Document Modeling with the TEI Critical Apparatus

A Panel for the TEI 2019 Conference in Graz, Austria

Link to these slides: http://bit.ly/crit-app-panel

Presenters: Hugh Cayless (@hcayless), Elisa Beshero-Bondar (@epyllia), Raffaele Viglianti (@raffazizzi)

Respondent: James Cummings (@jamescummings)

What is a Critical Apparatus, really?

Hugh Cayless (@hcayless)

What is a Critical Apparatus?

Latin: *apparatus criticus*, pl. *apparatūs critici*

- “Scholarly editions of texts…often record some or all of the known variations among different *witnesses* to the text.” —TEI Guidelines
- “[the apparatus]…records the work’s textual history over time” —Eggert (2007)
- “Editors are not always people who can be trusted, and critical apparatuses are provided so that readers are not dependent upon them.” —West (1973)

What is a Critical Apparatus?

A critical apparatus is the set of notes explaining an editor's (re)construction of a text. These notes may contain the readings of witnesses, conjectures not promoted to the text, explanatory notes, alternative spellings or punctuation, parallels from other works, and in general any information that might help a reader understand the background of the presented text.

What is a *TEI* Critical Apparatus?

> *A critical apparatus is the set of notes explaining an editor's (re)construction of a text.*

- In TEI, where these notes present alternate possibilities, they are modeled in such a way that they may be substituted for the readings in the default text.
- The <app>, <lem>, <rdg> structure places variants in parallel with the default readings.
- So in TEI, the apparatus is more than just notes, it is an actionable data structure.

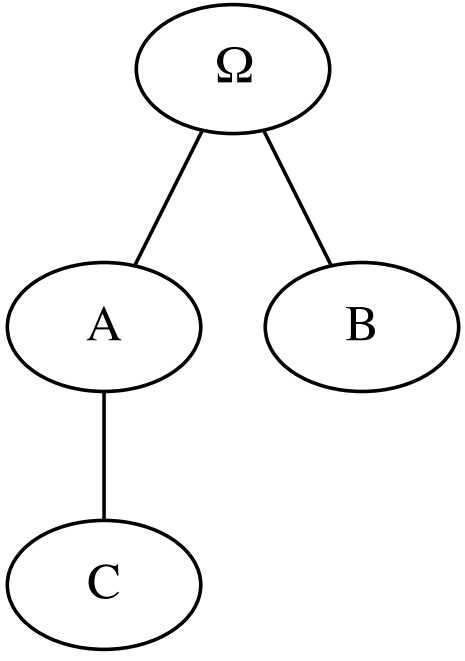
One view:

A TEI app. crit. represents a forking and rejoining of the text stream, a run of text for which there are multiple possibilities.

A: “The quick brown fox ju…”

B: “The quick brown mouse jumps over the lazy cat.”

C: “The quick brown cat jumps over the lazy dog.”



A: “The quick brown fox ju…”

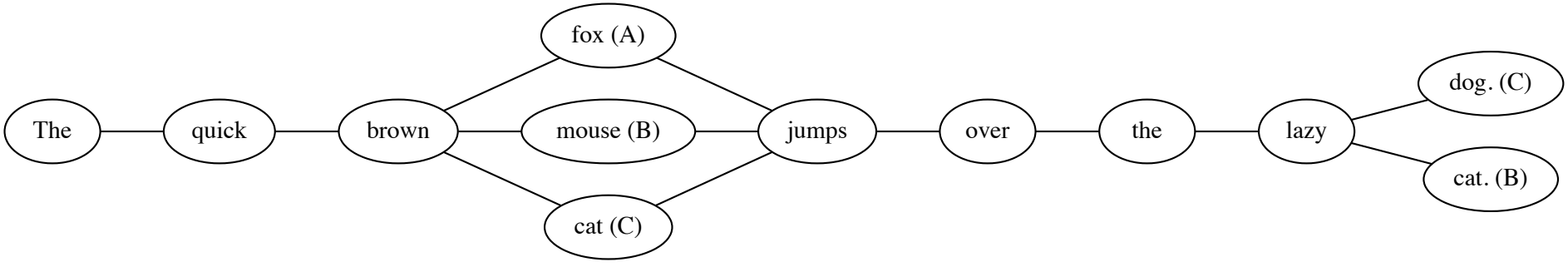
B: “The quick brown mouse jumps over the lazy cat.”

C: “The quick brown cat jumps over the lazy dog.”

We think A and B derive from the archetype via different routes, and C derives from A.

TEI app. crit. as variant graph

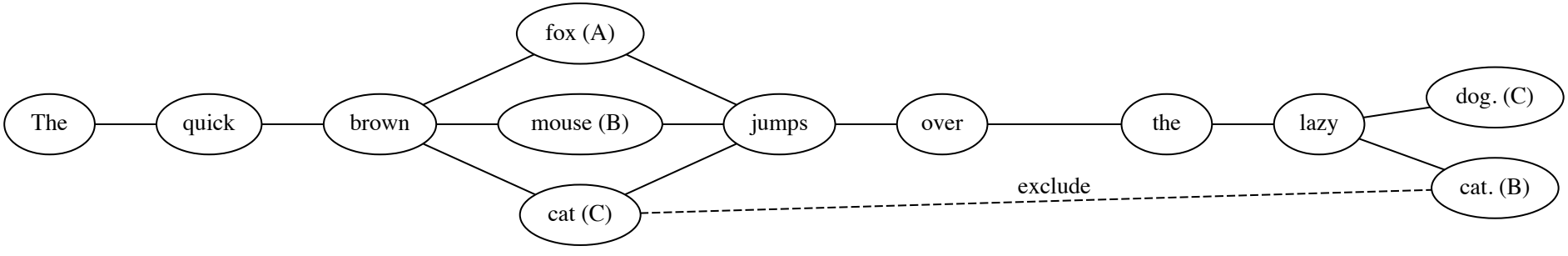
```
<p>The quick brown <app>
    <lem wit="#A">fox</lem>
    <rdg wit="#B">mouse</rdg>
    <rdg wit="#C">cat</rdg></app> jumps over
the lazy <app>
        <lem wit="#C">dog</lem>
        <rdg wit="#B">cat</rdg></app>.</p>
```



Implications

We might decide that, since the transmission of B and C was independent, you can't have two cats.

~~"The quick, brown cat jumps over the lazy cat."~~



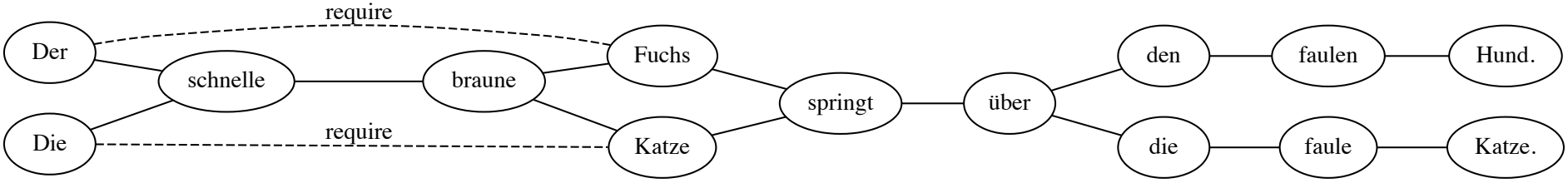

```
<p>The quick brown <app>
    <lem wit="#A">fox</lem>
    <rdg wit="#B">mouse</rdg>
    <rdg xml:id="C1" wit="#C" exclude="#C2">cat</rdg></app> jumps over the
lazy <app>
        <lem wit="#C">dog</lem>
        <rdg xml:id="C2" wit="#B" exclude="#C1">cat</rdg></app>.</p>
```

Implications

These aren’t simple, independent variations. There can be interdependencies. Imagine a German family of the tradition with two versions:

“Der schnelle braune Fuchs springt über den faulen Hund.”

“Die schnelle braune Katze springt über die faule Katze.”



If you have “Fuchs” the first word must be “Der”, if “Katze” then “Die”. “Die schnelle braune Fuchs…” would be another impossible text.

A TEI app. crit. represents a forking and rejoining of the text stream, a run of text for which there are multiple possibilities. These possibilities may be constrained by their context.

A TEI app. crit. represents a forking and rejoining of the text stream, a run of text for which there are multiple possibilities. These possibilities may be constrained by their context.

or...

A TEI app. crit. represents a forking and rejoining of the text stream, a run of text for which there are multiple possibilities. These possibilities may be constrained by their context.

or...

A TEI app. crit. entry is a type of annotation on the text, asserting that a particular source or authority has a different opinion about the text content.

TEI app. crit. as annotation

```
<p>The quick brown <app>
    <lem wit="#A">fox</lem>
    <rdg wit="#B">mouse</rdg>
    <rdg xml:id="C1" wit="#C" exclude="#C2">cat</rdg></app> jumps over the
lazy <app>
        <lem wit="#C">dog</lem>
        <rdg xml:id="C2" wit="#B" exclude="#C1">cat</rdg></app>.</p>
```

“A says, and the editor agrees, that the fourth word is ‘fox’. B says that it is ‘mouse’, and C says that it is ‘cat‘.”

Note that the apparatus doesn’t have to be inline. It could be standoff and say the same thing.

TEI app. crit. as (standoff) annotation

```
<p>The quick brown fox jumps over the lazy dog.</p>
...
<listApp>
  <app from="#match(//p[1],'fox')">
    <lem wit="#A">fox</lem>
    <rdg wit="#B">mouse</rdg>
    <rdg xml:id="C1" wit="#C" exclude="#C2">cat</rdg>
  </app>
  <app from="#match(//p[1],'dog')">
    <lem wit="#C">dog</lem>
    <rdg xml:id="C2" wit="#B" exclude="#C1">cat</rdg>
  </app>
</listApp>
```

What TEI app. crit. is *not*

- NOT a superimposition of two or more complete texts.
  - You shouldn't expect to be able to derive any individual source text from a TEI critical edition.
- Not a tool for comparing versions of a text.
- Not particularly automatable—designed to show a (human) editor's interpretation of a textual tradition.

All that said, it's a data structure, and can be repurposed. Collatex uses it as a collation export format, for example.

What it might be—a provocation

If we accept that a TEI critical apparatus can be viewed as a sort of (optionally standoff) assertive annotation, then we might imagine using it to describe things other than textual variation. What about variant markup?

Most annotation formats, including TEI <note> and things like Web Annotation, only allow you to associate the content of the annotation with the thing annotated, not to say something positive about it, like “I think this is a place name”.

I'll just leave this here...

```
<div type="textpart" subtype="chapter" n="1" xml:id="c1">
  <p type="textpart" subtype="section" n="1" xml:id="c1s1">
    <seg n="1" xml:id="c1s1p1">Gallia est omnis divisa in partes
tres, quarum unam incolunt Belgae, Aliam Aquitani, tertiam qui
ipsorum lingua Celtae, nostra Galli appellantur.</seg>...</p>
</div>...
<standoff>
  <listApp>
    <app from="#match(//seg[@xml:id='c1s1p1'],'Gallia')">
      <rdg><placeName ref="https://pleiades.stoa.org/places/993"
source="#Damon">Gallia</placeName></rdg>
    </app>
   </listApp>
</standoff>
```

"Damon says that 'Gallia' in chapter 1, paragraph 1, segment 1 is a place name referencing Pleiades #993."

This is (not) Spinal Tap:
Modeling to Prioritize Variance

"Spine 2" by Buzz Spector:

polaroid of 33 books aligned at the spines, one per human vertebra

Elisa Beshero-Bondar (@epyllia)

Spine work of a Stand-off Critical Apparatus

Spine work of a Stand-off Critical Apparatus

- express a holistic view structured according to variant locations

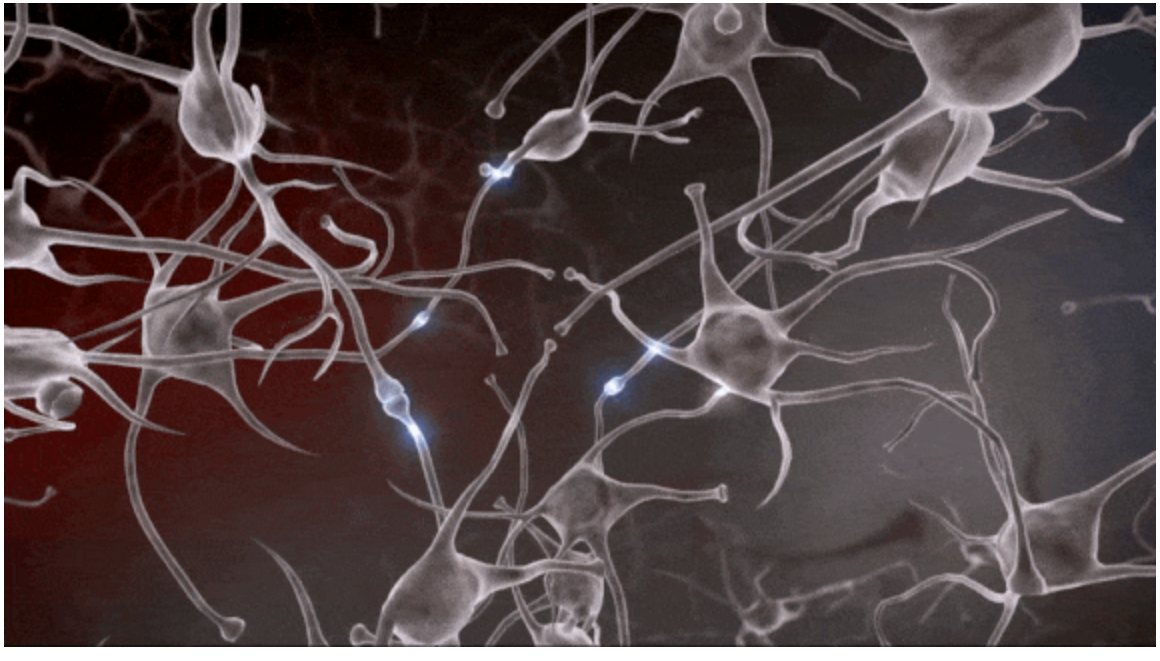
Spine work of a Stand-off Critical Apparatus

- express a holistic view structured according to variant locations
- serve as ”nerve plexus” of data pointers for dynamic coordination of multiple editions

Spine work of a Stand-off Critical Apparatus

- express a holistic view structured according to variant locations

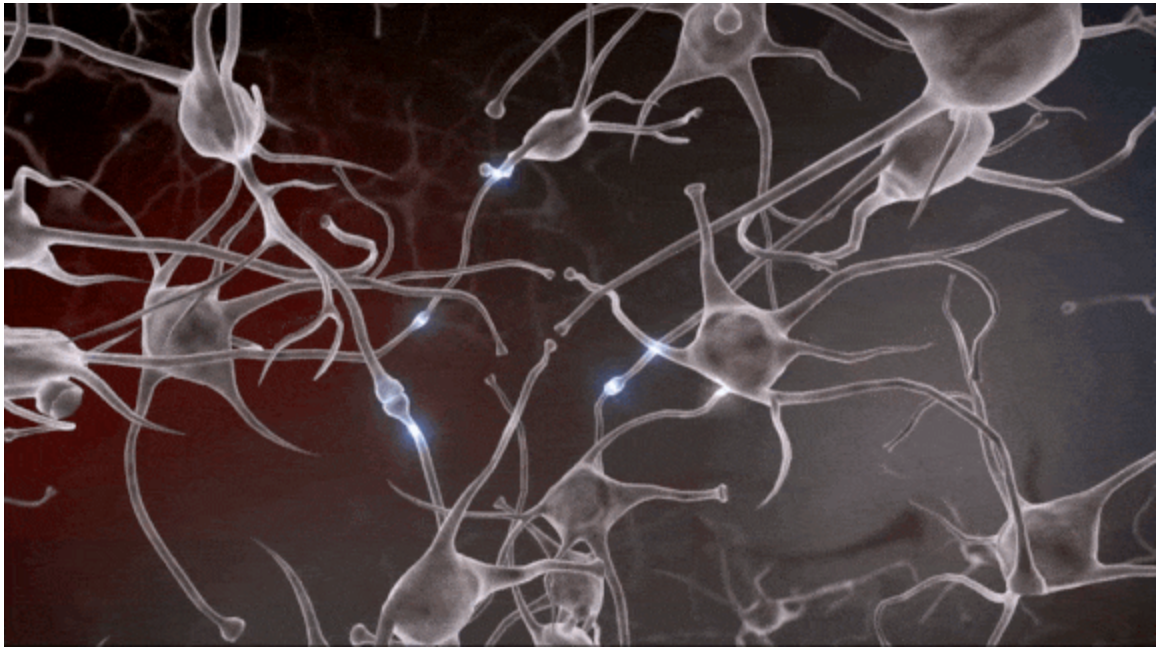
- serve as ”nerve plexus” of data pointers for dynamic coordination of multiple editions



Spine work of a Stand-off Critical Apparatus

- express a holistic view structured according to variant locations

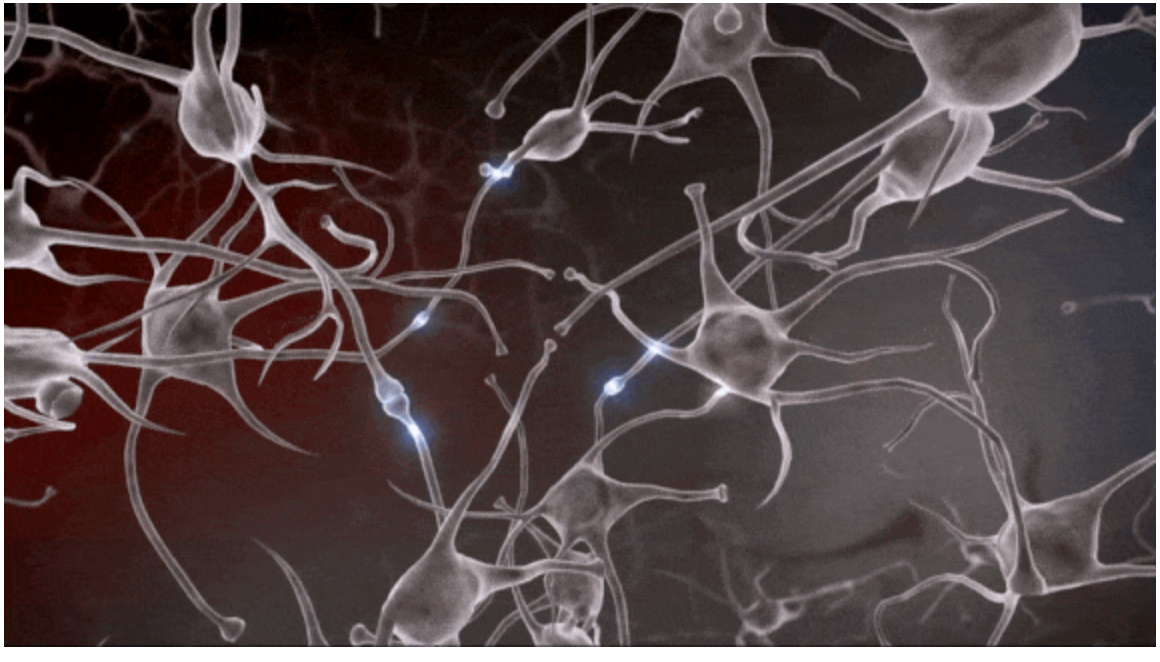
- serve as ”nerve plexus” of data pointers for dynamic coordination of multiple editions



- can be built up from computer-aided collation

Spine work of a Stand-off Critical Apparatus

- express a holistic view structured according to variant locations
- serve as ”nerve plexus” of data pointers for dynamic coordination of multiple editions



- can be built up from computer-aided collation
- case study (in the following slides) from *Frankenstein Variorum* project

Variorum - modeling change over time

Inspiration for *Frankenstein Variorum*: *Darwin Online* (ed. Barbara Bordalejo), except...

Variorum - modeling change over time

Inspiration for *Frankenstein Variorum*: *Darwin Online* (ed. Barbara Bordalejo), except...

DARWIN ONLINE

Publications | Manuscripts | Biography | Media | About us | Search | Advanced search

Online Variorum of Darwin's *Origin of Species*: first British edition (1859), page 1

1859 Page 1 or chapter None Compare 1869 Go! Introduction < Back Next >

See page in:
1859
1860
1861
1866
1869
1872

Compare with:
1860
1861
1866
1869
1872

ON THE ORIGIN OF SPECIES. — — — — —

INTRODUCTION.

WHEN on board H.M.S. 'Beagle,' as naturalist, I was much struck with certain facts in the distribution of the inhabitants of South America, and in the geological relations of the present to the past inhabitants of that continent. These facts seemed to me to throw some light on the origin of species— that mystery of mysteries, as it has been called by one of our greatest philosophers. On my return home, it occurred to me, in 1837, that something might perhaps be made out on this question by patiently accumulating and reflecting on all sorts of facts which could possibly have any bearing on it. After five years' work I allowed myself to speculate on the subject, and drew up some short notes; these I enlarged in 1844 into a sketch of the conclusions, which then seemed to me probable: from that period to the present day I have steadily pursued the same object. I hope that I may be excused for entering on these personal details, as I give them to show that I have not been hasty in coming to a decision.

My work is now nearly finished; but as it will take me two or three more years to complete it, and as my health is far from strong, I have been urged to publish this Abstract. I have more especially been induced to do this, as Mr. Wallace, who is now studying the

Comparison with 1869

ON THE ORIGIN OF SPECIES. . . . . .

INTRODUCTION.

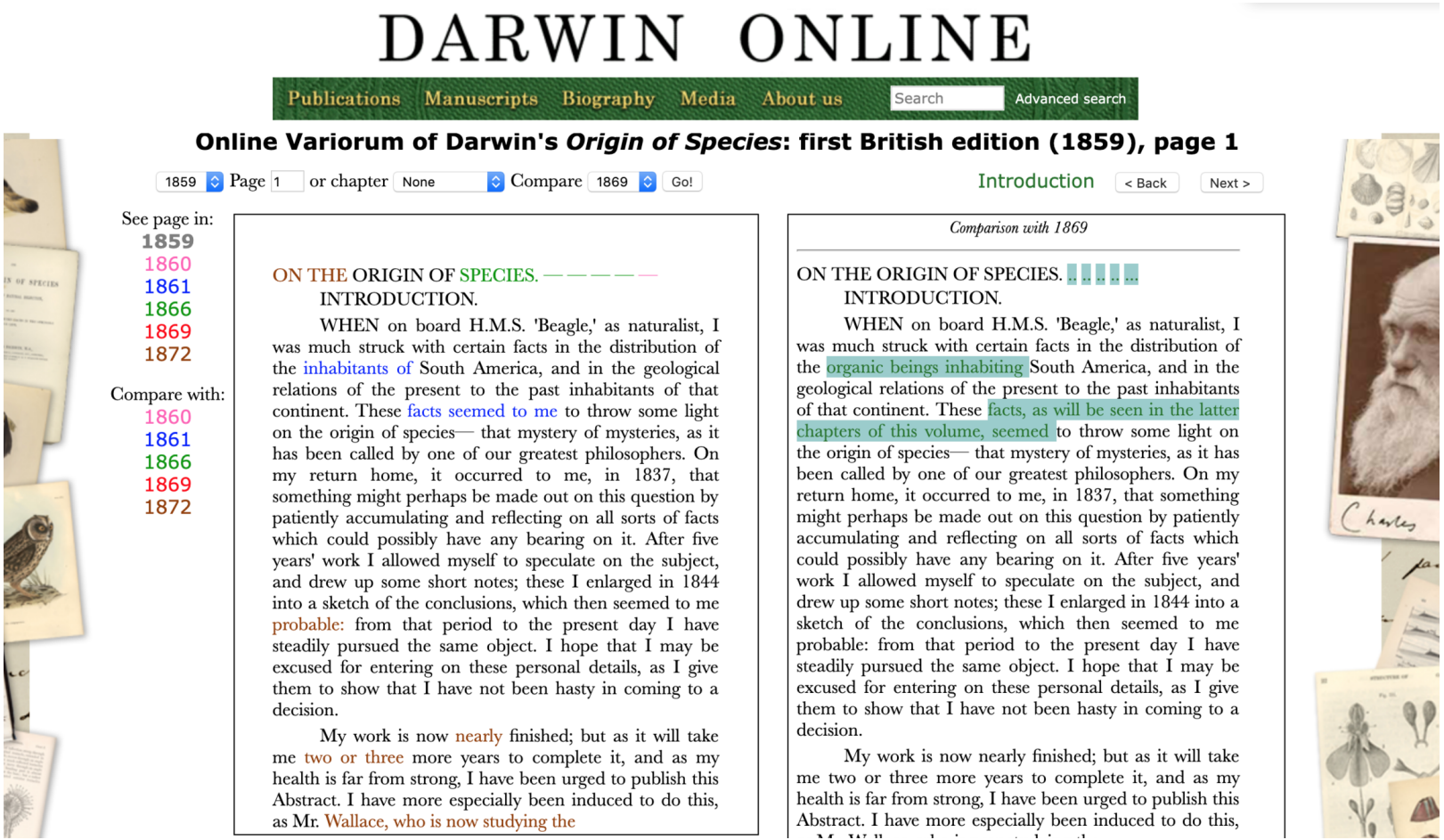
WHEN on board H.M.S. 'Beagle,' as naturalist, I was much struck with certain facts in the distribution of the organic beings inhabiting South America, and in the geological relations of the present to the past inhabitants of that continent. These facts, as will be seen in the latter chapters of this volume, seemed to throw some light on the origin of species— that mystery of mysteries, as it has been called by one of our greatest philosophers. On my return home, it occurred to me, in 1837, that something might perhaps be made out on this question by patiently accumulating and reflecting on all sorts of facts which could possibly have any bearing on it. After five years' work I allowed myself to speculate on the subject, and drew up some short notes; these I enlarged in 1844 into a sketch of the conclusions, which then seemed to me probable: from that period to the present day I have steadily pursued the same object. I hope that I may be excused for entering on these personal details, as I give them to show that I have not been hasty in coming to a decision.

My work is now nearly finished; but as it will take me two or three more years to complete it, and as my health is far from strong, I have been urged to publish this Abstract. I have more especially been induced to do this,

Variorum - modeling change over time

Inspiration for *Frankenstein Variorum*: *Darwin Online* (ed. Barbara Bordalejo), except...

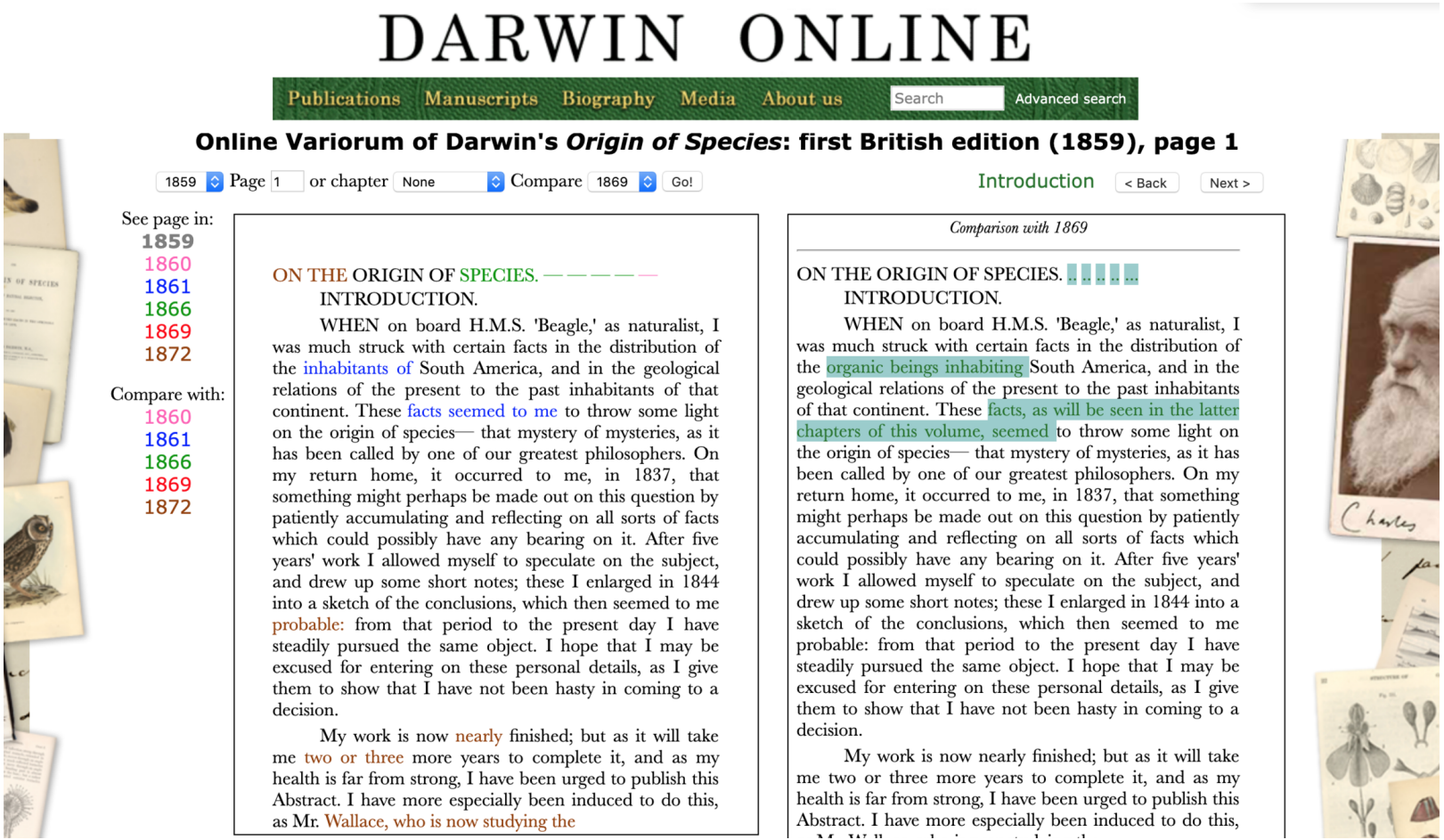
- *Frankenstein Variorum* only compares five witnesses



Variorum - modeling change over time

Inspiration for *Frankenstein Variorum*: *Darwin Online* (ed. Barbara Bordalejo), except...

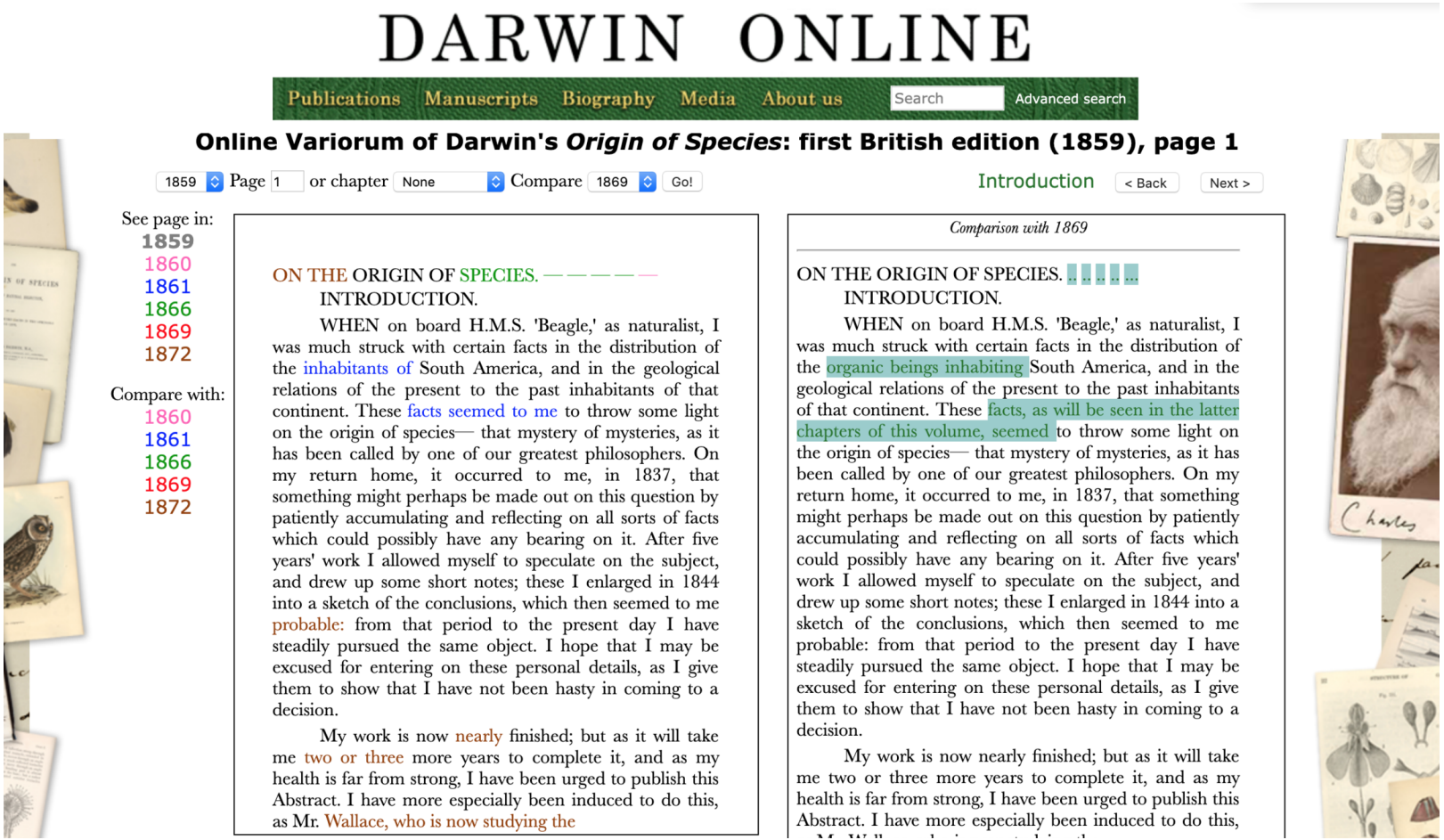
- *Frankenstein Variorum* only compares five witnesses
- *Frankenstein Variorum* incorporates two MS witnesses + three print editions



Variorum - modeling change over time

Inspiration for *Frankenstein Variorum*: *Darwin Online* (ed. Barbara Bordalejo), except...

- *Frankenstein Variorum* only compares five witnesses
- *Frankenstein Variorum* incorporates two MS witnesses + three print editions
- *Frankenstein Variorum* integrates by collation earlier digital editions made by others



Gothenburg Model

algorithm for computer-aided collation, developed in 2009 workshop of collateX and Juxta developers.

1. **Tokenization** :

2. **Normalization**

3. **Alignment**

4. **Analysis**

5. **Visualization**

Gothenburg Model

algorithm for computer-aided collation, developed in 2009 workshop of collateX and Juxta developers.

1. **Tokenization** :
   - Break down the smallest unit of comparison: (words--with punctuation, or character-by-character): FV tokenizes words and includes punctuation
2. **Normalization**

3. **Alignment**

4. **Analysis**

5. **Visualization**

Gothenburg Model

algorithm for computer-aided collation, developed in 2009 workshop of collateX and Juxta developers.

1. **Tokenization** :
   - Break down the smallest unit of comparison: (words--with punctuation, or character-by-character): FV tokenizes words and includes punctuation
2. **Normalization**
   - ('&' = 'and')
3. **Alignment**

4. **Analysis**

5. **Visualization**

Gothenburg Model

algorithm for computer-aided collation, developed in 2009 workshop of collateX and Juxta developers.

1. **Tokenization** :
   - Break down the smallest unit of comparison: (words--with punctuation, or character-by-character): FV tokenizes words and includes punctuation
2. **Normalization**
   - ('&' = 'and')
3. **Alignment**
   - Identify comparable divergence: what makes text sequences comparable units?

4. **Analysis**

5. **Visualization**

Gothenburg Model

algorithm for computer-aided collation, developed in 2009 workshop of collateX and Juxta developers.

1. **Tokenization** :
   - Break down the smallest unit of comparison: (words--with punctuation, or character-by-character): FV tokenizes words and includes punctuation
2. **Normalization**
   - ('&' = 'and')
3. **Alignment**
   - Identify comparable divergence: what makes text sequences comparable units?
   - “Chunking” text into comparable passages (chapters/paragraphs that line up with identifiable start and end points). Collation proceeds chunk by chunk.
4. **Analysis**

5. **Visualization**

Gothenburg Model

algorithm for computer-aided collation, developed in 2009 workshop of collateX and Juxta developers.

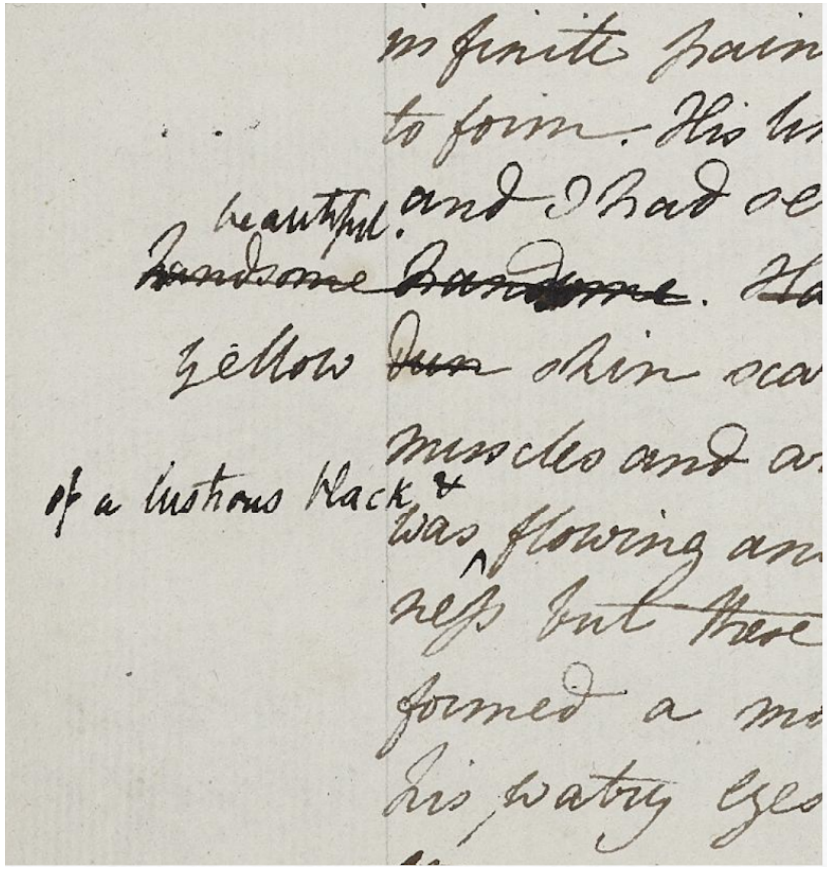
1. **Tokenization** :
   - Break down the smallest unit of comparison: (words--with punctuation, or character-by-character): FV tokenizes words and includes punctuation
2. **Normalization**
   - ('&' = 'and')
3. **Alignment**
   - Identify comparable divergence: what makes text sequences comparable units?
   - “Chunking” text into comparable passages (chapters/paragraphs that line up with identifiable start and end points). Collation proceeds chunk by chunk.
4. **Analysis**
   - (study output, correct, and re-align after machine process, AND refine automated processing)
5. **Visualization**

Gothenburg Model

algorithm for computer-aided collation, developed in 2009 workshop of collateX and Juxta developers.

1. **Tokenization** :
   - Break down the smallest unit of comparison: (words--with punctuation, or character-by-character): FV tokenizes words and includes punctuation
2. **Normalization**
   - ('&' = 'and')
3. **Alignment**
   - Identify comparable divergence: what makes text sequences comparable units?
   - “Chunking” text into comparable passages (chapters/paragraphs that line up with identifiable start and end points). Collation proceeds chunk by chunk.
4. **Analysis**
   - (study output, correct, and re-align after machine process, AND refine automated processing)
5. **Visualization**
   - critical edition apparatus, graph displays

FV: Tokenizing/normalizing S-GA diplomatic encoding



```
      <add place="sublinear"><metamark function="displace
        </mod> flowing and his teeth of a pearly white</l
    <line>ness but these luxurianc<add place="intralinear
    <line>formed a more horrid contrast with</line>
    <line>his watry eyes that seemed almost of</line>
    <line>the same colour as the dun white</line>
    <line>sockets in which they were set,</line>
  </zone>

  <zone type="left_margin" corresp="#c56-0045.01">
    <line><add><mod>
          <del rend="strikethrough">handsome</del>
          <add place="superlinear" hand="#pbs">beautiful.
        </mod></add></line>
  </zone>

  <zone type="left_margin" corresp="#c56-0045.02">
    <line><add>yellow</add></line>
  </zone>

  <zone type="left_margin" corresp="#c56-0045.03">
    <line><add hand="#pbs">of a lustrous black &amp;</add
  </zone>
</surface>
```

- required XSLT resequencing of margin zones (follow @corresp values to @xml:ids)
- required Python normalizing algorithm to suppress <line> from collation

Why collate the markup?

- Markup expresses conditions relevant for comparing texts
- Genetic markup with critical comparison:

Why collate the markup?

- Markup expresses conditions relevant for comparing texts
- Genetic markup with critical comparison:
  - genetic markup is not incomparable with markup of print editions

Why collate the markup?

- Markup expresses conditions relevant for comparing texts

- Genetic markup with critical comparison:
  - genetic markup is not incomparable with markup of print editions
  - genetic markup can answer scholarly research questions at critical scale
    - *MWS reworking the text: How guilty does Victor Frankenstein appear in 1816, 1818, 1820s after Percy's death, 1831?*
    - *Which passages underwent the most intense, ”molten” transformations over time?*
    - *What kind of influence did Percy Shelley have on Frankenstein‘s print editions?*

Preparing marked-up texts for collation

Preparing marked-up texts for collation

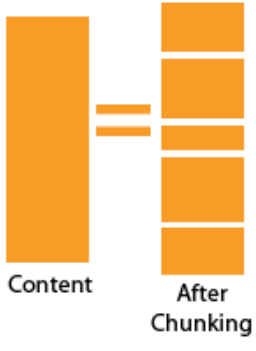
- **Determine comparable markup of text structures across Variorum editions:**
  - volume (print editions only), letter, chapter
  - paragraph, poetry line-groups and lines
  - notes

Preparing marked-up texts for collation

- **Determine comparable markup of text structures across Variorum editions:**
  - volume (print editions only), letter, chapter
  - paragraph, poetry line-groups and lines
  - notes
- **Markup of manuscript events included in Variorum comparison:**
  - deletion, insertion, gap

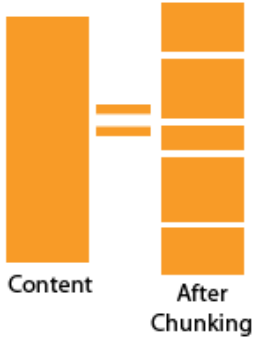
Preparing marked-up texts for collation

- **Determine comparable markup of text structures across Variorum editions:**
  - volume (print editions only), letter, chapter
  - paragraph, poetry line-groups and lines
  - notes
- **Markup of manuscript events included in Variorum comparison:**
  - deletion, insertion, gap
- **”Chunking” algorithm:** (limit possibility of major misalignments)
  - Locate ”seams” where all editions align
  - Divide into ”chunks” at the seams
  - Prep each edition as 33 collation ”chunks”, C01 - C33
  - All files identified as the same chunk are collated together



Preparing marked-up texts for collation

- **Determine comparable markup of text structures across Variorum editions:**
  - volume (print editions only), letter, chapter
  - paragraph, poetry line-groups and lines
  - notes
- **Markup of manuscript events included in Variorum comparison:**
  - deletion, insertion, gap
- **”Chunking” algorithm:** (limit possibility of major misalignments)
  - Locate ”seams” where all editions align
  - Divide into ”chunks” at the seams
  - Prep each edition as 33 collation ”chunks”, C01 - C33
  - All files identified as the same chunk are collated together



- **Normalizing algorithm:**
  - Decide what marks are equivalent
  - ignore but preserve other markup in collation process, also abbreviations, capitalization.

TEI App-Crit on its way to becoming a Spine

TEI App-Crit on its way to becoming a Spine

- output of computer-aided collation (not TEI, but like it)

TEI App-Crit on its way to becoming a Spine

- output of computer-aided collation (not TEI, but like it)
- build up variorum edition expressed in app-crit with flattened tags

Collating with markup: handsome” / “beautiful” passage processed by collateX

Collating with markup: handsome” / “beautiful” passage processed by collateX

```
 <app xml:id="C10_app44">
          <rdgGrp xml:id="C10_app44_rg1"
n="['&lt;del&gt;handsome&lt;del&gt;
&lt;del&gt;handsome&lt;
del&gt;beautiful.&lt;del&gt;handsome&lt;del&gt;beautiful;', 'great']"

            <rdg wit="fMS">&lt;lb n="c56-0045__main__23"/&gt;
  &lt;del rend="strikethrough" sID="c56-0045__main__d2e9837"/&gt;
handsome&lt;del eID="c56-0045__main__d2e9837"/&gt;
&lt;mdel&gt;.
&lt;/mdel&gt;&lt;lb n="c56-0045__left_margin__1"/&gt;
&lt;del rend="strikethrough" sID="c56-0045__left_margin__d2e9853"/&gt;handsome&lt;
del eID="c56-0045__left_margin__d2e9853"/&gt;beautiful.
&lt;del rend="strikethrough" sID="c56-0045__main__d2e9865"/&gt;
Handsome&lt;del eID="c56-0045__main__d2e9865"/&gt;
Beautiful; Great </rdg>
       </rdgGrp>

        <rdgGrp xml:id="C10_app44_rg2" n="['beautiful.', 'beautiful!—great']">
              <rdg wit="f1818">beautiful. Beautiful!—Great </rdg>
              <rdg wit="f1823">beautiful. Beautiful!—Great </rdg>
              <rdg wit="fThomas">beautiful. Beautiful!—Great </rdg>
              <rdg wit="f1831">beautiful. Beautiful!—Great </rdg>
        </rdgGrp>
</app>
```

Collating with markup: handsome” / “beautiful” passage processed by collateX

```
 <app xml:id="C10_app44">
          <rdgGrp xml:id="C10_app44_rg1"
n="['&lt;del&gt;handsome&lt;del&gt;
&lt;del&gt;handsome&lt;
del&gt;beautiful.&lt;del&gt;handsome&lt;del&gt;beautiful;', 'great']"

            <rdg wit="fMS">&lt;lb n="c56-0045__main__23"/&gt;
  &lt;del rend="strikethrough" sID="c56-0045__main__d2e9837"/&gt;
handsome&lt;del eID="c56-0045__main__d2e9837"/&gt;
&lt;mdel&gt;.
&lt;/mdel&gt;&lt;lb n="c56-0045__left_margin__1"/&gt;
&lt;del rend="strikethrough" sID="c56-0045__left_margin__d2e9853"/&gt;handsome&lt;
del eID="c56-0045__left_margin__d2e9853"/&gt;beautiful.
&lt;del rend="strikethrough" sID="c56-0045__main__d2e9865"/&gt;
Handsome&lt;del eID="c56-0045__main__d2e9865"/&gt;
Beautiful; Great </rdg>
       </rdgGrp>

        <rdgGrp xml:id="C10_app44_rg2" n="['beautiful.', 'beautiful!—great']">
              <rdg wit="f1818">beautiful. Beautiful!—Great </rdg>
              <rdg wit="f1823">beautiful. Beautiful!—Great </rdg>
              <rdg wit="fThomas">beautiful. Beautiful!—Great </rdg>
              <rdg wit="f1831">beautiful. Beautiful!—Great </rdg>
        </rdgGrp>
</app>
```

an ugly but powerful Frankenstein creature of collation!

TEI Interchangeability :: Collation of Markup

TEI advantage: Interchange (cf. Syd Bauman, “Interchange vs. Interoperability”):

”Human A” reading code written and documented by ”Human B” can understand how to adapt that code without consulting Human B.

Doing the work of interchange:

```
<milestone unit="tei:p"/>      ::      <p>. . . . . </p>
```

TEI Interchangeability :: Collation of Markup

TEI advantage: Interchange (cf. Syd Bauman, "Interchange vs. Interoperability"):

"Human A" reading code written and documented by "Human B" can understand how to adapt that code without consulting Human B.

Doing the work of interchange:

- **Determine** how to follow the "running stream" of semantically readable text to be compared with other editions.

```
<milestone unit="tei:p"/>      ::      <p>. . . . . </p>
```

TEI Interchangeability :: Collation of Markup

TEI advantage: Interchange (cf. Syd Bauman, "Interchange vs. Interoperability"):

"Human A" reading code written and documented by "Human B" can understand how to adapt that code without consulting Human B.

Doing the work of interchange:

- **Determine** how to follow the "running stream" of semantically readable text to be compared with other editions.
- **Map** the semantically comparable units in collation algorithm

```
<milestone unit="tei:p"/>      ::      <p>. . . . . </p>
```

TEI Interchangeability :: Collation of Markup

TEI advantage: Interchange (cf. Syd Bauman, "Interchange vs. Interoperability"):

"Human A" reading code written and documented by "Human B" can understand how to adapt that code without consulting Human B.

Doing the work of interchange:

- **Determine** how to follow the "running stream" of semantically readable text to be compared with other editions.
- **Map** the semantically comparable units in collation algorithm

```
<milestone unit="tei:p"/>      ::      <p>. . . . . </p>
```

- **Mask** the markup that isn't semantically comparable (MS surfaces, zones, lines)

TEI Interchangeability :: Collation of Markup

TEI advantage: Interchange (cf. Syd Bauman, "Interchange vs. Interoperability"):

"Human A" reading code written and documented by "Human B" can understand how to adapt that code without consulting Human B.

Doing the work of interchange:

- **Determine** how to follow the "running stream" of semantically readable text to be compared with other editions.
- **Map** the semantically comparable units in collation algorithm

```
<milestone unit="tei:p"/>      ::      <p>. . . . . </p>
```

- **Mask** the markup that isn't semantically comparable (MS surfaces, zones, lines)
- **Decide** on how to handle <add> and <del> markup:

TEI Interchangeability :: Collation of Markup

TEI advantage: Interchange (cf. Syd Bauman, "Interchange vs. Interoperability"): "Human A" reading code written and documented by "Human B" can understand how to adapt that code without consulting Human B.

Doing the work of interchange:

- **Determine** how to follow the "running stream" of semantically readable text to be compared with other editions.
- **Map** the semantically comparable units in collation algorithm

```
<milestone unit="tei:p"/>      ::      <p>. . . . . </p>
```

- **Mask** the markup that isn't semantically comparable (MS surfaces, zones, lines)
- **Decide** on how to handle <add> and <del> markup:
    - Do you want your critical apparatus to include deleted material?

TEI Interchangeability :: Collation of Markup

TEI advantage: Interchange (cf. Syd Bauman, "Interchange vs. Interoperability"): "Human A" reading code written and documented by "Human B" can understand how to adapt that code without consulting Human B.

Doing the work of interchange:

- **Determine** how to follow the "running stream" of semantically readable text to be compared with other editions.
- **Map** the semantically comparable units in collation algorithm

```
<milestone unit="tei:p"/>      ::      <p>. . . . . </p>
```

- **Mask** the markup that isn't semantically comparable (MS surfaces, zones, lines)
- **Decide** on how to handle <add> and <del> markup:
    - Do you want your critical apparatus to include deleted material?
    - Or only the "finished" MS? (Mask the <del> elements, and preserve the <add> material)

XPointer Challenge: find the locations expressed in each app in the original editions

- Method 1: produce edition files from the app-crit with XSLT
  - Plant TEI element (e.g. <seg>) to indicate variant locations, give each an @xml:id
  - Build Spine by generating @target directly accessing <seg> elements
- Method 2: point to pre-existing editions
  - Programmatic search-work to find variant passages (not signalled in the edition markup)
  - Build Spine with XPath and string-range indicators
    - See TEI Guidelines 16.2.4.1

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>
```

Markup is text, after all!

Summary of Spine-Making:

Markup is text, after all!

Summary of Spine-Making:

- Flatten markup for computer assisted collation

Markup is text, after all!

Summary of Spine-Making:

- Flatten markup for computer assisted collation
- Edit the output collation (Gothenberg Model process)

Markup is text, after all!

Summary of Spine-Making:

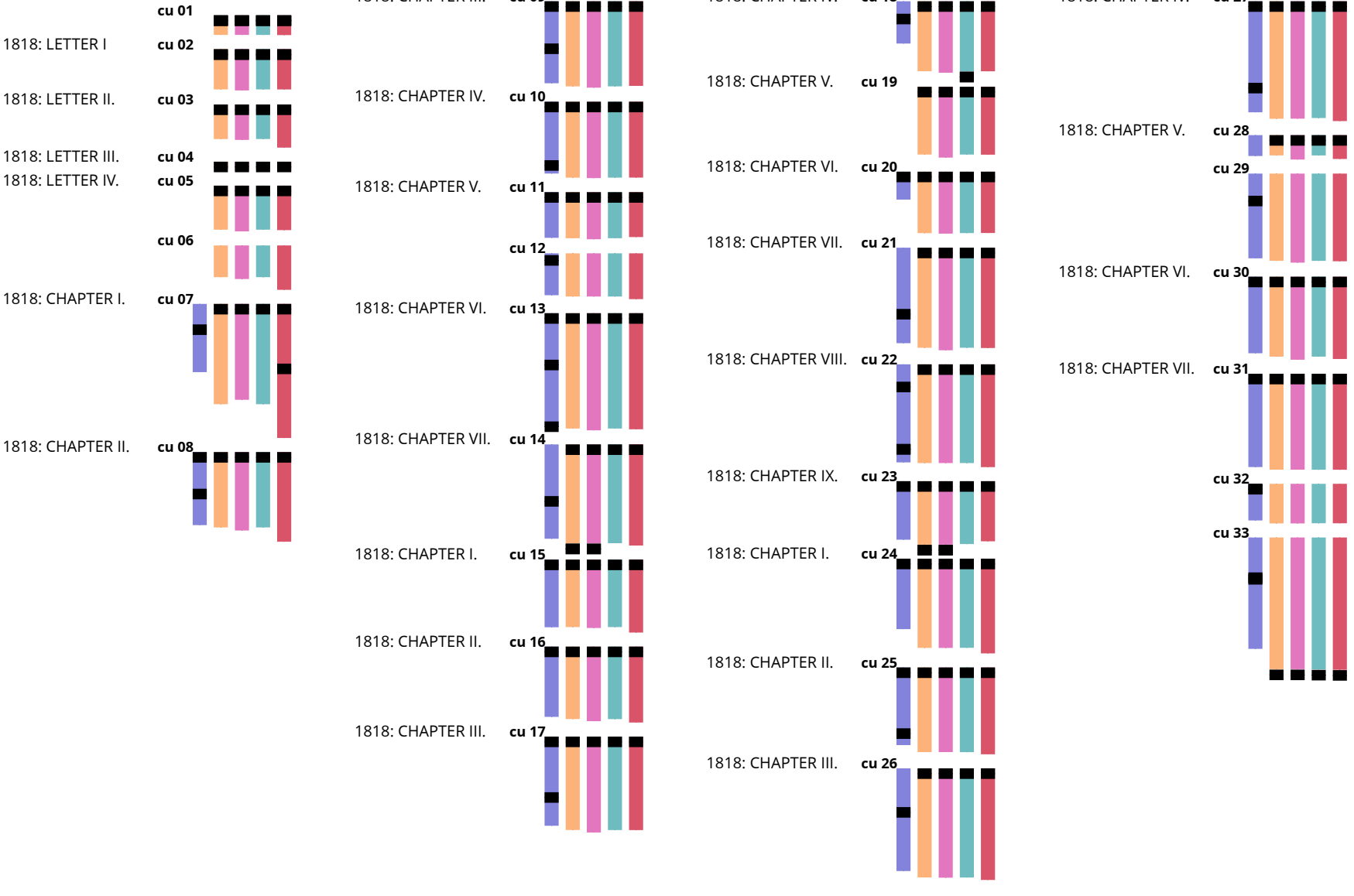
- Flatten markup for computer assisted collation
- Edit the output collation (Gothenberg Model process)
- XSLT Transformation A (pipeline): raise editions with “hotspots”
  - Raise the flattened markup to reconstruct some editions, with marked <seg> elements
  - Deal with overlapping hierarchies: (e.g. Molten passages cross paragraph boundaries): Output editions break into fragments around up-raised markup.

Markup is text, after all!

Summary of Spine-Making:

- Flatten markup for computer assisted collation
- Edit the output collation (Gothenberg Model process)
- XSLT Transformation A (pipeline): raise editions with “hotspots”
  - Raise the flattened markup to reconstruct some editions, with marked <seg> elements
  - Deal with overlapping hierarchies: (e.g. Molten passages cross paragraph boundaries): Output editions break into fragments around up-raised markup.
- XSLT Transformation B: construct the standoff spine with pointers:
  - Convert collateX output critical apparatus to ”spine nerve plexus” holding XML pointers
  - These **point to the marked hotspots** in the editions reconstructed in Pipeline A
  - And **point to xml:ids + string-ranges in external editions** that were not generated by the process (e.g. *FV* pointing to *Shelley-Godwin Archive*)

- **“Spine” data model** = standoff use of TEI critical apparatus:
  - can include processed data, like maximum edit-distance, at each location
  - can include data on normalization: e.g. normalized tokens used in collation process
  - coordinates data on variance,
  - points to specific locations in separate edition files


cu 01
1818: LETTER I
cu 02
1818: LETTER II.
cu 03
1818: LETTER III.
cu 04
1818: LETTER IV.
cu 05
cu 06
1818: CHAPTER I.
cu 07
1818: CHAPTER II.
cu 08
1818: CHAPTER III.
cu 09
1818: CHAPTER IV.
cu 10
1818: CHAPTER V.
cu 11
cu 12
1818: CHAPTER VI.
cu 13
1818: CHAPTER VII.
cu 14
1818: CHAPTER I.
cu 15
1818: CHAPTER II.
cu 16
1818: CHAPTER III.
cu 17
1818: CHAPTER IV.
cu 18
1818: CHAPTER V.
cu 19
1818: CHAPTER VI.
cu 20
1818: CHAPTER VII.
cu 21
1818: CHAPTER VIII.
cu 22
1818: CHAPTER IX.
cu 23
1818: CHAPTER I.
cu 24
1818: CHAPTER II.
cu 25
1818: CHAPTER III.
cu 26
1818: CHAPTER IV.
cu 27
1818: CHAPTER V.
cu 28
cu 29
1818: CHAPTER VI.
cu 30
1818: CHAPTER VII.
cu 31
cu 32
cu 33

”Preparing diversely encoded documents for collation challenges us to consider inconsistent and overlapping hierarchies as a tractable matter for computational alignment—where alignment becomes an organizing principle that fractures hierarchies, chunking if not atomizing them at the level of the smallest meaningfully sharable semantic features.”

”We have negotiated interchangeability by cutting across individual text hierarchies to emphasize lateral connections and commonalities—making a new TEI whose hierarchy serves as a stand-off ”spine” or ”switchboard” permitting comparison and sharing of common data. Our goal of pointing to aligned data required us to locate the interchangeable structural markers in our source documents.”

For more on our document data modeling, see

Beshero-Bondar, Elisa E., and Raffaele Viglianti. “Stand-off Bridges in the Frankenstein Variorum Project: Interchange and Interoperability within TEI Markup Ecosystems.” *Balisage Series on Markup Technologies*, vol. 21 (2018). https://doi.org/10.4242/BalisageVol21.Beshero-Bondar01.

0

Publishing a Stand-off Critical Apparatus: Leveraging isomorphic representations across text and music notation

Raff Viglianti (@raffazizzi)

Early Modern Songscapes

English ayres & their dynamic acoustic environments

songscapes.org

Stand-off apparatus and the representation of primary sources

BL Add. MS 53723 `<l>alas forsaken I Complaine;</l>`

C 709 `<l>Alas deserted I Complain,</l>`

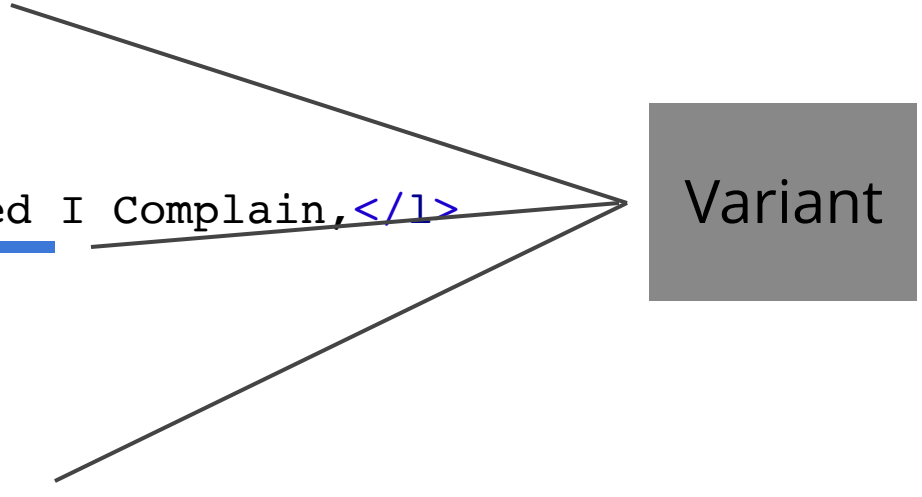
Folger L638 `<l>Alas deserted I complain;</l>`

Stand-off apparatus and the representation of primary sources

BL Add. MS 53723 `<l>alas forsaken I Complaine;</l>`

C 709 `<l>Alas deserted I Complain,</l>`

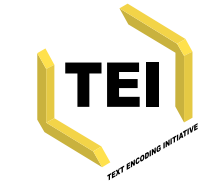
Folger L638 `<l>Alas deserted I complain;</l>`



Songscapes stand-off collation

TEI (no XPointer in this case)

BL Add. MS 53723



+


```
1  <TEI>
2    <div>
3      <head>Text Collation</head>
4      <app>
5        <rdgGrp>
6          <rdg wit="#BL_53723">
7            <ptr target="tei/Ariadne-BL_53723.xml#v1"/>
8          </rdg>
9          <rdg wit="#L638">
10           <ptr target="tei/Ariadne-L638.xml#v1"/>
11         </rdg>
12       </rdgGrp>
13       <rdg wit="#C709">
14         <ptr target="tei/Ariadne-C709.xml#v1"/>
15       </rdg>
16     </app>
17   </div>
18 </TEI>
```

Folger L638

+

C 709

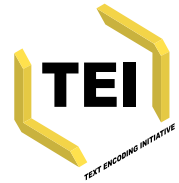
Adapted from:
https://github.com/EarlyModernSongscapes/songscapes/blob/master/data/collations/Theseus%2C_O_Theseus%2C_hark!.xml

Songscapes stand-off collation

MEI

BL Add. MS 53723

Folger L638



+

+

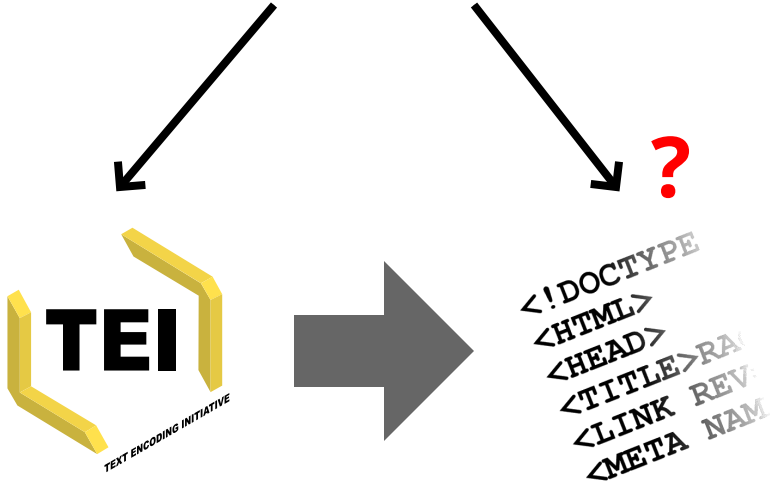

```
<TEI>
  <div>
    <head>Music Collation</head>
    <notatedMusic>
      <mei:mei> <!-- header -->
        <mei:music><mei:body><mei:mdiv><mei:score>
          <mei:app>
            <mei:rdg source="#M-BL_53723"
                 target="mei/Ariadne-BL_53723.xml#m-101
                         mei/Ariadne-BL_53723.xml#m-106"/>
            <mei:rdg source="#M-L638"
                target="mei/Ariadne-L638.xml#m-101
                        mei/Ariadne-L638.xml#m-106"/>

          </mei:app>
        </mei:score></mei:mdiv></mei:body></mei:music>
    </mei:mei>
  </div>
</TEI>
```

Adapted from:
https://github.com/EarlyModernSongscapes/songscapes/blob/master/data/collations/Theseus%2C_O_Theseus%2C_hark!.xml

Publishing this kind of model (including Frankenstein Variorum!)

- Typical TEI to HTML transformation would require transforming pointers too.
- Pointers need to be followed in response to user interaction.

```
<ptr target="MSC56.xml#string-range(//line[13],0,21)" />
```



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>
```

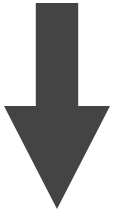
Thomas edition

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

Isomorphic representations (TEI)

CETEIcean 🐋 (/sɪˈtiːʃn/)
https://github.com/TEIC/CETEIcean

```
<lg type="stanza">
  <l>Theseous! ô theseus! heark! but yet in vaine,</l>
  <l>alas <seg xml:id="v4">forsaken</seg> I Complaine;</l>
  <l>it was some Neighb'ringe Rock / more softe then he, /</l>
  <l rend="indent1">whose hollow Bowels pittyed me,</l>
  <!-- ... -->
</lg>
```



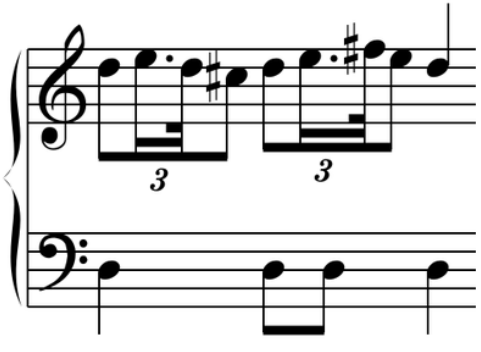
HTML5 Custom Elements

```
<tei-lg type="stanza">
  <tei-l>Theseous! ô theseus! heark! but yet in vaine,</tei-l>
  <tei-l>alas <tei-seg xml:id="v4">forsaken</tei-seg> I Complaine;</tei-l>
  <tei-l>it was some Neighb'ringe Rock / more softe then he, /</tei-l>
  <tei-l rend="indent1">whose hollow Bowels pittyed me,</tei-l>
  <!-- ... -->
</tei-lg>
```

Isomorphic representations (MEI)

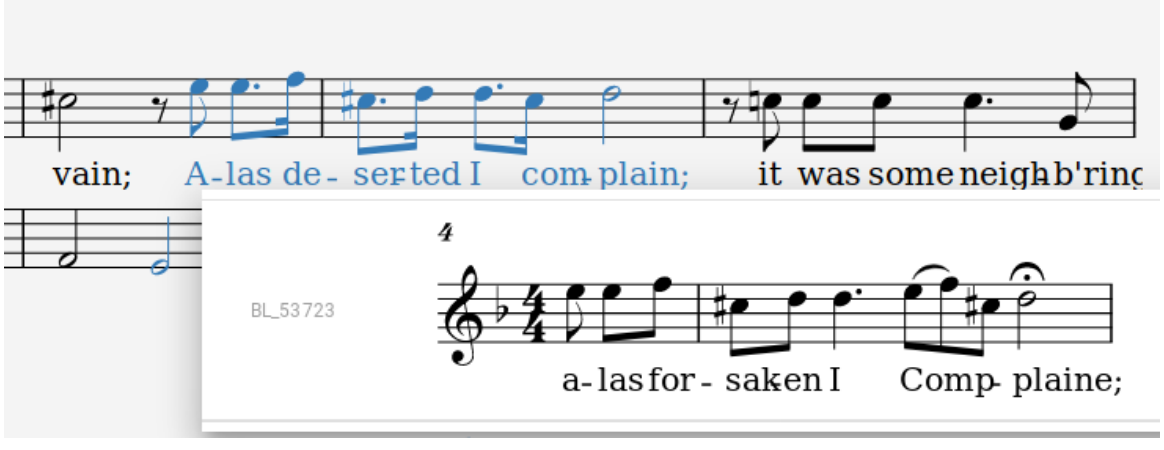
- Verovio: SVG as isomorphic surrogate of MEI

```
<tuplet xml:id="t1" num="3" numbase="2">
  <beam xml:id="b1">
    <note xml:id="n1" pname="d" oct="5" dur="8" />
    <note xml:id="n2" pname="e" oct="5" dur="16" dots="1"/>
    <note xml:id="n3" pname="d" oct="5" dur="32" />
    <note xml:id="n4" pname="c" oct="5" dur="8" accid="s"/>
  </beam>
</tuplet>
<beam xml:id="b2">
  <tuplet xml:id="t2" num="3" numbase="2">
    <note xml:id="n5" pname="d" oct="5" dur="8" />
    <note xml:id="n6" pname="e" oct="5" dur="16" dots="1"/>
    <note xml:id="n7" pname="f" oct="5" dur="32" accid="s"/>
    <note xml:id="n8" pname="e" oct="5" dur="8"/>
  </tuplet>
</beam>
```

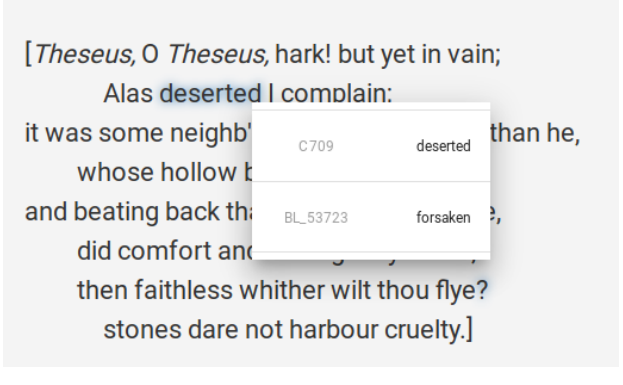



```
<g class="tuplet" id="svg-t1" >
  <g class="beam" id="svg-b1" >
    <g class="note" id="svg-n1" >...</g>
    <g class="note" id="svg-n2" >...</g>
    <g class="note" id="svg-n3" >...</g>
    <g class="note" id="svg-n4" >...</g>
  </g>
</g>
<g class="beam" id="svg-b2" >
  <g class="tuplet" id="svg-t2" >
    <g class="note" id="svg-n5" >...</g>
    <g class="note" id="svg-n6" >...</g>
    <g class="note" id="svg-n7" >...</g>
    <g class="note" id="svg-n8" >...</g>
  </g>
</g>
```

Songscapes viewer



music

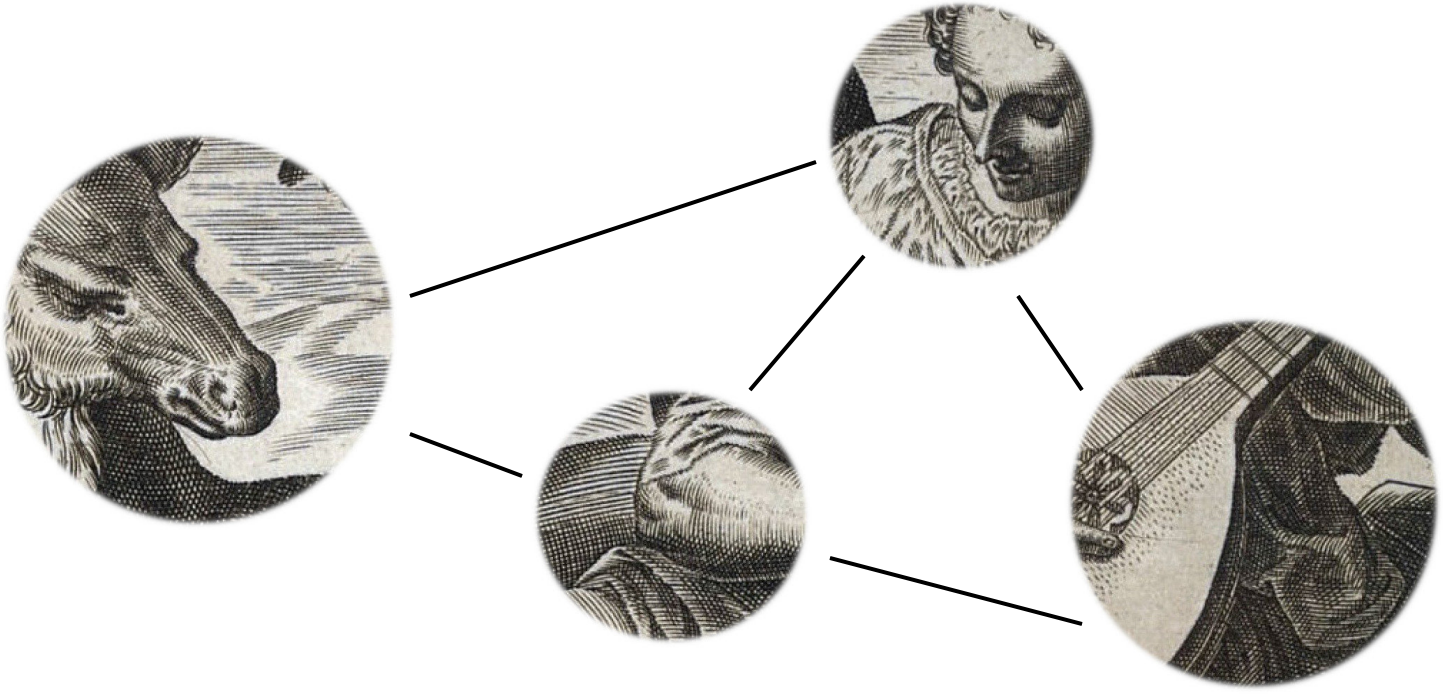


text

From: ems.digitalscholarship.utsc.utoronto.ca/islandora/object/ems%3A102

Addressability beyond a single project

- What if a stand-off collation pointed to TEI / MEI resources from other projects?
  - breaking silos (further)
  - building on existing resources / editions
- We need well thought out and flexible stand-off support in TEI



Data models, many-witness texts, and the future of apparatus markup: **a response**

James Cummings (@jamescummings)

What *really* is a critical apparatus

- Hugh Cayless started us out with an excellent (re-)introduction to critical apparatus and ways to view it specifically:
  - TEI critical apparatus as variant graph
  - TEI critical apparatus as annotation
- He suggested some things TEI critical apparatus is not
- And tried to provoke us with TEI critical apparatus as standoff assertive annotation providing variant markup

A TEI app. crit. represents a forking and rejoining of the text stream, a run of text for which there are multiple possibilities. These possibilities may be constrained by their context.

or...

A TEI app. crit. entry is a type of annotation on the text, asserting that a particular source or authority has a different opinion about the text content.

A TEI app. crit. represents a forking and rejoining of the text stream, a run of text for which there are multiple possibilities. These possibilities may be constrained by their context.

or...

- **'And'? Are these mutually exclusive viewpoints or can we use both in the same document?**

A TEI app. crit. entry is a type of annotation on the text, asserting that a particular source or authority has a different opinion about the text content.

TEI app. crit. as (standoff) annotation

```
<p>The quick brown fox jumps over the lazy dog.</p>
...
<listApp>
  <app from="#match(//p[1],'fox')">
    <lem wit="#A">fox</lem>
    <rdg wit="#B">mouse</rdg>
    <rdg xml:id="C1" wit="#C" exclude="#C2">cat</rdg>
  </app>
  <app from="#match(//p[1],'dog')">
    <lem wit="#C">dog</lem>
    <rdg xml:id="C2" wit="#B" exclude="#C1">cat</rdg>
  </app>
</listApp>
```

TEI app. crit. as (standoff) annotation

```
<p>The quick brown fox jumps over the lazy dog.</p>
...
<listApp>
  <app from="#match(//p[1],'fox')">
    <lem wit="#A">fox</
    <rdg wit="#B">mouse
    <rdg xml:id="C1" wi
  </app>
  <app from="#match(//p
    <lem wit="#C">dog</
    <rdg xml:id="C2" wi
  </app>
</listApp>
```

- **Is a <lemrom> necessary or is that determined by source? (if we own it?) But I suppose it provides metadata of witness?**
- **Standoff apparatus seems _much_ easier if word-level markup exists. (e.g. from="#w4") Should we be encouraging this?**

What TEI app. crit. is *not*

- NOT a superimposition of two or more complete texts.
  - You shouldn‘t expect to be able to derive any individual source text from a TEI critical edition.
- Not a tool for comparing versions of a text.
- Not particularly automatable—designed to show a (human) editor‘s interpretation of a textual tradition.

All that said, it’s a data structure, and can be repurposed. Collatex uses it as a collation export format, for example.

What TEI app. crit. is *not*

- NOT a superimposition of two or more complete texts.
  - You shouldn‘t expect to be able to derive any individual source text from a TEI critical edition.
- Not a tool for comparing versions of a text.
- Not particularly automatable—designed to show a (human) editor‘s interpretation of a textual tradition.

- **The key word is 'expect'... plenty of projects do precisely this with their markup because it has been created with this in mind. And software (c.f. Versioning Machine) works this way.**
- **How do we document that this is a possibility in our metadata?**
- **It will *always* be editor's version of witness**

He'll just leave this here...

```
<div type="textpart" subtype="chapter" n="1" xml:id="c1">
  <p type="textpart" subtype="section" n="1" xml:id="c1s1">
    <seg n="1" xml:id="c1s1p1">Gallia est omnis divisa in partes
tres, quarum unam incolunt Belgae, Aliam Aquitani, tertiam qui
ipsorum lingua Celtae, nostra Galli appellantur.</seg>...</p>
</div>...
<standoff>
  <listApp>
    <app from="#match(//seg[@xml:id='c1s1p1'],'Gallia')">
      <rdg><placeName ref="https://pleiades.stoa.org/places/993"
source="#Damon">Gallia</placeName></rdg>
    </app>
   </listApp>
</standoff>
```

"Damon says that 'Gallia' in chapter 1, paragraph 1, segment 1 is a place name referencing Pleiades #993."

I'll just change this here...

```
<div type="textpart" subtype="chapter" n="1" xml:id="c1">
  <p type="textpart" subtype="section" n="1" xml:id="c1s1">
    <seg n="1" xml:id="c1s1p1">Gallia est omnis divisa in partes tres,
quarum unam incolunt Belgae.</seg>...</p></div>...
<standoff>
  <listApp>
    <app from="#match(//seg[@xml:id='c1s1p1'],'Gallia')">
      <rdg>
        <div>
         <head>Does standoff have to result in valid TEI?
         Should this only be used for assertive annotation?</head>
            <!-- Lots of random stuff here -->
        </div>
      </rdg>
    </app>
  </listApp>
</standoff>
```

Reminder: and <floatingText> now allowed inside <rdg>... for better or worse

This is (not) Spinal Tap

- Elisa Beshero-Bondar described the impressive 'nerve plexus' spine as central coordinating structure in Frankenstein Variorum
- How much can be derived? Are general systems for collation spine construction possible?
- Pointer based systems like this highlight the lack of good support for working in stand-off / out-of-line methods in most XML editors
- Worry about fragility of string-ranges, while reasonable in closed ecosystem, how much should we worry about this with networked distributed systems not under our control?

Gothenburg Model

algorithm for computer-aided collation, developed in 2009 workshop of collateX and Juxta developers.

1. **Tokenization** :

2. **Normalization**

3. **Alignment**

4. **Analysis**

5. **Visualization**

Gothenburg Model

algorithm for computer-aided collation, developed in 2009 workshop of collateX and Juxta developers.

1. **Tokenization** :
   - Break down the smallest unit of comparison: (words--with punctuation, or character-by-character): FV tokenizes words and includes punctuation
2. **Normalization**

3. **Alignment**
   - **I like systems based on tokenized words (characters seems too overkill for me)**
   - **'includes punctuation' -- in the word or as <pc>?**
4. - **Always worried about normalization steps... what is lost? (Assuming nothing here, as is merely for collation?)**

5. **Visualization**

Gothenburg Model

algorithm for computer-aided collation, developed in 2009 workshop of collateX and Juxta developers.

1. **Tokenization** :
   - Break down the smallest unit of comparison: (words--with punctuation, or character-by-character): FV tokenizes words and includes punctuation
2. **Normalization**
   - ('&' = 'and')
3. **Alignment**
   - **I like systems based on tokenized words (characters seems too overkill for me)**
   - **'includes punctuation' -- in the word or as <pc>?**
4. - **Always worried about normalization steps... what is lost? (Assuming nothing here, as is merely for collation?)**
5. **Visualization**

Gothenburg Model

algorithm for computer-aided collation, developed in 2009 workshop of collateX and Juxta developers.

1. **Tokenization** :
   - Break down the smallest unit of comparison: (words--with punctuation, or character-by-character): FV tokenizes words and includes punctuation
2. **Normalization**
   - ('&' = 'and')
3. **Alignment**
   - **I like systems based on tokenized words (characters seems too overkill for me)**
   - **'includes punctuation' -- in the word or as <pc>?**
4. - **Always worried about normalization steps... what is lost? (Assuming nothing here, as is merely for collation?)**
5. **Visualization**

Gothenburg Model

algorithm for computer-aided collation, developed in 2009 workshop of collateX and Juxta developers.

1. **Tokenization** :
   - Break down the smallest unit of comparison: (words--with punctuation, or character-by-character): FV tokenizes words and includes punctuation
2. **Normalization**
   - ('&' = 'and')
3. **Alignment**
   - **I like systems based on tokenized words (characters seems too overkill for me)**
   - **'includes punctuation' -- in the word or as <pc>?**
4. - **Always worried about normalization steps... what is lost? (Assuming nothing here, as is merely for collation?)**
5. **Visualization**

Gothenburg Model

algorithm for computer-aided collation, developed in 2009 workshop of collateX and Juxta developers.

1. **Tokenization** :
   - Break down the smallest unit of comparison: (words--with punctuation, or character-by-character): FV tokenizes words and includes punctuation
2. **Normalization**
   - ('&' = 'and')
3. **Alignment**
   - **I like systems based on tokenized words (characters seems too overkill for me)**
   - **'includes punctuation' -- in the word or as <pc>?**
   - **Always worried about normalization steps... what is lost? (Assuming nothing here, as is merely for collation?)**
4. automated processing)
5. **Visualization**

Gothenburg Model

algorithm for computer-aided collation, developed in 2009 workshop of collateX and Juxta developers.

1. **Tokenization** :
   - Break down the smallest unit of comparison: (words--with punctuation, or character-by-character): FV tokenizes words and includes punctuation
2. **Normalization**
   - ('&' = 'and')
3. **Alignment**
   - **I like systems based on tokenized words (characters seems too overkill for me)**
   - **'includes punctuation' -- in the word or as <pc>?**
   - **Always worried about normalization steps... what is lost? (Assuming nothing here, as is merely for collation?)**
4. automated processing)
5. **Visualization**
   - critical edition apparatus, graph displays

Collating with markup: "handsome" / "beautiful" passage

```
 <app xml:id="C10_app44">
          <rdgGrp xml:id="C10_app44_rg1"
n="['&lt;del&gt;handsome&lt;del&gt;
&lt;del&gt;handsome&lt;
del&gt;beautiful.&lt;del&gt;handsome&lt;del&gt;beautiful;', 'great']"

            <rdg wit="fMS">&lt;lb n="c56-0045__main__23"/&gt;
  &lt;del rend="strikethrough" sID="c56-0045__main__d2e9837"/&gt;
handsome&lt;del eID="c56-0045__main__d2e9837"/&gt;
&lt;mdel&gt;.
&lt;/mdel&gt;&lt;lb n="c56-0045__left_margin__1"/&gt;
&lt;del rend="strikethrough" sID="c56-0045__left_margin__d2e9853"/&gt;handsome&lt;
del eID="c56-0045__left_margin__d2e9853"/&gt;beautiful.
&lt;del rend="strikethrough" sID="c56-0045__main__d2e9865"/&gt;
Handsome&lt;del eID="c56-0045__main__d2e9865"/&gt;
Beautiful; Great </rdg>
       </rdgGrp>

        <rdgGrp xml:id="C10_app44_rg2" n="['beautiful.', 'beautiful!—great']">
              <rdg wit="f1818">beautiful. Beautiful!—Great </rdg>
              <rdg wit="f1823">beautiful. Beautiful!—Great </rdg>
              <rdg wit="fThomas">beautiful. Beautiful!—Great </rdg>
              <rdg wit="f1831">beautiful. Beautiful!—Great </rdg>
        </rdgGrp>
</app>
```

Collating with markup: "handsome" / "beautiful" passage

```
 <app xml:id="C10_app44">
          <rdgGrp xml:id="C10_app44_rg1"
n="['&lt;del&gt;handsome&lt;del&gt;
&lt;del&gt;handsome&lt;
del&gt;beautiful.&lt;del&gt;handsome&lt;del&gt;beautiful;', 'great']"

            <rdg wit="fMS">&lt;lb n="c56-0045__main__23"/&gt;
  &lt;del rend="strikethrough" sID="c56-0045__main__d2e9837"/&gt;
handsome&lt;del eID="c56-0045__main__d2e9837"/&gt;
&lt;mdel&gt;.
&lt;/mdel&gt;&lt;lb n="c56-0045__left_margin__1"/&gt;
&lt;del rend="strikethrough" sID="c56-0045__left_margin__d2e9853"/&gt;handsome&lt;
del eID="c56-0045__left_margin__d2e9853"/&gt;beautiful.
&lt;del rend="strikethrough" sID="c56-0045__main__d2e9865"/&gt;
Handsome&lt;del eID="c56-0045__main__d2e9865"/&gt;
Beautiful; Great </rdg>
      </rdgGrp>
```

- **Escaping XML like this frightens me. If our collation systems need to do this, maybe we need to improve our systems!**
- **That said, I don't necessarily have a better solution for this use case.**

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>
```

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>
```

Publishing a Stand-off Critical Apparatus

Publishing a Stand-off Critical Apparatus

- Raffaele Viglianti provides us with interesting information on the publication of a spine-based model similar to the Frankenstein Variorum but also using Early Modern Soundscapes project

Publishing a Stand-off Critical Apparatus

- Raffaele Viglianti provides us with interesting information on the publication of a spine-based model similar to the Frankenstein Variorum but also using Early Modern Soundscapes project
- Important to note separation of encoding choices and editorial decisions from the system of modelling variance

Publishing a Stand-off Critical Apparatus

- Raffaele Viglianti provides us with interesting information on the publication of a spine-based model similar to the Frankenstein Variorum but also using Early Modern Soundscapes project
- Important to note separation of encoding choices and editorial decisions from the system of modelling variance
- Useful reminder that critical apparatus might not be of text, but potentially of music, or text&music

Publishing a Stand-off Critical Apparatus

- Raffaele Viglianti provides us with interesting information on the publication of a spine-based model similar to the Frankenstein Variorum but also using Early Modern Soundscapes project
- Important to note separation of encoding choices and editorial decisions from the system of modelling variance
- Useful reminder that critical apparatus might not be of text, but potentially of music, or text&music
- Demonstrates again that we still need much better tools for standOff critical appartus and creating such spines

Songscapes TEI stand-off collation

TEI (no XPointer)

BL Add. MS 53723

+


```
1  <TEI>
2    <div>
3      <head>Text Collation</head>
4      <app>
5        <rdgGrp>
6          <rdg wit="#BL_53723">
7            <ptr target="tei/Ariadne-BL_53723.xml#v1"/>
8          </rdg>
9          <rdg wit="#L638">
10           <ptr target="tei/Ariadne-L638.xml#v1"/>
11         </rdg>
12       </rdgGrp>
13       <rdg wit="#C709">
14         <ptr target="tei/Ariadne-C709.xml#v1"/>
15       </rdg>
16     </app>
17   </div>
18 </TEI>
```

Folger L638

+

C 709

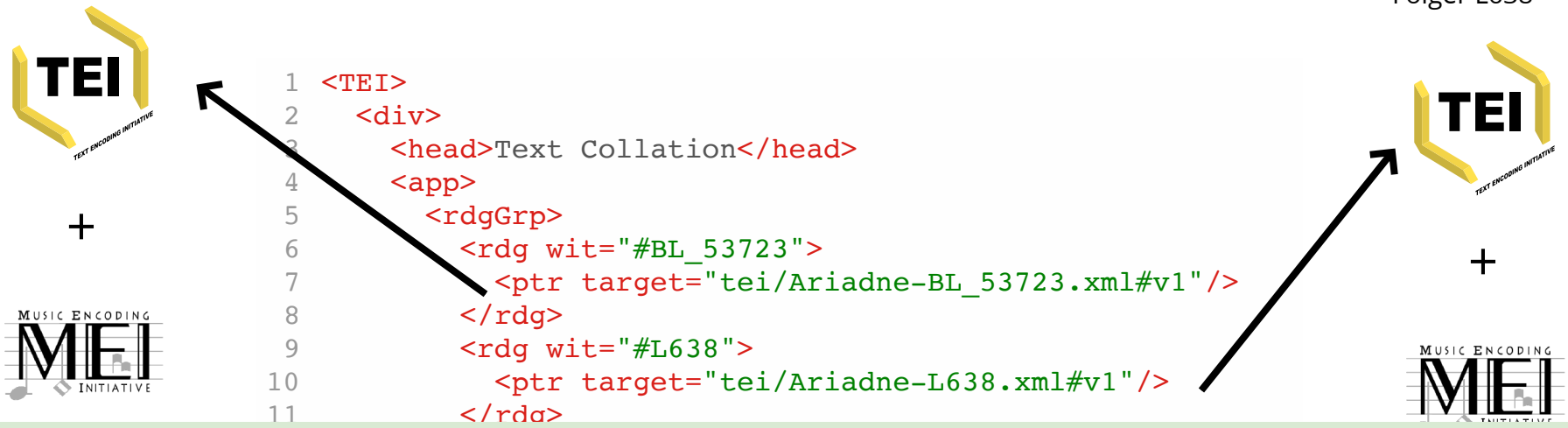
Adapted from:
https://github.com/EarlyModernSongscapes/songscapes/blob/master/data/collations/Theseus%2C_O_Theseus%2C_hark!.xml

Songscapes TEI stand-off collation

TEI (no XPointer)

BL Add. MS 53723

Folger L638



```
<TEI>
  <div>
    <head>Text Collation</head>
    <app>
      <rdgGrp>
        <rdg wit="#BL_53723">
          <ptr target="tei/Ariadne-BL_53723.xml#v1"/>
        </rdg>
        <rdg wit="#L638">
          <ptr target="tei/Ariadne-L638.xml#v1"/>
        </rdg>
```

- **Feel more comfortable with ID-based pointers**
- **Why not use a pointing attribute on <rdg> for compact markup? (But which one? @corresp?)**
  - **Do we need @target on <rdg>? Or @from/@to as on <app>?**

Adapted from:
https://github.com/EarlyModernSongscapes/songscapes/blob/master/data/collations/Theseus%2C_O_Theseus%2C_hark!.xml

Publishing this kind of model

- implications of using a stand-off apparatus for driving a digital publication:
  - typical TEI -> HTML transformation requires transforming pointers too
  - pointers may need to be followed in response to user interaction

Publishing this kind of model

- implications of using a stand-off apparatus for driving a digital publication:
  - typical TEI -> HTML transformation requires transforming pointers too
  - pointers may need to be followed in response to user interaction

- **What happens when pointers can no longer be followed?**
- **In examples shown is spine truly needed? Or could same data be generated and stored in minimal if redundant copies in each edition?**
- **Useful beyond single project; Does this lead to a need for meta-spine edition of spine-based editions?**

Some General Concerns

Some General Concerns

- **Scalability:** How do these approaches scale, not only to hundreds of witnesses but to thousands?

Some General Concerns

- **Scalability:** How do these approaches scale, not only to hundreds of witnesses but to thousands?
- **Fragility:** For any stand-off solution does how we point to things depend on how likely they are to change, move, or die?

Some General Concerns

- **Scalability:** How do these approaches scale, not only to hundreds of witnesses but to thousands?
- **Fragility:** For any stand-off solution does how we point to things depend on how likely they are to change, move, or die?
- **Practicality:** How easy to adopt with less technical burden?

Some General Concerns

- **Scalability:** How do these approaches scale, not only to hundreds of witnesses but to thousands?
- **Fragility:** For any stand-off solution does how we point to things depend on how likely they are to change, move, or die?
- **Practicality:** How easy to adopt with less technical burden?
- **Readability:** Just possible with id-based systems and small number of witnesses, becomes opaque with string-ranges or many witnesses. *We need better stand-off encoding software.*

Some General Concerns

- **Scalability:** How do these approaches scale, not only to hundreds of witnesses but to thousands?
- **Fragility:** For any stand-off solution does how we point to things depend on how likely they are to change, move, or die?
- **Practicality:** How easy to adopt with less technical burden?
- **Readability:** Just possible with id-based systems and small number of witnesses, becomes opaque with string-ranges or many witnesses. *We need better stand-off encoding software.*
- **Incompatible granularity:** TEI critical apparatus now enables you to have a <rdg> with phrase content next to one with a or <floatingText> inside. Does this cause limitations when making comparisons?

Some General Concerns

- **Scalability:** How do these approaches scale, not only to hundreds of witnesses but to thousands?
- **Fragility:** For any stand-off solution does how we point to things depend on how likely they are to change, move, or die?
- **Practicality:** How easy to adopt with less technical burden?
- **Readability:** Just possible with id-based systems and small number of witnesses, becomes opaque with string-ranges or many witnesses. *We need better stand-off encoding software.*
- **Incompatible granularity:** TEI critical apparatus now enables you to have a <rdg> with phrase content next to one with a or <floatingText> inside. Does this cause limitations when making comparisons?
- **Future:** What is next for TEI critical apparatus? How much should TEI legislate form of particular stand-off approaches? Tradeoffs between flexibility and constraint?

Document Modeling with the TEI Critical Apparatus

A Panel for the TEI 2019 Conference in Graz, Austria

Presenters: Hugh Cayless (@hcayless), Elisa Beshero-Bondar (@epyllia), Raffaele Viglianti (@raffazizzi)

Respondent: James Cummings (@jamescummings)