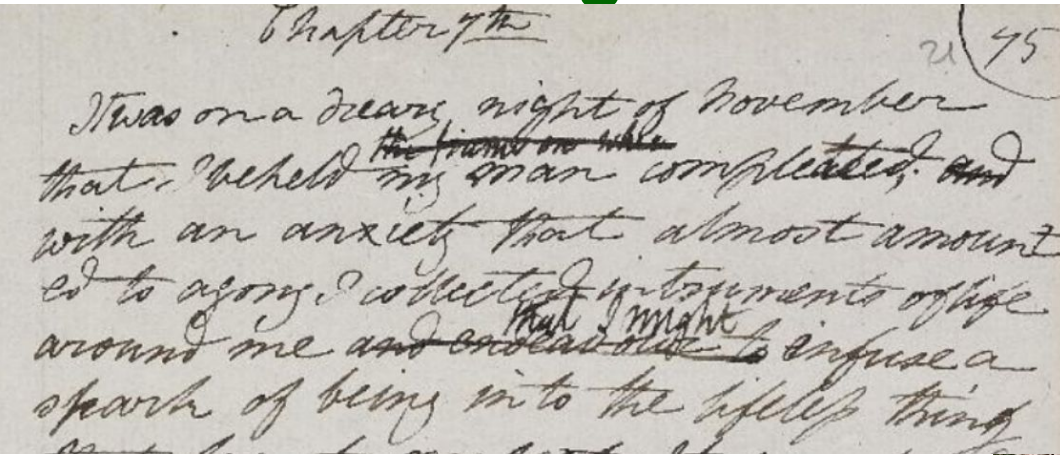


Stand-off Bridges in the Frankenstein Variorum



Interchange and Interoperability...with TEI?

Link to these slides: <http://bit.ly/frankenStandOff>

Elisa Beshero-Bondar (@epyllia) and Raffaele Viglianti (@raffazizzi)

[Balisage 2018](#) Wednesday 1 August @ 4pm

The Frankenstein Variorum: who we are / where we work

GitHub: https://github.com/PghFrankenstein/Pittsburgh_Frankenstein

Textual Editing, TEI, Variorum Interface

- Elisa Beshero-Bondar (Pitt-Greensburg)
- Raffaele Viglianti (MITH)
- Rikk Mulligan (Carnegie Mellon)

Contextual Annotations

- Jon Klancher (Carnegie Mellon)
- Jack Quirk (Carnegie Mellon)
- Avery Wiscomb (Carnegie Mellon)

GIS and Data Visualization

- Emma Slayton (Carnegie Mellon)
- Rikk Mulligan (Carnegie Mellon)
- Scott Weingart (Carnegie Mellon)
- Avery Wiscomb (Carnegie Mellon)

Project Interface

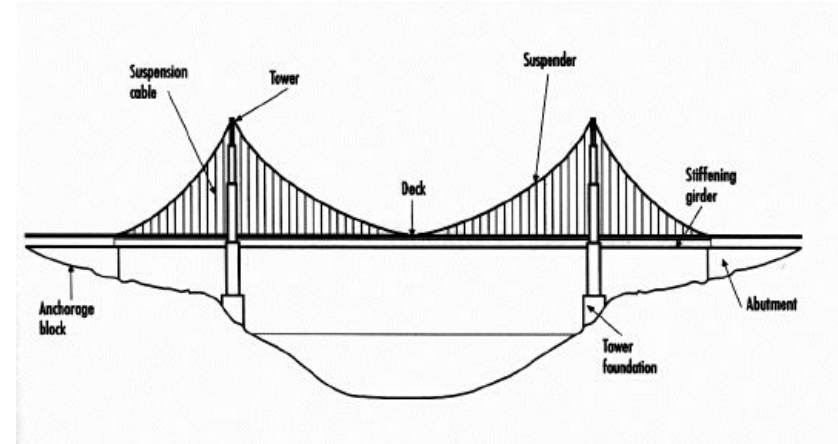
- Rikk Mulligan (Carnegie Mellon)
- Elisa Beshero-Bondar (Pitt-Greensburg)
- Raffaele Viglianti (MITH)
- Emma Slayton (Carnegie Mellon)
- Scott Weingart (Carnegie Mellon)

Frankenstein Variorum: Motivating Questions:

- Can we make an edition that conveniently compares the manuscripts to the print publications of *Frankenstein*?
- Can we make a comprehensive “collation” to show changes to the novel over time, from 1816 to 1831?
 - What parts of the novel were most mutable?
 - Which editorial interventions persist from 1816 to 1831?
- Can we build on and work with existing digital editions of *Frankenstein*, instead of “reinventing the wheel”?

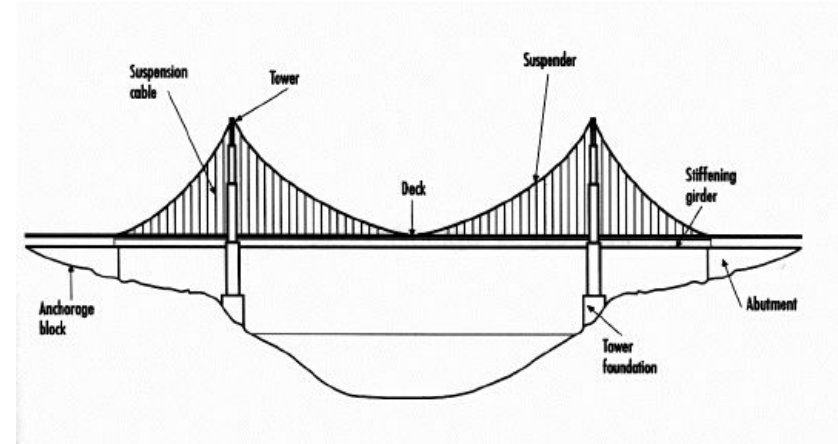
Bridge Building for Interchange and Interoperation

- *Frankenstein* as site of experiments with computers and texts
- Opportunity to **interoperate**?
 - Unmediated computer interpretation
 - Prepare code to be parsed by machines in the same way
- Opportunity for **interchange**?
 - Human exchange of ideas
 - facilitated by legible, systematic, documented code
 - Blind interchange: don't need to meet the coder!
- [Syd Bauman, “Interchange vs. Interoperability” \(Balisage 2011\)](#)



Bridge Building for Interchange and Interoperation

- **Computer-assisted collation:**
 - rebuilding disparate code for interoperation
 - Requires retooling "bridge formats" of code
- **Variorum edition with standoff pointers:**
 - rebuilding again to point back to the originals
 - Respects, supports, requires interchangeability

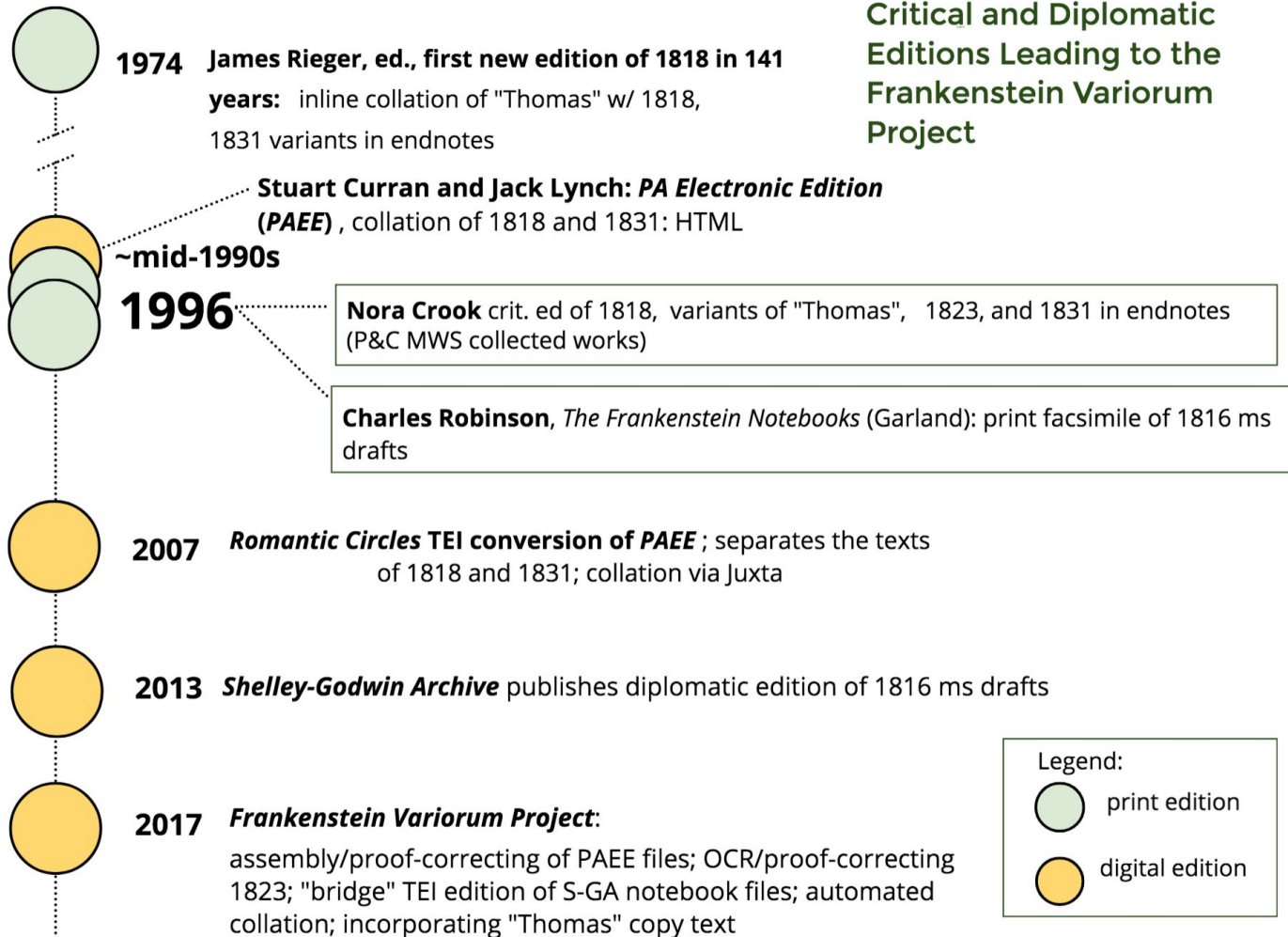


Frankenstein's Variations



- **1816 Draft MS-Notebooks:** 3 boxes at the Bodleian Library:
- **1818 Edition:** Lackington: 3 volumes (from the [PA Electronic Edition, 1994](#))
- **"Thomas copy":** hand-written revisions by MWS in a copy of 1818 (new)
- **1823 Edition:** edited from 1818 ed. by William Godwin (new)
- **1831 Edition** (1/2 of a volume): bound with Friedrich von Schiller's *The Ghost Seer* in Bentley's Standard Series of novels (from the [PA Electronic Edition, 1994](#))

Critical and Diplomatic Editions Leading to the Frankenstein Variorum Project



PA Electronic Edition: hypertext collation experiment

1818

Chapter

Contents

Index

Frame

1831

CHAPTER II.

WHEN I had attained the age of seventeen, my parents resolved that I should become a student at the university of Ingolstadt. I had hitherto attended the schools of Geneva; but my father thought it necessary, for the completion of my education, that I should be made acquainted with other customs than those of my native country. My departure was therefore fixed at an early date; but, before the day resolved upon could arrive, the first misfortune of my life occurred -- an omen, as it were, of my future misery.

Elizabeth had caught the scarlet fever; but her illness was not severe, and she quickly recovered. During her confinement, many arguments had been urged to persuade my mother to refrain from attending upon her. She had, at first, yielded to our entreaties; but when she heard that her favourite was recovering, she could no longer debar herself from her society, and entered her chamber long before the danger of infection was past.

CHAPTER III.

WHEN I had attained the age of seventeen, my parents resolved that I should become a student at the university of Ingolstadt. I had hitherto attended the schools of Geneva; but my father thought it necessary, for the completion of my education, that I should be made acquainted with other customs than those of my native country. My departure was therefore fixed at an early date; but, before the day resolved upon could arrive, the first misfortune of my life occurred -- an omen, as it were, of my future misery.

Elizabeth had caught the scarlet fever; her illness was severe, and she was in the greatest danger. During her illness, many arguments had been urged to persuade my mother to refrain from attending upon her. She had, at first, yielded to our entreaties; but when she heard that the life of her favourite was menaced, she could no longer control her anxiety. She attended her sick bed -- her watchful attentions triumphed over the malignity of the

CHAPTER [II.] <III>

WHEN I had attained the age of seventeen, my parents resolved that I should become a student at the university of Ingolstadt. I had hitherto attended the schools of Geneva; but my father thought it necessary, for the completion of my education, that I should be made acquainted with other customs than those of my native country. My departure was therefore fixed at an early date; but, before the day resolved upon could arrive, the first misfortune of my life occurred -- an omen, as it were, of my future misery.

Elizabeth had caught the scarlet fever; **[but]** her illness was **[not]** severe, and she **[quickly recovered]** <was in the greatest danger>. During her **[confinement]** <illness>, many arguments had been urged to persuade my mother to refrain from attending upon her. She had, at first, yielded to our entreaties; but when she heard that <the life of> her favourite was **[recovering, she could no longer debar herself from her society, and entered her chamber long before the danger of infection was past. The]** <menaced, she could no longer control her anxiety. She attended her sick bed, -- her watchful attentions triumphed over the malignity of the distemper, -- Elizabeth was saved, but the> consequences of this imprudence were fatal <to her preserver>. On the third day my mother sickened: her fever was **[very malignant]** <accompanied by the most alarming

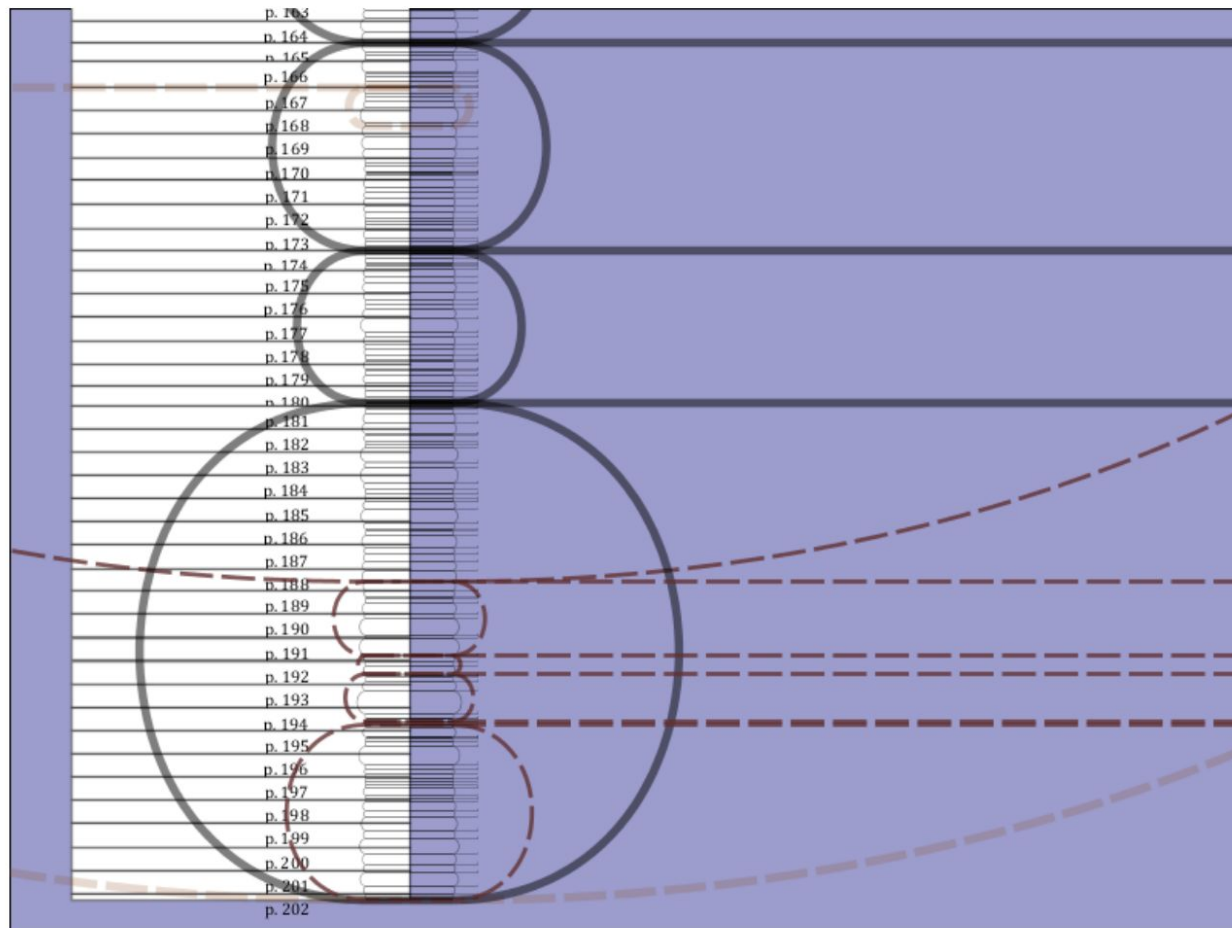
hundreds of small HTML files juxtaposed in frames

LMNL *Frankenstein*

Frankenstein as
locus for digital
experiments on
textual bodies

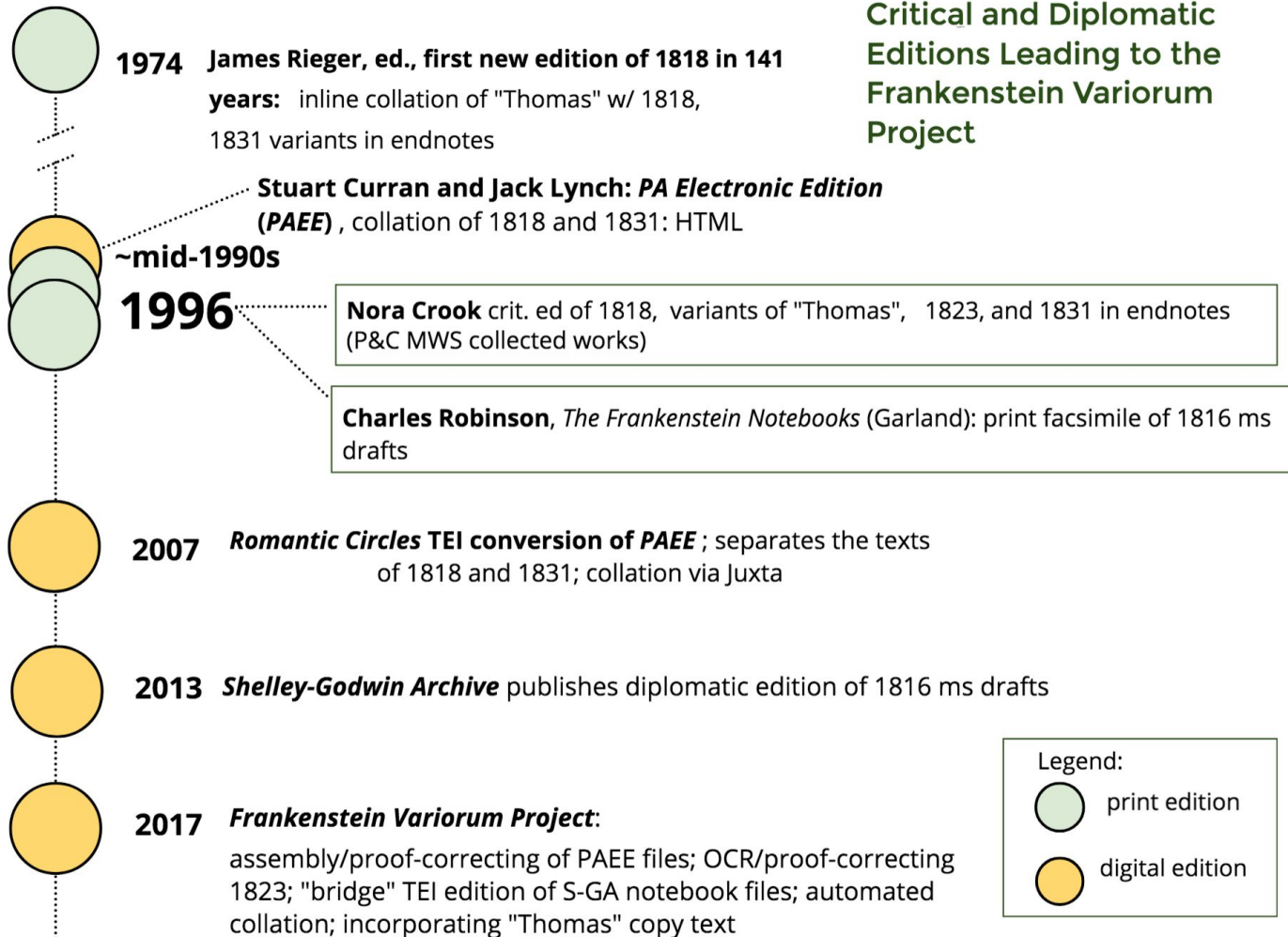
- [Wendell Piez,](#)
“[Hierarchies within](#)
[range space](#)”
([Balisage 2014](#))

Figure 7: Overlap in Chapter 24



This closeup of the map of the novel shows where Chapter 24 overlaps with the framing structures (the four letters from the main narrator, Walton, to his sister) and the entries in Letter IV.

Critical and Diplomatic Editions Leading to the Frankenstein Variorum Project



Shelley-Godwin Notebooks: c56 (p.11)

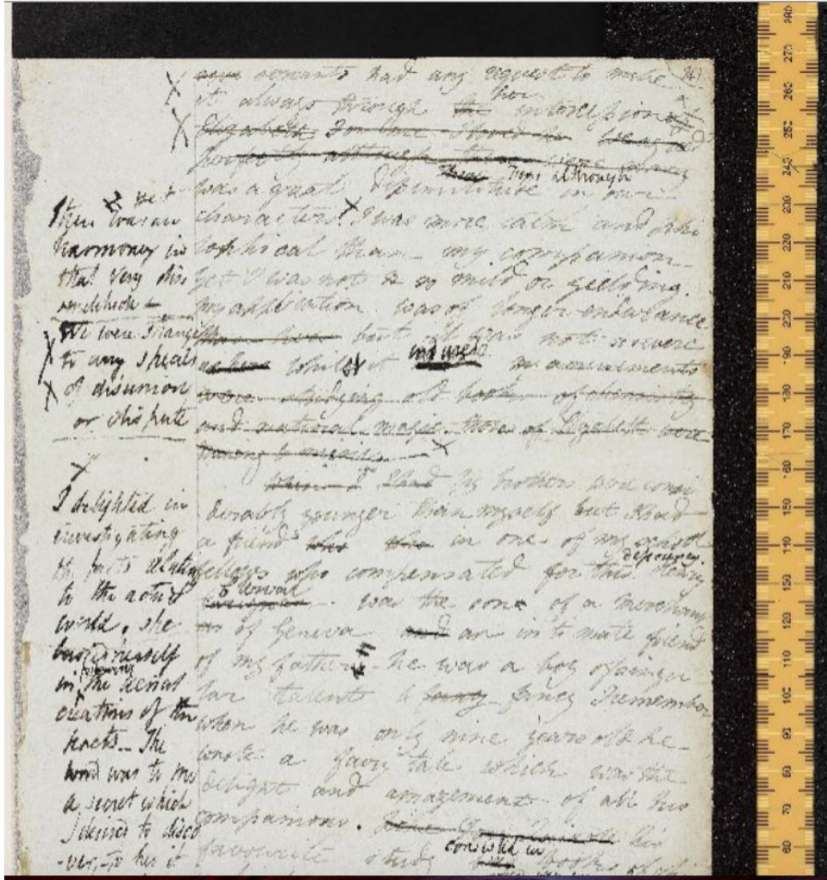


Search



LIMIT VIEW:

☒ All ☐ Mary Shelley ☐ Percy Shelley



X
X

X yet
there was an
harmony in
that very diss
imilitude—

X We were strangers
to any species
X of disunion
or dispute

X

sevr servants had any request to make

her
it always through the intercession of
Elizabeth For me I loved he We agreed
perfectly although there were many

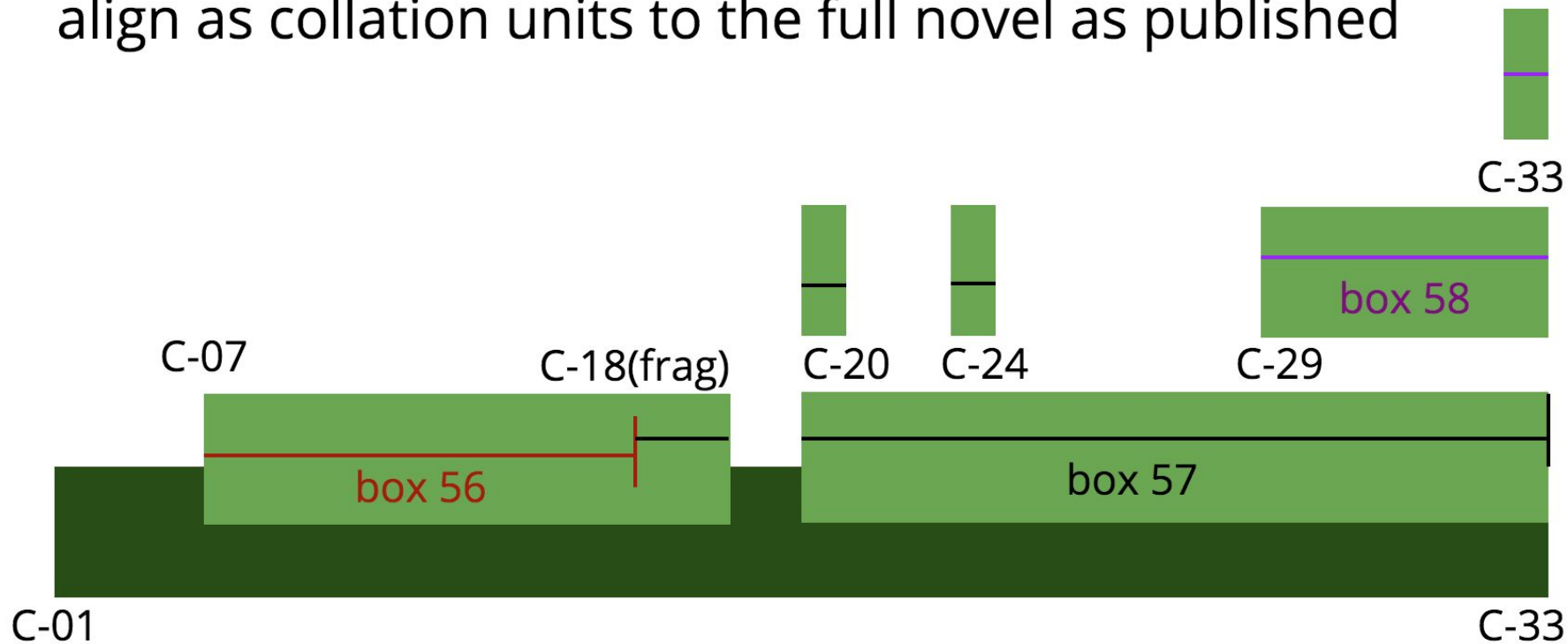
There For, although
was a great dissimilitude in our
characters. X I was more calm and phi
losphical than my companion
Yet I was not n so mild or yielding.
My application was of longer endurance
than hers but it was not so severe

endured
as hers whilst it lasted my amusements
were studying old books of chemistry
and natural magic those of Elizabeth were
X
dra wing & music.

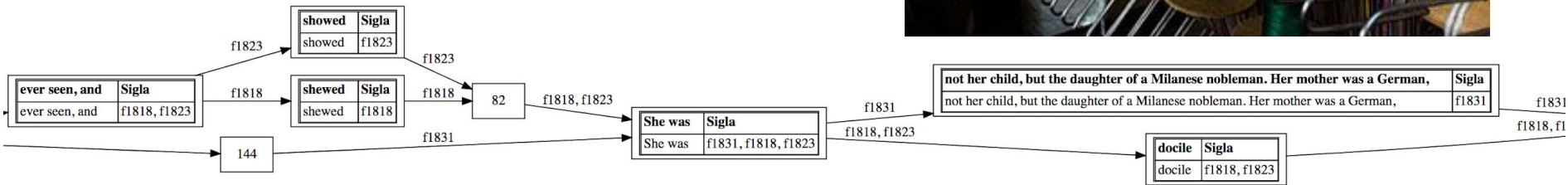
When I had My brothers were consi
derably younger than myself but I had
a friend who who in one of my school
deficiency.
fellows who compensated for this. Henry

Clerval
Carignan was the sons of a merchant
an of Geneva and an intimate friend

How the Bodleian Library *Frankenstein* ms notebooks align as collation units to the full novel as published



Collation is weaving...



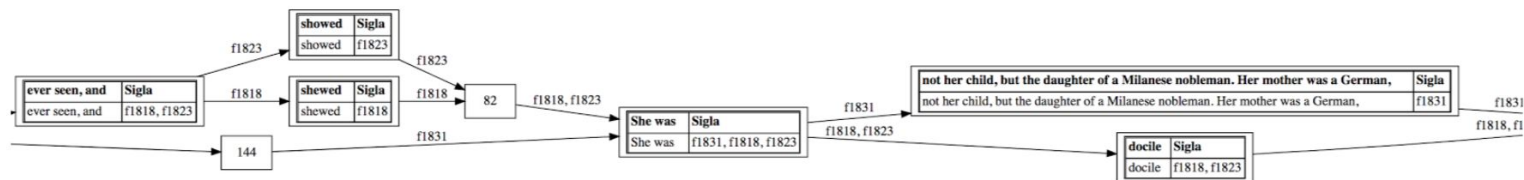
...Can we weave diplomatic ms encoding into a variorum apparatus representing print editions?

Weaving trees . . .



```
...running stream <milestone ana="start"
/> stream of <lb ana="start"/> sequential text <lb ana="end"> nodes
to collate with markup sign-posts</milestone ana="end"> along the way...
```

```
...Another running stream <milestone ana="start"
/> stream of <del ana="start"/> sequential text <del ana="end"> nodes
to collate with markup sign-posts</milestone ana="end"> along the way...
```



Computer-aided collation: Gothenburg Model

Gothenburg model : algorithm for computer-assisted collation, developed in 2009 workshop of collateX and Juxta developers.

1. Tokenization

- (words, characters)

2. Normalization

- ('&' = 'and')

3. Alignment

- Identify comparable divergence: what makes text sequences comparable units?
- Macro and micro scale: organize text into comparable passages (paragraphs/chapters/strings)
- What differences should be ignored as "noise"?

4. Analysis

- (study output, correct, and re-align after machine process, AND refine automated processing)

5. Visualization

- critical edition apparatus, graph displays

Inside the aligned "chunks", flattening markup for collation

collateX: powerful tool for computational alignment:

- Permits normalization of word/character tokens (& = and)
- From multiple documents, produces one output of the original source tokens
- XML tags are read as text tokens, converted to JSON lists

`<markup/>`  `<markup/>`

JSON input to collateX: tokens, and normalized tokens

```
{'id': 'fMS', 'tokens': [{'t': 'dreary ', 'n': 'dreary'}, {'t': 'night ', 'n': 'night'}, {'t': 'of ', 'n': 'of'}, {'t': 'November ', 'n': 'november'}, {'t': ' <lb n="c56-0045__main__3"/> ', 'n': ' '}, {'t': 'that ', 'n': 'that'}, {'t': 'I ', 'n': 'i'}, {'t': 'beheld ', 'n': 'beheld'}, {'t': ' <del loc="c56-0045__main__d1e8429__Start"/> the ', 'n': ' the'}, {'t': 'frame ', 'n': 'frame'}, {'t': 'on ', 'n': 'on'}, {'t': 'whic <del loc="c56-0045__main__d1e8429__End"/> ', 'n': 'whic '}, {'t': 'my ', 'n': 'my'}, {'t': 'man ', 'n': 'man'}, {'t': 'completeed,. ', 'n': 'completeed,.'}]}
```

collateX reads and aligns the *normalized* tokens, but outputs the original tokens from the input documents

collateX XML critical apparatus output: preserves “flattened markup”

```
<app type="invariant">
  <rdg wit="#f1818">that I beheld the </rdg>
  <rdg wit="#f1823">that I beheld the </rdg>
  <rdg wit="#f1831">that I beheld the </rdg>
  <rdg wit="#fMS">that I beheld &lt;del loc="c56-0045__main__d1e8429__Start"/&gt;the </rdg>
  <rdg wit="#fThomas">that I beheld the </rdg>
</app>
<app>
  <rdg wit="#f1818">accomplishment of </rdg>
  <rdg wit="#f1823">accomplishment of </rdg>
  <rdg wit="#f1831">accomplishment of </rdg>
  <rdg wit="#fMS">frame on whic&lt;del loc="c56-0045__main__d1e8429__End"/&gt; </rdg>
  <rdg wit="#fThomas">accomplishment of </rdg>
</app>
<app type="invariant">
  <rdg wit="#f1818">my </rdg>
  <rdg wit="#f1823">my </rdg>
  <rdg wit="#f1831">my </rdg>
  <rdg wit="#fMS">my </rdg>
  <rdg wit="#fThomas">my </rdg>
</app>
```


Collation process: Challenges for TEI editions

Breaking hierarchies in interests of collation:

- Prepare “bridge editions”
 - Each edition must share comparable encoding with the others
- Pre-collation “chunking” and “flattening”:
 - Break the macro-superstructure (volumes/chapters):
Reconstruct around “alignment chunks”
 - 33 collation units (“chunk”) files for each edition in our Variorum, roughly divided at chapter boundaries
 - “Flatten” structural XML elements so collation process can **read around** hierarchies (e.g. paragraph boundaries)
 - Tokens must represent the “running stream” of sequential text
 - Collation process can align hierarchical boundaries as sign-posts, working with significant flattened tags.

Frankenstein “bridge” editions prepped for collation:

- Transform web 1.0 PA Electronic Edition (1818 and 1831) into TEI structural markup
- Give all editions the same elements for structural divisions
 - <milestone/> with attributes for volume, letter, and chapter divisions
 - <p>, <lg>, <l> for paragraphs and poetry
 - and <add> for ms editions
 - kept other markup from S-GA (surface/zone information is preserved **but not collated**)
- Flatten all the elements prior to collation
- Align collation "chunk" units
- Re-sequence S-GA encoding
 - Margin zone markup moved from the ends of page files into sequential position to give collateX a unified "running stream" of text
 - Automated by XSLT thanks to S-GA's precise locational markup

... transforming the collation output

- **collateX XML output:**
 - outputs each alignment area whether "equivalent" or not
 - adapted to output information about whether readings were "invariant"
 - supports building the "bridges" or "spine" architecture of our edition
- Up-convert separate individual editions from the collation "weave":
Stages:
 - XSLT: Rebuild individual editions, separating their <rdg> witnesses from the output
 - Apply **<seg>....</seg>** to identify the "hotspots" of variance
 - Reconstruct flattened elements (regular expression search and replace), and reconstruct hotspot <seg> elements to fit the edition's document hierarchy

* **Documentation/Tutorials for working with collateX and XML:** see <https://collatex.net/doc/#xml-output> and https://github.com/DiXiT-eu/collatex-tutorial/blob/master/unit5/3_collation-outputs.ipynb

Tracing the Concurrent Hierarchies: the "spine"

```
<app type="invariant">
  <rdg wit="#f1818">black </rdg>
  <rdg wit="#f1823">black </rdg>
  <rdg wit="#f1831">black </rdg>
  <rdg wit="#fMS">black </rdg>
  <rdg wit="#fThomas">black </rdg>
</app>
<app>
  <rdg wit="#f1818">lips.
  &lt;p eID="novell1_letter4_chapter4_div4_div4_p2"/&gt;
  &lt;p sID="novell1_letter4_chapter4_div4_div4_p3"/&gt;The </rdg>
    <rdg wit="#f1823">lips.&lt;p eID="novell1_letter4_chapter4_div4_div4_p2"/&gt;
  &lt;p sID="novell1_letter4_chapter4_div4_div4_p3"/&gt;The </rdg>
    <rdg wit="#f1831">lips.&lt;p eID="novell1_letter4_chapter5_div4_div5_p2"/&gt;
  &lt;p sID="novell1_letter4_chapter5_div4_div5_p3"/&gt;The </rdg>
    <rdg wit="#fMS">lips.&lt;milestone unit="tei:p"/&gt;&lt;
  lb n="c56-0046__main__3"/&gt;The</rdg>
    <rdg wit="#fThomas">lips. And the</rdg>
  </app>
<app>
  <rdg wit="#fThomas">contortions that ever and anon con
  &lt;del sID="novell1_letter4_chapter4_div4_div4_p2_add1_dell1"/&gt;puls&lt;
  del eID="novell1_letter4_chapter4_div4_div4_p2_add1_dell1"/&gt;vulsed &amp; deformed his un-human features.
  &lt;p eID="novell1_letter4_chapter4_div4_div4_p2"/&gt;
  &lt;p sID="novell1_letter4_chapter4_div4_div4_p3"/&gt;The </rdg>
  </app>
<app type="invariant" xml:id="C10_app93">
  <rdg wit="#f1818">different accidents of life are </rdg>
  <rdg wit="#f1823">different accidents of life are </rdg>
  <rdg wit="#f1831">different accidents of life are </rdg>
  <rdg wit="#fMS">different accidents of life are </rdg>
  <rdg wit="#fThomas">different accidents of life are </rdg>
</app>
```

Tracing the Concurrent Hierarchies: (Thomas ed.)

```
<p xml:id="novell_letter4_chapter4_div4_div4_p2">
  <seg part="F" xml:id="C10_app44-fThomas_Pt2">How </seg>can I describe my <seg
    xml:id="C10_app46-fThomas">emotions </seg>at this <seg xml:id="C10_app48-fThomas"
      >catastrophe, </seg>or how delineate the wretch whom with such infinite pains and
      care I had endeavoured to <seg xml:id="C10_app54-fThomas">form? </seg>His limbs were
      in <seg xml:id="C10_app56-fThomas">proportion, </seg>and I had selected his features
      as <seg xml:id="C10_app60-fThomas">beautiful. Beautiful!—Great </seg>God! His <seg
        xml:id="C10_app62-fThomas">yellow </seg>skin scarcely covered the work of muscles
        and arteries beneath; his hair <seg xml:id="C10_app66-fThomas">was of </seg>a
        lustrous <seg xml:id="C10_app68-fThomas">black, </seg>and <seg
          xml:id="C10_app70-fThomas">flowing; </seg>his teeth of a pearly <seg
            xml:id="C10_app72-fThomas">whiteness; </seg>but these <seg
              xml:id="C10_app74-fThomas">luxuriences </seg>only formed a more horrid contrast
              with his <seg xml:id="C10_app80-fThomas">watery eyes, </seg>that seemed almost of the
              same colour as the dun white sockets in which they were set, his shrivelled <seg
                xml:id="C10_app86-fThomas">complexion, </seg>
            <seg xml:id="C10_app87-fThomas">
              <del xml:id="novell_letter4_chapter4_div4_div4_p2_dell1">and</del>
            </seg>
            <seg xml:id="C10_app88-fThomas">straight </seg>black <seg xml:id="C10_app90-fThomas"
              >lips. And </seg>
            <seg xml:id="C10_app91-fThomas">the </seg>
            <seg part="I" xml:id="C10_app92-fThomas_Pt1">contortions that ever and anon con<del
              xml:id="novell_letter4_chapter4_div4_div4_p2_add1_dell1">puls</del>vulsed &amp;
              deformed his un-human features.</seg>
          </p>
<p xml:id="novell_letter4_chapter4_div4_div4_p3">
  <seg part="F" xml:id="C10_app92-fThomas_Pt2">The </seg>different accidents of life are
  not so changeable as the feelings of human nature. I had worked hard for
```


From CollateX to a TEI "spine"

1818 (chunk)

```
<p xml:id="p1">
```

I believed that exercise and
amusement <seg

xml:id="v1">would soon

drive away such

symptoms;</seg>

```
</p>
```

1823 (chunk)

```
<p xml:id="p1">
```

I believed that exercise and
amusement <seg

xml:id="v1">would soon

drive away such

symptoms;</seg>

```
</p>
```

1831 (chunk)

```
<p xml:id="p1">
```

I believed that exercise and
amusement would <seg

xml:id="v1">then drive away

incipient disease;</seg>

```
</p>
```

Collation of variants (generated from CollateX)

```
<app> ... </app>
<app>
  <rdgGrp n="1">
    <rdg wit="#P1818">
      <ptr target="P1818.xml#v1">
    </rdg>
    <rdg wit="#P1823">
      <ptr target="P1823.xml#v1">
    </rdg>
    <rdg wit="#MSC56">
      <ptr target="MSC56.xml#string-range(//line[13],0,21)"/>
      <ptr target="MSC56.xml#string-range(//line[14],5,12)"/>
      <ptr target="MSC56.xml#string-range(//line[15],0,5)"/>
    </rdg>
  </rdgGrp>
  <rdg wit="#P1831">
    <ptr target="P1831.xml#string-range(//p[@xml:id='p1'],39,73)"/>
  </rdg>
</app>
<app> ... </app>
```

MS Abinger c. 56 (chunk)

```
<line>I </line>
```

```
<line>believed that exercise and  
amusement</line>
```

```
<line>would soon drive
```

```
away</line>
```

```
<line><del
```

```
rend="strikethrough">these
```

```
</del><add hand="#pbs"
```

```
place="superlinear">such
```

```
</add>sym</line>
```

```
<line>ptoms</line>
```

MS Abinger c. 57

```
<TEI> .... </TEI>
```

MS Abinger c. 58

```
<TEI> .... </TEI>
```

Thomas edition

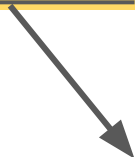
```
<TEI> .... </TEI>
```

Compare with: <https://goo.gl/G9Uku5>

Pointers to S-GA



We only point to stable files, by choosing a specific version (commit) of S-GA TEI data on GitHub (currently the latest commit)



```
<ptr  
target="https://raw.githubusercontent.com/umd-mith/sga/fa4b1db104df3214077c3675146576b3763ce22f/data/tei/ox/ox-ms_abinger_c56/ox-ms_abinger_c56-0045.xml#string-range(//tei:zone[@type='main']///tei:line[1],  
8,12)"/>
```

XPointer



- XPointer is not well supported out-of-the-box, but implementing string-range() is not too complex and worked well for pointing to S-GA data.
- TEI defines its own XPointer scheme (a superset of the official W3C recommendation that was not widely adopted)

[Guidelines 16.2.4](#)

```
<ptr  
target="https://raw.githubusercontent.com/umd-mith/sga/fa4b1db104df3214077c3675146576b3763ce22f/data/tei/ox/ox-ms_abinger_c56/ox-ms_abinger_c56-0045.xml#string-range(//tei:zone[@type='main']///tei:line[1],  
8,12)"/>
```

XPointer - whitespace




- We assume that `string-range()` would refer to normalized text (this is unclear in the TEI Guidelines)
- `xpath()` scheme: “When an XPath is interpreted by a TEI processor, the information set of the referenced document **is interpreted without any additional information** supplied by any schema processing that may or may not be present. In particular this means that **no whitespace normalization is applied** to a document **before** the XPath is interpreted.” [Guidelines 16.2.4.2](#)
- Does the same apply to `string-range()`?
- We assume whitespace normalization to simplify considerably the string range count

<ptr

target="https://raw.githubusercontent.com/umd-mith/sga/fa4b1db104df3214077c3675146576b3763ce22f/data/tei/ox/ox-ms_abinger_c56/ox-ms_abinger_c56-0045.xml#string-range(//tei:zone[@type='main']///tei:line[1],8,12)"/>

Building a first web publication of the variorum

Work in progress!

 [home](#) [Reader](#) [Compare](#) [About](#) [Team](#) [Next steps](#)

Frankenstein Variorum: Reader

[MS](#) [1818](#) [1823](#) [1831](#) [Thomas](#)

Source Edition: fMS

Choose “base text” to read

75

21

Chapter **7th**

It was on a dreary night of **November**

that I beheld ^{the **come on who**} my **man comple** **at te ed**. And

with an anxiety that almost amount

ed to **agony** I collected instruments of life

around **me** ^{that I might} **refuse** a

spark of being into the lifeless thin k g

that lay at my feet. It was already

one in the **morning**, the rain pattered

dismally against the window panes, &

my candle was nearly burnt out, **when**

This is the reader interface. You may view one chapter of one text. Highlighted text indicates words or passages that vary across texts. Click on a highlighted selection to view these variations below. Relevant annotations will also appear in the lower right window.


Variations

1818: IV.
1823: IV.
1831: V.
MS: 7th
Thomas: IV.

Annotations

There are no annotations in this group.

Variant readings from
other sources

TEI rendered with
CETELcean 

<https://goo.gl/d7Fei8>

The future is distributed, and hierarchies are bridgeable

- Our edition simulates a future in which digital editions can be re-used outside of their project scopes.
- New editions need not eclipse earlier ones, but build on and point to their work.
- TEI XML need not impose monolithic hierarchy. Instead, TEI can express negotiation.
 - Overlap is tractable for alignment.
 - Alignment can be an organizing principle that cuts across hierarchies
 - Requires "chunking" and atomizing hierarchies, finding the smallest meaningfully sharable semantic features
- Scholarly editions can construct hierarchies *designed for breaking* at their seams.

