional layer

ePUB Format



- An ePUB file is a ZIP archive that contains, in effect, a website
  - HTML files, images, CSS style sheets, metadata and other assets.
  - By using HTML5, publications can contain video, audio, and interactivity, just like websites in web browsers.
- The ePUB container must contain:
  - At least one content document.
  - One navigation document.
  - One package document listing all publication resources including a “spine” ie an ordered sequence of ID references defining the default reading order.

PDF

Google reports 2.3bn PDF documents vs 9bn HTML

- Important document format for the Web
- Structured for rendering of pre-formatted documents
  - Painting onto a screen in a device-independent way
    - Set CHARACTERS from FONTS at POSITION
    - Draw lines, arcs, images at POSITION
    - No concept of paragraphs, line breaking, headings, lists etc
- Often used as the FINAL, OFFICIAL format of record
- PDF documents are structured as
  - a set of objects
  - a final index that points to the position of each object for efficiency (image objects may be multi-Gb)

See PDF tutorial!

Sample PDF... objects

%PDF-1.0

1 0 obj

Root Object

<<

/Type /Catalog /Pages 3 0 R /Outlines 2 0 R >>

endobj

2 0 obj

Outlines Object

<<

/Type /Outlines /Count 0

>>

endobj

3 0 obj

Page List

<<

/Type /Pages /Count 1 /Kids [4 0 R] >>

endobj

4 0 obj

First Page

<<

/Type /Page

/Parent 3 0 R

/Resources << /Font << /F1 7 0 R >> /ProcSet 6 0 R >> /MediaBox [0 0 612 792]

/Contents 5 0 R >>

endobj

5 0 obj

<< /Length 44 >>

Drawing Commands for First Page

stream

BT /F1 24 Tf

100 100 Td (Hello There) Tj ET

endstream

endobj

6 0 obj

Definitions for First Page

[/PDF /Text]

endobj

7 0 obj

Fonts for First Page

<<

/Type /Font /Subtype /Type1 /Name /F1 /BaseFont /Courier >>

endobj

PTO for rest of file...

Sample PDF… index

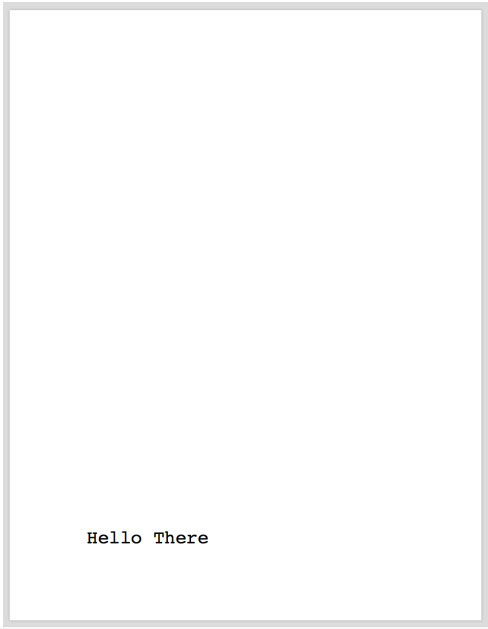
xref
08
0000000000 65535 f
0000000009 00000 n
0000000074 00000 n
0000000120 00000 n
0000000179 00000 n
0000000322 00000 n
0000000415 00000 n
0000000445 00000 n
trailer
<<
/Size 8 /Root 1 0 R >>
startxref 553
%%EOF

Index

Count of Objects, ID of Root Object

Pointer to Start of Index

End Of File



Simple PDF file rendered by PDF Viewer

Key instructions from data stream in object #5

BT /F1 24 Tf
100 100 Td (Hello There) Tj ET

BeginText use font1 at size 24, move to (100, 100), draw the text "Hello There" EndText

Uses of PDF

- Developed before the Web to facilitate the exchange of hardcopy documents
- PDF is the format people use when they need an electronic “hard copy” document.
- Many business, publishing and records-keeping applications require a reliable, flexible and capable AND IMMUTABLE analog for paper.
- Alternative is scanned images (e.g. TIFF) but those are UNSEARCHABLE.
- PDF features
  - Archival quality control (PDF/A)
  - Extensible document and content-level metadata
  - Annotations and fillable forms
  - Security (passwords) and authenticity
  - Accessibility
  - Controllable content re-use
  - Redaction
  - Watermarking
  - 3D, video and other rich content
  - Scripting

PDF and HTML – the pros

PDF	HTML
Consistent layout	Customized to device (incl. mobile)
Offline accessibility	Enriched and interactive content
Easy to store and organize	Always latest version
Similar to print version	Up-to-date and linked context
Easy to print	Linked with data repositories
Displays images well	Easy to search
Easy to share by email (when small)	Easy to share by link (also when large)
Easy to annotate	Fast access from lists
	Includes supplementary material

This list is taken TerraXML.com

Simple Web Data

- Tab separated

```
time    Height above sea level: [13,129]        Temperature, w
09/01/2011 23:00        -0.011628735    NA      NA
09/02/2011 00:00        -0.0076217903   NA      NA
09/02/2011 01:00        9.15126E-4      NA      NA
09/02/2011 02:00        -0.0048775193   NA      NA
```

- Comma separated

```
Jan's Illustrated Computer Literacy 101,,,,,,,,
,Total Hits,increase over previous month %,Total Hits on Pages,increase over
2010,,,,,,,,
Jan ,"14,283,059",#REF!,"937,606",#REF!,60.30,#REF!,"386,240",#REF!
Feb,"20,358,731",43%,"1,190,643",27%,81.11,35%,"519,694",35%
Mar,"21,403,930",5%,"1,237,711",4%,88.06,9%,"564,030",9%
Apr,"18,758,304",-12%,"1,039,302",-16%,78.74,-11%,"504,235",-11%
```

- JSON

```
"firstName": "John",
"lastName": "Smith",
"age": 25,
"address": {
    "streetAddress": "21 2nd S
    "city": "New York",
    "state": "NY",
    "postalCode": 10021
},
"phoneNumbers": [
    {
        "type": "home",
```

Implementing Web Data

- Some data standards are supported natively by all browsers
  - HTML, CSS, plain text
- Some are unevenly supported
  - MathML, MPEG
- Some are given to external applications
  - Word
- Some are displayed by third party browser plugins
  - Flash
  - Java
- Problems: software maintenance, efficiency, security, multi-platform development