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). It is not strictly a markup language but a formal grammar for the design and specification of markup languages.

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).

1999: World Wide Web Consortium (W3C) superseded the IETF, published the HTML 4 standards.

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):
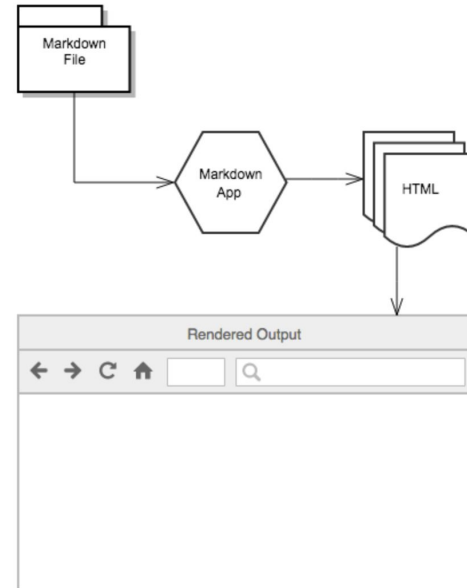
```
<measure unit=" pounds" value=" -1000">
One thousand pounds in debt</measure>
```

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/station-metro)]

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.

Some Markdown authoring and editing applications

- Mac: Texts, iA Writer, Ulysses, Marked, or MacDown
- iOS / Android: iA Writer or Ulysses (iOS only)
- Windows: Texts, MarkdownPad, Typora
- Linux: Typora
- Web: Dillinger or StackEdit
- Looking to self-publish a book? Leanpub, which converts Markdown files into an electronic book (PDF, EPUB, and MOBI file format); Bookdown (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

GitHub, Inc. [US] | https://github.com/atnu-ies/digital-editing-2019/blob/master/README.md

Apps | Gmail | School of Advanc... | MEL Versions of ... | BBC - iPlayer Radio | Log in – Nationwid... | TextLab | Login – British Aca... | BL My Account | Senate House Libr... | UL intranet | R Programming – ... | Statistical Method... | TEI Web Editor | City of London Lib...

Textual Editing in the Digital Age

A 5-day training event organised by Animating Texts Newcastle University and the Institute of English Studies, University of London

Welcome to the repo for the "Textual Editing in the Digital Age" training event (1-5 April 2019) at Newcastle University.

The sessions will be convened by James Cummings (Newcastle, Text Encoding Initiative) [JC], Tiago Sousa Garcia (Newcastle) [TSG], and Christopher Ohge (London) [CO], with guest tutors Charlotte Tupman (Exeter) [CT] and Francesca Benatti (Open U) [FB]. The training will consist of alternating lectures and practical workshop sessions on textual editing with XML technologies, including a thorough grounding in TEI as well as introductions to XPath searching and TEI publication tools. Also planned are presentations of case studies, a visit to Newcastle Library's Special Collections (to hear about their interesting digital editing projects). There may also be unofficial social visits to the pub some evenings.

Registration is open to all students and staff, but preference will be given to MA and PhD students and early career staff from Northern England, Northern Ireland, and Scotland. There will be no charge for the workshop.

Applications for this workshop are now closed.

If you have any questions, please don't hesitate to contact us at animating.text@newcastle.ac.uk.

Venue

The workshop will take place at Newcastle University, at the Armstrong Building room 2.96. A map of its location and a selection of nearby places to eat lunch is available here.

Elements | Console | Sources | Network | Performance | Memory

```html
<!doctype html>
<html lang="en">
▶<head>…</head>
▼<body class="logged-out env-production emoji-size-boost page-blob">
   ▶<div class="position-relative js-header-wrapper ">…</div>
   <div id="start-of-content" class="show-on-focus"></div>
   <div id="js-flash-container">

   </div>
   ▼<div class="application-main " data-commit-hovercards-enabled>
      ▼<div itemscope itemtype="http://schema.org/SoftwareSourceCode" class>
         ▼<main id="js-repo-pjax-container" data-pjax-container>
            ▶<div class="pagehead repohead instapaper_ignore readability-menu experiment-repo-nav "
            </div>
            ▼<div class="container new-discussion-timeline experiment-repo-nav ">
               ::before
               ▼<div class="repository-content ">
                  <a class="d-none js-permalink-shortcut" data-hotkey="y" href="/atnu-ies/digital-editing-2019/blob/13e430e52e34461ac05a6957a7eeb89bb342c3a8/README.md">Permalink</a>
                  <!-- blob contrib key: blob_contributors:v21:86db389c5dc66ad4c6aa1215e930d377 -->
                  ▶<div class="signup-prompt-bg rounded-1">…</div>
                  ▶<div class="d-flex flex-shrink-0 flex-items-center mb-3">…</div>
                  ▶<div class="Box Box--condensed d-flex flex-column flex-shrink-0">…</div>
                  ▼<div class="Box mt-3 position-relative">
                     ▼<div class="Box-header py-2 d-flex flex-justify-between flex-items-center">
                        ▶<div class="text-mono f6">…</div>
                        ▶<div class="d-flex">…</div>
```

Tentative Schedule

| Time | Mon 1 April | Tue 2 April | Wed 3 April | Thu 4 April | Fri 5 April |
|---------------|--
---	-----------------------
--------------------------------------	-----------------------------------
09:00 - 09:30	*Registration*				
09:30 - 10:30	Welcome; Introduction to Workshop; Textual Scholarship
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:
Manuscript Description	Exercise: TEI Customisation
11:00 - 11:15	*Break*	*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
12:45 - 13:45	*Lunch Break*	*Lunch Break*	*Lunch Break*	*Lunch
Break*	*Lunch Break*
13:45 - 14:45	Introduction to XML and TEI [JC]	TEI Metadata [JC]
TEI for Critical Editions [JC]	Markup for Correspondence [CO]	Transforming
TEI [JC]
14:45 - 15:15	Exercise: XML Markup	Exercise: Enrich Your TEI Header
Exercise: Critical Apparatus Markup	Exercise: Correspondence Markup
Exercise: Transform Your TEI
15:15 - 15:30	*Break*	*Break*	*Break*	*Break*	*Break*
15:30 - 16:30	Case Study: Mark Twain, April Fool: TEI Correspondence,
Personography, and Network Graphs [CO]	Case Study: Encoding Ancient
```

```html
<h2 id="tentative-schedule">Tentative Schedule</h2>
<table>
<thead>
<tr class="header">
<th>Time</th>
<th>Mon 1 April</th>
<th>Tue 2 April</th>
<th>Wed 3 April</th>
<th>Thu 4 April</th>
<th>Fri 5 April</th>
</tr>
</thead>
<tbody>
<tr class="odd">
<td><strong>09:00 - 09:30</strong></td>
<td><em>Registration</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr class="even">
<td><strong>09:30 - 10:30</strong></td>
<td>Welcome; Introduction to Workshop; Textual Scholarship Survey [CO]</td>
<td>Introduction to TEI Text Structure [TSG]</td>
<td>Names, People, and Places [CO]</td>
<td>Introduction to Manuscript and Object Description [JC]</td>
<td>TEI Customisation [TSG]</td>
</tr>
<tr class="odd">
<td><strong>10:30 - 11:00</strong></td>
<td>Group Work: How would you edit this text?</td>
<td>Exercise: Structure a TEI Document</td>
<td>Exercise: Marking Up Named Entities</td>

# Markdown  >  HTML  &  CSS

GitHub, Inc. [US] | https://github.com/atnu-ies/digital-editing-2019/blob/master/README.md

Apps | Gmail | School of Advanc... | MEL Versions of ... | BBC - iPlayer Radio | Log in - Nationwid... | TextLab | Login - British Aca... | BL My Account | Senate House Libr... | UL intranet | R Programming - ... | Statistical Method... | TEI Web Editor | City of London Lib...

## Tentative Schedule

| Time | Mon 1 April | Tue 2 April | Wed 3 April | Thu 4 April | Fri 5 April |
|------|-------------|-------------|-------------|-------------|-------------|
| 09:00 - 09:30 | *Registration* | | | | |
| 09:30 - 10:30 | Welcome; Introduction to Workshop; Textual Scholarship 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: Manuscript Description | Exercise: TEI Customisation |
| 11:00 - 11:15 | *Break* | *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 |
| 12:45 - 13:45 | *Lunch Break* | *Lunch Break* | *Lunch Break* | *Lunch Break* | *Lunch Break* |
| 13:45 - 14:45 | Introduction to XML and TEI [JC] | TEI Metadata [JC] | TEI for Critical Editions [JC] | Markup for Correspondence [CO] | Transforming TEI [JC] |
| 14:45 - 15:15 | Exercise: XML Markup | Exercise: Enrich Your TEI Header | Exercise: Critical Apparatus Markup | Exercise: Correspondence Markup | Exercise: Transform Your TEI |

Elements | Console | Sources | Network | Performance | Memory

```
▼<table>
  ▼<thead>
    ▼<tr>
        <th>Time</th>
        <th>Mon 1 April</th>
        <th>Tue 2 April</th>
        <th>Wed 3 April</th>
        <th>Thu 4 April</th>
        <th>Fri 5 April</th>
      </tr>
    </thead>
  ▼<tbody>
    ▼<tr>
      ▼<td>
          <strong>09:00 - 09:30</strong>
        </td>
      ▼<td>
          <em>Registration</em>
        </td>
        <td></td>
        <td></td>
        <td></td>
        <td></td>
      </tr>
    ▼<tr>
      ▶<td>…</td>
        <td>Welcome; Introduction to Workshop; Textual Scholarship Survey [CO]</td>
        <td>Introduction to TEI Text Structure [TSG]</td>
        <td>Names, People, and Places [CO]</td>
        <td>Introduction to Manuscript and Object Description [JC]</td>
        <td>TEI Customisation [TSG]</td>
      </tr>
    ▶<tr>…</tr>
```

Styles | Computed

Filter | :hov | .cls

```
element.style {
}

.readme                readme.scss:14
.markdown-body,
.readme .plain {
    word-wrap: break-word;
}

.markdown-        markdown-body.scss:8
body {
    font-family: -apple-
        system,BlinkMacSystemFont,
        UI,Helvetica,Arial,sans-
        serif,Apple Color
        Emoji,Segoe UI
        Emoji,Segoe UI Symbol;
    font-size: 16px;
    line-height: 1.5;
    word-wrap: break-word;
}

.p-5 {                   padding.scss:11
    padding: ▶ 32px!important;
}

article,               normalize.scss:43
aside,
details, figcaption, figure,
footer, header, main, menu,
nav, section {
    display: block;
}

* {                        base.scss:3
    box-sizing: border-box;
}

article,     user agent stylesheet
aside,
footer, header, hgroup, main,
nav, section {
    display: block;
}
```

... | div | div | div | #readme | article.markdown-body.entry-content.p-5

Markdown > HTML & CSS

GitHub, Inc. [US] | https://github.com/atnu-ies/digital-editing-2019/blob/master/README.md

Apps · Gmail · School of Advanc... · MEL Versions of ... · BBC - iPlayer Radio · Log in - Nationwid... · TextLab · Login - British Aca... · BL My Account · Senate House Libr... · UL intranet · R Programming - ... · Statistical Method... · TEI Web Editor · City of London Lib...

Tentative Schedule

Time	Mon 1 April	Tue 2 April	Wed 3 April	Thu 4 April	Fri 5 April
09:00 - 09:30	*Registration*				
09:30 - 10:30	Welcome; Introduction to Workshop; Textual Scholarship 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: Manuscript Description	Exercise: TEI Customisation
11:00 - 11:15	*Break*	*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
12:45 - 13:45	*Lunch Break*	*Lunch Break*	*Lunch Break*	*Lunch Break*	*Lunch Break*
13:45 - 14:45	Introduction to XML and TEI [JC]	TEI Metadata [JC]	TEI for Critical Editions [JC]	Markup for Correspondence [CO]	Transforming TEI [JC]
14:45 - 15:15	Exercise: XML Markup	Exercise: Enrich Your TEI Header	Exercise: Critical Apparatus Markup	Exercise: Correspondence Markup	Exercise: Transform Your TEI

DevTools panels

Elements · Console · Sources · Network · Performance · Memory

```
▼<table>
  ▼<thead>
    ▼<tr>
        <th>Time</th>
        <th>Mon 1 April</th>
        <th>Tue 2 April</th>
        <th>Wed 3 April</th>
        <th>Thu 4 April</th>
        <th>Fri 5 April</th>
      </tr>
  </thead>
  ▼<tbody>
    ▼<tr>
      ▼<td>
          <strong>09:00 - 09:30</strong>
        </td>
      ▼<td>
          <em>Registration</em>
        </td>
        <td></td>
        <td></td>
        <td></td>
        <td></td>
      </tr>
    ▼<tr>
      ▶<td>…</td>
        <td>Welcome; Introduction to Workshop; Textual Scholarship Survey [CO]</td>
        <td>Introduction to TEI Text Structure [TSG]</td>
        <td>Names, People, and Places [CO]</td>
        <td>Introduction to Manuscript and Object Description [JC]</td>
        <td>TEI Customisation [TSG]</td>
      </tr>
    ▶<tr>…</tr>
```

Styles · Computed

Filter :hov .cls +

```
element.style {
}

.readme                          readme.scss:14
.markdown-body,
.readme .plain {
    word-wrap: break-word;
}

.markdown-         markdown-body.scss:8
body {
    font-family: -apple-
      system,BlinkMacSystemFont,
      UI,Helvetica,Arial,sans-
      serif,Apple Color
      Emoji,Segoe UI
      Emoji,Segoe UI Symbol;
    font-size: 16px;
    line-height: 1.5;
    word-wrap: break-word;
}

.p-5 {                           padding.scss:11
    padding: ▶ 32px!important;
}

article,                         normalize.scss:43
aside,
details, figcaption, figure,
footer, header, main, menu,
nav, section {
    display: block;
}

* {                              base.scss:3
    box-sizing: border-box;
}

article,    user agent stylesheet
aside,
footer, header, hgroup, main,
nav, section {
    display: block;
}
```

... div div div #readme article.markdown-body.entry-content.p-5

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.jstor.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.jstor.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://commonmark.org/help/
- Tutorial: https://commonmark.org/help/tutorial/