Introduction to Markup and Markdown

Christopher Ohge Institute of English Studies, University of London

christopher.ohge@sas.ac.uk







This presentation

Part I

- What is Markup and why should you care?
- How markup languages have changed.

Part II

- A Markdown primer for web authoring.
- Compiling Markdown: pandoc, GitHub Pages, Jekyll, Hugo...

What is Markup?

Two kinds of *annotations* regarding *instructions*, per the *Oxford English Dictionary*:

Bibliographical (instructions to the compositor using abbreviations or signs):

"The process or result of marking corrections on copy or proofs in preparation for printing" (2b).

Computational (instructions to the computer using tags):

"The process of embedding tags in an electronic text so as to distinguish the text's logical, syntactic, or structural components; the tags so embedded" (2c).

- Markup of a non-digital text is a reductive sampling of complex writing systems (ASCII and Unicode)
- Text is represented by sequences of binary digits according to encoding schemes (made by people)

What is Markup?

Markup is essential to annotating many activities, from data storage, information management, data structuring, metadata preservation and delivery, and Web publishing.

It is also a form of translation:

- Markup is a way of making our knowledge or understanding about a text explicit
- Markup makes strives to make explicit (to a machine) what is implicit (to a person)
- Markup assists us in facilitating re-use of the same material:
 - in different formats
 - in different contexts
 - by different sorts of users

A brief history of markup

1967: "Generic coding," a text tagging system by William Tunnicliffe and the Graphic Communication Association's (GCA) Composition Committee, was presented at a conference for the Canadian Government Printing Office.

1970s: Development of roff, nroof, and TeX, all different kinds of presentational markup for digital typesetting.

1969/1973: Charles Goldfarb and a team at IBM developed Generalized Markup language (GML). Assumes a 'document type'.

1986: Goldfarb released the Standard Generalized Markup Language (SGML). Not strictly a markup language but a formal grammar for the design and specification of markup.

1989: Text Encoding Initiative Guidelines released, expressed

in SGML (they are now expressed in XML, in v. 5)

1991: Tim Berners-Lee released a document called "HTML tags," which proposed 18 tags based on the SGML language for authoring web pages. Thus the HyperText Markup Language was born.

1994: First official specification of HTML 2.0, published by the Internet Engineering Task Force (IETF).

IETF, published the HTML 4 standards.

1999: World Wide Web Consortium (W3C) superseded the

Early 2000s: Schism between XHTML and HTML.

2006: Berners-Lee wrote a blog post that stated the HTML to XML transition had not worked. Changes to web standards need to incremental.

Present: most of us still see (and use) HTML5. XHTML still used, though (epub, for example).

Kinds of Markup

Procedural Markup (e.g. <u>troff</u>, <u>TeX</u>, and <u>PostScript</u>): RED INK ON; print "-£1000"; RED INK OFF

Presentational Markup (e.g. css, html):

\textcolor{red}{-£1000}

Descriptive Markup (e.g. html, sgml, xml):

<measure unit=" pounds" value=" -1000">

One thousand pounds in debt</measure>

Descriptive Markup

- It is usually more useful to mark up what we think the elements of data represent rather than what they look like.
- Descriptive markup enables us to make explicit distinctions for processing a string of characters.
- It formalises the naming, characterising, and annotating of textual data which enables processing and re-use.

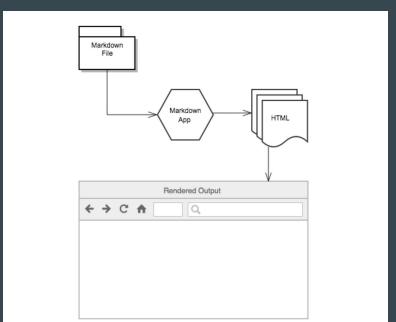
Kinds of Markup

• Whereas presentational markup concerns the appearance of a text, descriptive markup is a metalanguage that deals with the meaning of the text (what it is).

An encoded source file is not immediately processed, so that makes sharing much easier and

improves workflow.

 Take this simple example of a Markdown file workflow:



Markdown

• The computer scientist John Gruber created this simple markup language in 2004 after becoming frustrated by writing tedious HTML tags to properly format his content.

- Markdown is a text-to-HTML conversion tool for web writing.
 - It is an easy-to-use syntax for quickly styling and structuring all forms of writing on the web.
 - It is often used on GitHub's README files, but it is also usable on other blog services like WordPress. It easily converts to .html, .pdf, and .docx with Pandoc.

Markdown and WYSIWYG in text editing

- Markdown editing does not involve a "what-you-see-is-what-you-get" (often abbreviated to WYSIWYG) text editor like Microsoft Word, in which you click on buttons and format the document immediately.
- With Markdown you enter in the syntax in-text to instruct the computer how a document should look.
- Basically, Markdown is just regular text with a few non-alphabetic characters thrown
 in, like # or *. So *The Tempest* is rendered as *The Tempest*. If we put # in front of
 The Tempest, it becomes a first-level header.
- Why write a file in Markdown (.md)?

Markdown continued.

The importance structure and of workflow: not only writing flow, but the flow of a computational pipeline. It is also good for documenting and demoing code.

Using the Atom text editor (and other text editors), you can preview your markdown for readability. In Atom, you enter Markdown preview by entering control+shift+m. An example—take some plain text:

In a Station of the Metro

By Ezra Pound, 1885 - 1972

The apparition of these faces in the crowd;

Petals on a wet, black bough.

(Source: Poets.org, https://www.poets.org/poetsorg/poem/station-metro)

Markdown structure

```
# *In a Station of the Metro*
## By Ezra Pound, 1885 - 1972
The apparition of these faces in the
crowd;
Petals on a wet, black bough.
(**Source:**
[Poets.org] (https://www.poets.org/poetsorg
/poem/station-metro))
```

Markdown preview

```
# *In a Station of the Metro*
## By Ezra Pound, 1885 - 1972
```

```
The apparition of these faces in the crowd;
```

```
Petals on a wet, black bough.
```

```
(**Source:**
[Poets.org] (https://www.poets.org/
poetsorg/poem/station-metro))
```

In a Station of the Metro

By Ezra Pound, 1885 - 1972

The apparition of these faces in the crowd;

Petals on a wet, black bough.

Source: Poets.org

Markdown preview enhanced

In a Station of the Metro^[Source:
[Poets.org](https://www.poets.org/poetsorg/poem/stationmetro)]

By Ezra Pound^[Ezra Pound (1885 - 1972) was an American poet...]

The apparition of these faces in the crowd;

Petals on a wet, black bough.

[*Note:* you may need to install a preview package that creates the automatic footnotes. In Atom, it is markdown-preview-enhanced. (Go to Preferences > Install and enter "markdown-preview-enhanced".)]

In a Station of the Metro [1]

By Ezra Pound^[2]

The apparition of these faces in the crowd;

Petals on a wet, black bough.

- 1. Source: Poets.org ←
- 2. Ezra Pound (1885 1972) was an American poet...

The Philosophy of Markdown.

To quote John Gruber, the creator of Markdown (https://daringfireball.net/projects/markdown/):

> The idea for Markdown is to make it easy to read, write, and edit prose. HTML is a **publishing** format; Markdown is a **writing** format. Thus, Markdown's formatting syntax only addresses issues that can be conveyed in plain text.



The Philosophy of Markdown.

To quote John Gruber, the creator of Markdown:

The idea for Markdown is to make it easy to read, write, and edit prose. HTML is a *publishing* format; Markdown is a *writing* format. Thus, Markdown's formatting syntax only addresses issues that can be conveyed in plain text.

Some Markdown authoring and editing applications

- Mac: <u>Texts</u>, <u>iA Writer</u>, <u>Ulysses</u>, <u>Marked</u>, or <u>MacDown</u>
- iOS / Android: <u>iA Writer</u> or <u>Ulysses</u> (iOS only)
- Windows: <u>Texts</u>, <u>MarkdownPad</u>, <u>Typora</u>
- Linux: Typora
- Web: <u>Dillinger</u> or <u>StackEdit</u>
- Looking to self-publish a book? <u>Leanpub</u>, which converts Markdown files into an electronic book (PDF, EPUB, and MOBI file format); <u>Bookdown</u> (based on R Markdown).
- You can also incorporate code into Markdown (R Markdown, e.g.).

Compiling Markdown

- Use Pandoc (a universal document converter from the command line): https://pandoc.org/
 - Converts md to docx, html, PDF, and pretty much anything else...
 - Sample syntax: pandoc text.md -o text.html
 - Perhaps most importantly for text encoders: it converts to TEI (simple).
 - Sample syntax: pandoc -o text.xml --to=tei
 - See the user manual: https://pandoc.org/MANUAL
- Push to GitHub Pages: https://pages.github.com/
- Use a static website generator
 - Jekyll: https://jekyllrb.com/docs/
 - Hugo: https://gohugo.io/
- Copy and paste into your content management system (using html code editor), e.g. WordPress

html to Markdown?

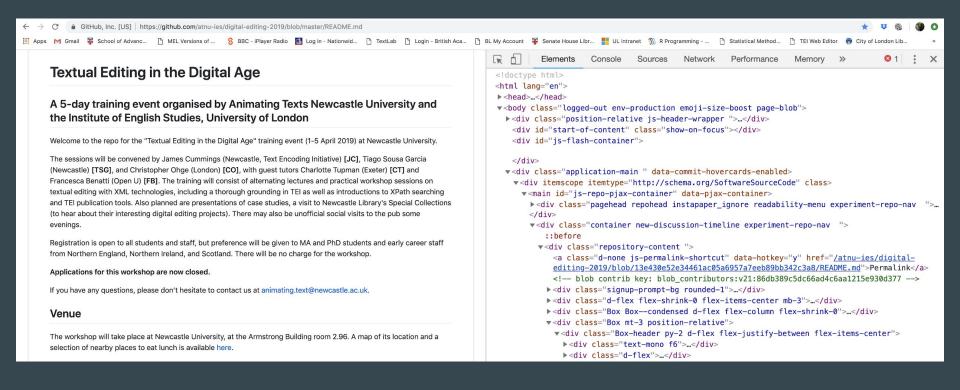
- What if you already have an html file?
- Use Aaron Aaron Swartz 's text2html tool:

http://www.aaronsw.com/2002/html2text/

• Use Pandoc (in the command line):

pandoc text.html -o text.md

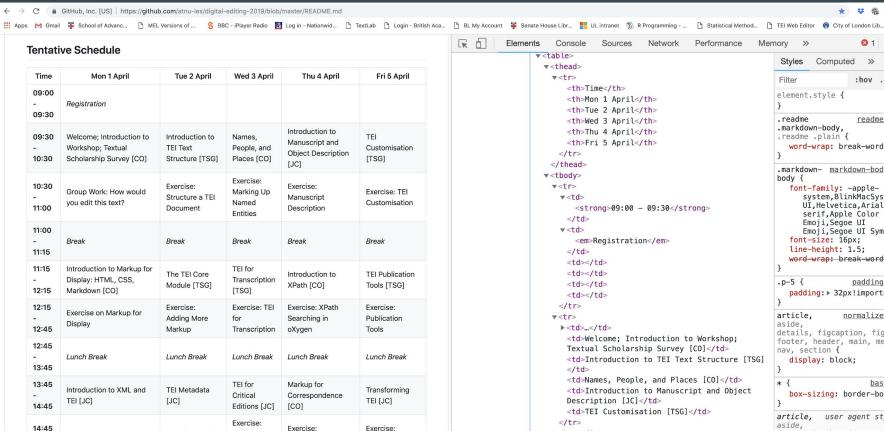
Example of Markdown > HTML: A GitHub page



## Tentative Schedule	<h2 id="tentative-schedule">Tentative Schedule</h2>
Time Mon 1 April Tue 2 April Wed 3 April Thu 4 April Fri 5 April	thead>
	Time
	Mon 1 April
	Tue 2 April
	Wed 3 April
09:00 - 09:30 *Registration*	Thu 4 April
09:30 - 10:30 Welcome; Introduction to Workshop; Textual Scholarship	Fri 5 April
Survey [CO] Introduction to TEI Text Structure [TSG] Names, People, and	
Places [CO] Introduction to Manuscript and Object Description [JC] TEI	
Customisation [TSG]	
10:30 - 11:00 Group Work: How would you edit this text? Exercise:	
Structure a TEI Document Exercise: Marking Up Named Entities Exercise:	09:00 - 09:30
Manuscript Description Exercise: TEI Customisation	Registration
11:00 - 11:15 *Break* *Break* *Break* *Break*	<
11:15 - 12:15 Introduction to Markup for Display: HTML, CSS, Markdown	<
[CO] The TEI Core Module [TSG] TEI for Transcription [TSG] Introduction	<
to XPath [CO] TEI Publication Tools [TSG]	<
12:15 - 12:45 Exercise on Markup for Display Exercise: Adding More	
Markup Exercise: TEI for Transcription Exercise: XPath Searching in oXygen	
Exercise: Publication Tools	09:30 - 10:30
12:45 - 13:45 *Lunch Break* *Lunch Break* *Lunch Break* *Lunch	Welcome; Introduction to Workshop; Textual Scholarship Survey [CO]
Break* *Lunch Break*	Introduction to TEI Text Structure [TSG]
13:45 - 14:45 Introduction to XML and TEI [JC] TEI Metadata [JC]	Names, People, and Places [CO]
TEI for Critical Editions [JC] Markup for Correspondence [CO] Transforming	Introduction to Manuscript and Object Description [JC]
TEI [JC]	TEI Customisation [TSG]
14:45 - 15:15 Exercise: XML Markup Exercise: Enrich Your TEI Header	
Exercise: Critical Apparatus Markup Exercise: Correspondence Markup	
Exercise: Transform Your TEI	10:30 - 11:00
15:15 - 15:30 *Break* *Break* *Break* *Break*	Group Work: How would you edit this text?
15:30 - 16:30 Case Study: Mark Twain, April Fool: TEI Correspondence,	Exercise: Structure a TEI Document
Personography, and Network Graphs [CO] Case Study: Encoding Ancient	Exercise: Marking Up Named Entities

Markdown

HTML



Exercise: Enrich

Your TEI Header

Exercise: XML Markup

15:15

Critical

Apparatus

Correspondence

Markup

Transform Your

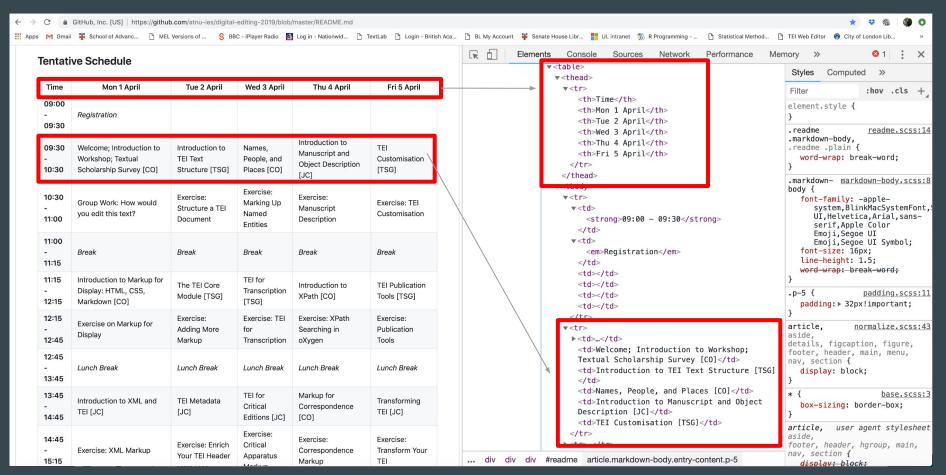
TEI

Elements Console Sources Network Performance Me	emory » 81 : >
▼ <thead></thead>	Styles Computed >>
▼	Filter :hov .cls -
Time	
Mon 1 April	element.style {
Tue 2 April	}
Wed 3 April	<pre>.readme</pre> <pre>readme.scss:</pre>
Thu 4 April	<pre>.markdown-body, .readme .plain {</pre>
Fri 5 April	word-wrap: break-word;
	}
	,
▼	<pre>.markdown- markdown-body.scs: body {</pre>
▼	font-family: -apple-
▼>	system,BlinkMacSystemFor
09:00 - 09:30	UI, Helvetica, Arial, sans-
	serif,Apple Color Emoji,Segoe UI
▼>	Emoji, Segoe UI Symbol;
<pre>Registration</pre>	font-size: 16px;
	line-height: 1.5;
<	word-wrap: break-word;
	}
<	.p-5 { padding.scss
<	<pre>padding: ▶ 32px!important;</pre>
	}
▼	article, <u>normalize.scss</u>
▶	aside, details, figcaption, figure,
Welcome; Introduction to Workshop;	footer, header, main, menu,
Textual Scholarship Survey [CO]	nav, section {
Introduction to TEI Text Structure [TSG]	display: block;
	}
>td>Names, People, and Places [CO]	* { base.scs
Introduction to Manuscript and Object	<pre>box-sizing: border-box;</pre>
Description [JC]	}
TEI Customisation [TSG]	article, user agent styleshe
	aside,
▶	footer, header, hgroup, main, nav, section {
div div div #readme article.markdown-body.entry-content.p-5	display: block:

Markdown

HTML

k CSS



Learn more

READINGS (all available on the GitHub site as PDFs)

- GOLDFARB, CHARLES F. "The Roots of SGML: A Personal Recollection." *Technical Communication* 46, no. 1 (1999): 75-78. www.istor.org/stable/43088604.
- HUITFELDT, CLAUS. "Markup Technology and Textual Scholarship." In *Digital Critical Editions*, edited by Apollon Daniel, Bélisle Claire, and Régnier Philippe, 157-78. University of Illinois Press, 2014. www.istor.org/stable/10.5406/j.ctt6wr6r8.10.
- SCHLOEN, DAVID and SANDY SCHLOAN. "Beyond Gutenberg: Transcending the Document Paradigm in Digital Humanities", Digital Humanities Quarterly 8.4 (2014). http://digitalhumanities.org//dhq/vol/8/4/000196/000196.html.

MORE ON MARKDOWN

- Cheatsheet: https://guides.github.com/pdfs/markdown-cheatsheet-online.pdf
- Tutorial: https://commonmark.org/help/tutorial/