

# Bicentennial Bits and Bytes



## The Pittsburgh Digital Frankenstein Project

Rikk Mulligan | Elisa Beshero-Bondar | Matt Lavin | Jon Klancher  
@CritRikk | @epyllia | @mjLavin80 | @jklancher

MLA 2018: **Saturday Jan. 6 @ 3:30pm**; Sheraton Riverside Suite

Link to these slides: <http://bit.ly/BicFrankMLA18>

## Speaker notes

Hello. Thank you for sharing your time with us, especially late on a Saturday afternoon. Just to make sure you are in the right place, this is session 632: Bits and Bytes: The Pittsburgh Digital Frankenstein Project.

I and my fellow panelists will be describing several aspects of our project over the next forty minutes or so to set up our roundtable discussion with you.

I'm Rikk Mulligan; my fellow panelists are Elisa Beshero-Bondar, Matthew Lavin, and Jon Klancher. I'll begin by introducing our project and explaining how we came together, as well as what each of us are bringing to the project and how this helped us define our initial project plans. My co-panelists will go in depth into our current phase and we'll end with some discussion of our next steps and May 2018 milestone.

# A Patchwork Team

- **Elisa Beshero-Bondar**, Director, Center for the Digital Text, University of Pittsburgh at Greensburg
- **Jon Klancher**, English Department, Carnegie Mellon University
- **Matt Lavin**, Director, Digital Media Lab, University of Pittsburgh
- **Rikk Mulligan**, University Libraries, Carnegie Mellon University
- **Raff Vigiante**, Maryland Institute for Technology in the Humanities (MITH), University of Maryland
- **Scott Weingart**, Program Director of Digital Humanities, University Libraries, Carnegie Mellon University



## Speaker notes

Most Humanities research is done by individuals, but Digital Scholarship tends to require a team, often directed by a principal investigator with a well-defined research agenda. Our team is somewhat different--we formed around the opportunity to contribute to a project focused on the bicentennial anniversary of the publication of Mary Shelley's novel Frankenstein. We came together organically through online and face-to-face relationships because of the possibilities of working together rather than as part of a specific goal or design.

The Pittsburgh Digital Frankenstein Project began to coalesce during October 2016, but its beginnings go back to August 2016 when Neil Fraistat of MITH--the Maryland Institute for Technology in the Humanities--contacted Scott Weingart to ask if he might be interested in doing some visualizations on the materials in the Shelley Godwin Archive.

A couple of weeks later, while Scott and I were discussing an unrelated science fiction project, he asked if I had any interest in Frankenstein. I'd just learned that CMU has a copy of the first edition in our special collections and thought this would be a great opportunity to explore new DH methods and highlight our special collections through a DH project.

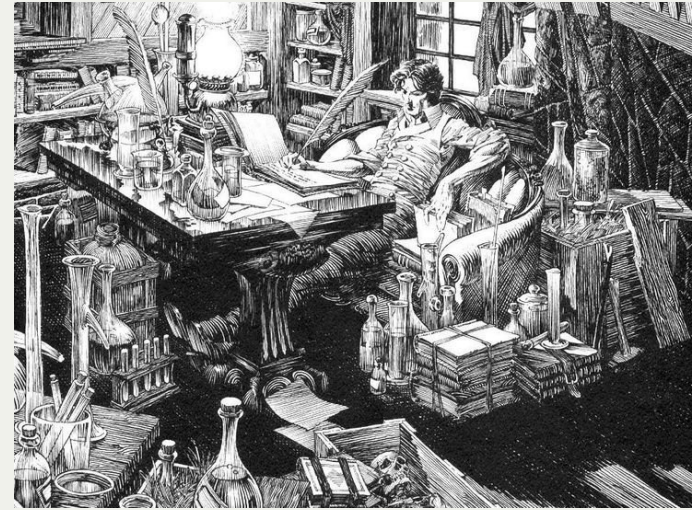
Scott then video chatted with Neil, Raff Vigilanti, David Rettenmaier, and Purdom Lindblad at MITH in September to begin scoping out the project. He then sent out an email to the four of us in the Pittsburgh area (Elisa, Matt, Jon, and myself), introducing us and asking if we might explore potential goals.

Although Raff was not involved in our first meetings, Elisa was already in contact with him and had previously arranged for him to speak at Pitt-Oakland and Pitt-Greensburg on music encoding and its applications. Elisa and Raff had also been in touch with Neil about contributing an update to the edition of Frankenstein on the Romantic Circles website, one that would interweave with the Shelley-Godwin Archive's edition of the Frankenstein manuscript Notebooks. Because Elisa and Raff work together as members of the TEI Technical Council, Neil hoped they might work together on the interconnection of those editions, as part of the "Pittsburgh group" collaboration.

Jon accepted Scott's invitation because of his involvement in the Romantic Bicentennials as Co-Director of Networked Events, which has several upcoming Frankenstein projects. As the only Romanticist at CMU we were hoping he would participate and he was looking at this as a chance to work with a DH team.

Scott and I are members of CMU's new digital center, dSHARP, and Matt is one of those who attended our open office/open consulting hours last year. After hearing about the project after our first couple of meetings he opted to join in the late fall of 2016.

# What We Contribute



Illustrations by Bernie Wrightson.  
*Frankenstein*. Mary Wollstonecraft Shelley. Marvel Comics 1983.  
images from Dark Horse Comics 2008 reprint.

- **Elisa Beshero-Bondar:** *Romanticist*; Textual Scholar; [TEI architecture and collation](#)
- **Jon Klancher:** *Romanticist*; Book Historian; [Annotations](#)
- **Matt Lavin:** 19th Century Americanist; [Textual Analysis](#), [Stylometry](#)
- **Rikk Mulligan:** 20th Century Americanist; [Web Coding](#), [Interface Design](#)
- **Raffaele Vigiante:** Research Programmer; [Shelley-Godwin Archive encoding](#);  
[TEI pointers to S-GA Notebooks](#)
- **Scott Weingart:** Early Modernist, *History of Science*; [Textual Analysis](#), [Stylometry](#)

## Speaker notes

Our team is eclectic--a patchwork of institutions, roles, and levels of technical expertise and DH experience. This is our strength. Even as Victor's "patchwork" construct is greater than the sum of its parts (especially in modern reimaginations), so is our goal to construct a digital edition with features that bridge print and digital resources. We also approach the project from very different perspectives and can learn a lot from each other. As Romanticists Jon and Elisa are our subject matter experts on Shelley's era and the text; both have several years experience teaching a variety of Frankenstein editions. Scott expertise in the history and philosophy of science will also contribute to Jon's annotations of the text. Elisa's extensive experience with the Text Encoding Initiative technologies including TEI XML and XSLT transformations are providing the framework for our online texts. She has worked with Raff to architect our collations and consulted with David Birnbaum at the University of Pittsburgh on some of the thornier coding issues. Raff is also contributing from MITH to help us integrate the Shelley-Godwin Archive tags and annotations, as well as later helping with the TEI pointers for our future interface efforts.

Both Matt and Scott are DH generalists with extensive experience with a range of tools, methods, and projects. They began working on the textual analysis of our materials once Elisa and Raff got them in the shape they required. They are currently working on the stylometrics. [In this case, programmatic approaches to the study of measurable features of (literary) style, such as sentence length, vocabulary richness and various frequencies (of words, word lengths, word forms, etc.) with practical applications in authorship attribution research.]

I only had a brief flirtation with TEI XML before this project. I've been working closely with Elisa to create the clean text files of the print editions and to integrate the notebooks into our corpus. My expertise lies more in web and interface design, which will come more into play as we evolve the presentation of our results.

# Print publications

known/authorized by MWS

- 1818 Edition (3 volumes)
- 1823 Edition (2 volumes)
- 1831 Edition (1/2 of a volume)
  - bound with Friedrich von Schiller's *The Ghost Seer* in Bentley's Standard Series of novels)



Illustrations by Bernie Wrightson.  
*Frankenstein*. Mary Wollstonecraft Shelley. Marvel Comics 1983.  
images from Dark Horse Comics 2008 reprint.

## Speaker notes

But I'm getting ahead of myself. Before we could decide what directions we might go in for our digital project, we needed to survey the current Frankenscape. Although Elisa, Jon, and Raff were well acquainted with Frankenstein, the rest of us were only passing familiar with an edition of the novel. We began by looking at the three print editions to see what we might contribute to the digital scholarship.

It is estimated that 500 copies of Frankenstein were published anonymously on January 1, 1818, in three small volumes. The current Romantic Circles edition attributes the Preface to Percy Shelley although neither the dedication or preface are signed or initialed in the actual copy. As a number of scholars have pointed out (Charles Robinson, Susan Tyler Hitchcock) the novel quickly inspired a number of stage adaptations. These proved so successful that Mary's father, William Godwin, supervised the editing and republication of a two volume edition in 1823. The changes in this edition are so minor, which may explain why we could find no digital edition and had to digitize it ourselves from a photo facsimile. The last edition released during Mary Shelley's lifetime appeared in 1831. It is important to note that the changes from the 1818 and 1823 to the 1831 are fairly extensive, and that everyone who involved in the ghost story contest and earlier text other than Mary had died by the time she released this edition with her story of its genesis in the introduction.



# Digital Sources

- **Pennsylvania Electronic Edition**

- 1994 start (early HTML, frame-based)
- <http://knarf.english.upenn.edu/>

- **Romantic Circles**

- 2009 update from HTML to TEI XML
- <https://www.rc.umd.edu/editions/frankenstein>

- **Manuscript Notebooks**

Bodleian Arbinger c56, c57, c58

- Shelley-Godwin Archive
- TEI XML
- <http://shelleygodwinarchive.org/contents/frankenstein/>



Illustrations by Bernie Wrightson.  
*Frankenstein*. Mary Wollstonecraft Shelley. Marvel Comics 1983.  
images from Dark Horse Comics 2008 reprint.

## Speaker notes

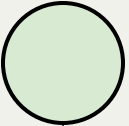
We started with a goal to prepare an updated and improved digital edition of Mary Shelley's *Frankenstein* that conforms to the TEI P5 standard. Elisa and Raff knew that much of this material already existed online.




The earliest work dates back to the 1990s, in Stuart Curran and Jack Lynch's University of Pennsylvania Electronic Edition. This website uses frames to display the 1818 and 1831 editions, as well as a "variorum" frame to show differences. This site also hosts hundreds of additional files as hypertext annotations and connections.

The Romantic Circles website is a refereed scholarly resource; it published versions of the 1818 and 1831 editions upcoded from the PAEE HTML to TEI 3 in 2009. This site also offers a visual comparison of the texts, hosted by the Juxta Commons.

Finally, the Shelley-Godwin Archive has the transcribed Abinger Manuscript Notebooks, c56, c57 and c58 currently in the Bodleian Library of Oxford. The archive focuses on the manuscripts, providing access to the extant fair copy of *Frankenstein* in three forms and a range of critical materials. [(1) the physical order of Bodleian MS Abinger c.58; (2) the virtually reconstituted order of Notebooks C1 and C2; and (3) the linear chapter sequence of the the three-volume fair copy.]


# Critical and Diplomatic Editions Leading to the Pgh Frankenstein Project

 **1974** **Rieger:** inline collation of "Thomas" w/ 1818, 1831 variants in endnotes


  
  
 **~mid-1990s** **Curran and Lynch: *PA Electronic Edition (PAEE)***, collation of 1818 and 1831: HTML

**1996** **Crook** crit. ed of 1818, variants of "Thomas", 1823, and 1831 in endnotes (P&C MWS collected works)


**C. Robinson**, *The Frankenstein Notebooks* (Garland): print facsimile of 1816 ms drafts

 **2007** **Romantic Circles TEI conversion of PAEE**; separates the texts of 1818 and 1831; collation via Juxta

 **2013** **Shelley-Godwin Archive** publishes diplomatic edition of 1816 ms drafts

 **2017** **Pittsburgh Bicentennial Frankenstein Project** begins: assembly/proof-correcting of PAEE files; OCR/proof-correcting 1823; "bridge" TEI edition of S-GA notebook files; automated collation; incorporating "Thomas" copy text

Legend:

 print edition

 digital edition

## Speaker notes

To develop our “improved and updated” digital edition we needed to think ask the question: “what are the authoritative scholarly and critical editions of Frankenstein”?

We've gone back to looking at James Rieger's University of Chicago Press annotated edition from 1974. It is one of the first to emphasize the 1818 text; it also included the “Thomas” edition of 1823 (Mary Shelley annotated 1818 copy) in its analysis.

The first online collated edition, Curran and Lynch had only compared the 1818 and 1831 editions.

Nora Crook and Charles Robinson made path-breaking scholarly editions--one critical, and one of the Abinger ms notebooks--in print, not connected to Curran's work.

Romantic Circles might be considered an update if not an upgrade of the PAEE, using programmatic tools in Juxta Commons rather than hand-collation.

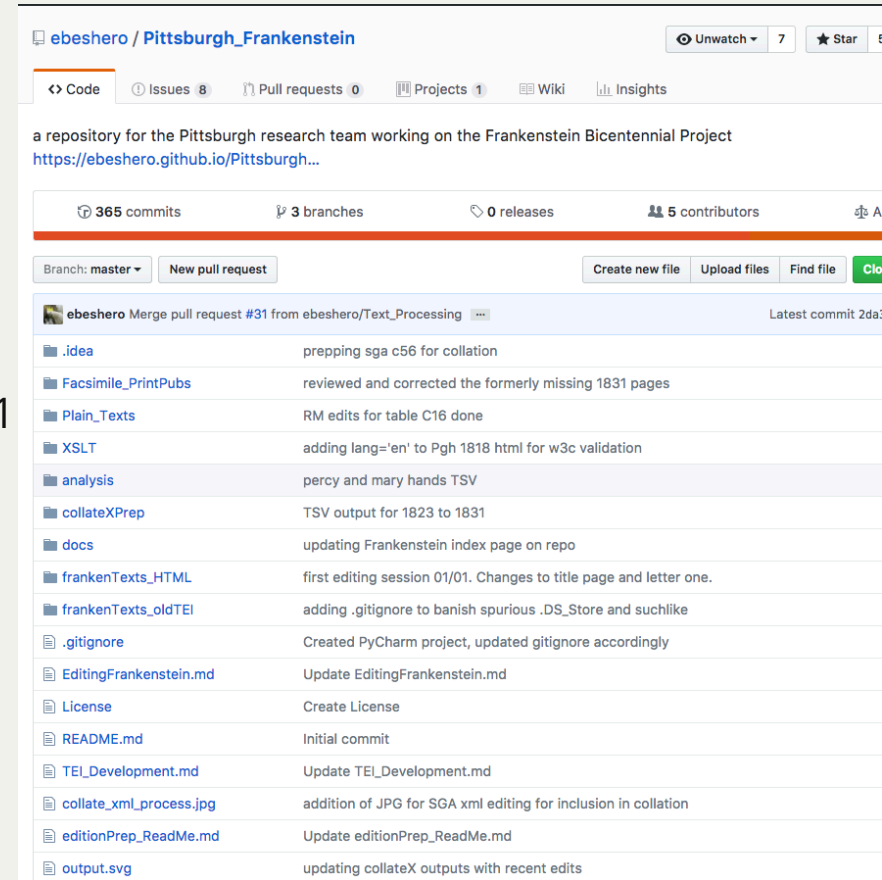
The Shelley-Godwin Archive went online with the Notebooks digital edition in 2013, making a diplomatic edition publicly available.

Our work is a bridge: We bring all of these nodes of scholarship together by considering the Abinger manuscript notebooks, the 1818, 1823, and 1831 editions, and Mary Shelley's annotations in the “Thomas” edition. We do not seek to replace the scholarly apparatus created by others, but to build upon, correct, update, and connect this fine work. We are cognizant that some of the work has errors we can fix or was built with almost-obsolete web technologies--needing a more stable digital architecture.

# Evolving Project

Returning to the original texts to produce:

- Clean Text files for each edition (1818, 1823, 1831)
- TEI XML files for each edition
- Comprehensive collations from ms through 1831 to bridge and build on previous critical editions (print and digital)
- Variorum interface to show changes over time
- Stylometric analysis
- Annotations



[https://github.com/ebeshero/Pittsburgh\\_Frankenstein](https://github.com/ebeshero/Pittsburgh_Frankenstein)

## Speaker notes

We have completed several steps toward preparing a new, scholarly digital edition of Frankenstein in TEI 5. We have also started the work to offer additional scholarly resources online as part of this project.

We maintain a GitHub repository to track and share our work and render our efforts as transparent as possible.

As of now we have completed plain text editions of the 1818, 1823, and 1831 editions on our GitHub. We have added comments about our initial attempts to use the PAEE and RC materials, and explained the process we used to produce the clean text files.

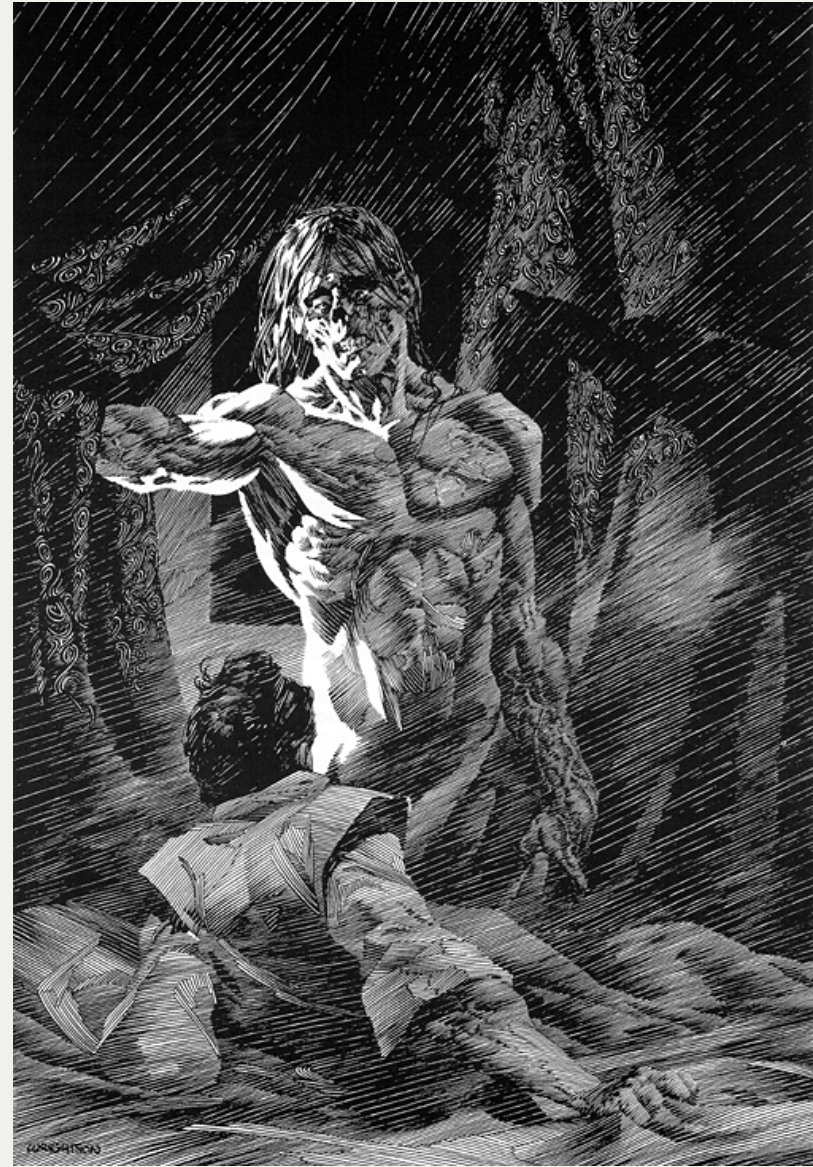
Elisa will go into more detail regarding our efforts to prepare the TEI XML editions and collation files. She will also present our work to date on developing a comprehensive collation of all five sources, and sketch out our goals in developing a visual variorum interface to display the differences between them.

Matt will discuss our current work on the stylometric analysis. For those who are unfamiliar, stylometrics in this cases uses programmatic approaches to study the measurable features of (literary) style; this can include sentence length, vocabulary richness and the various frequencies of words, word lengths, and word forms as part of authorship attribution research.

Jon will talk to you about how we hope to augment the annotations of previous scholarly editions and both the online resources with something different.

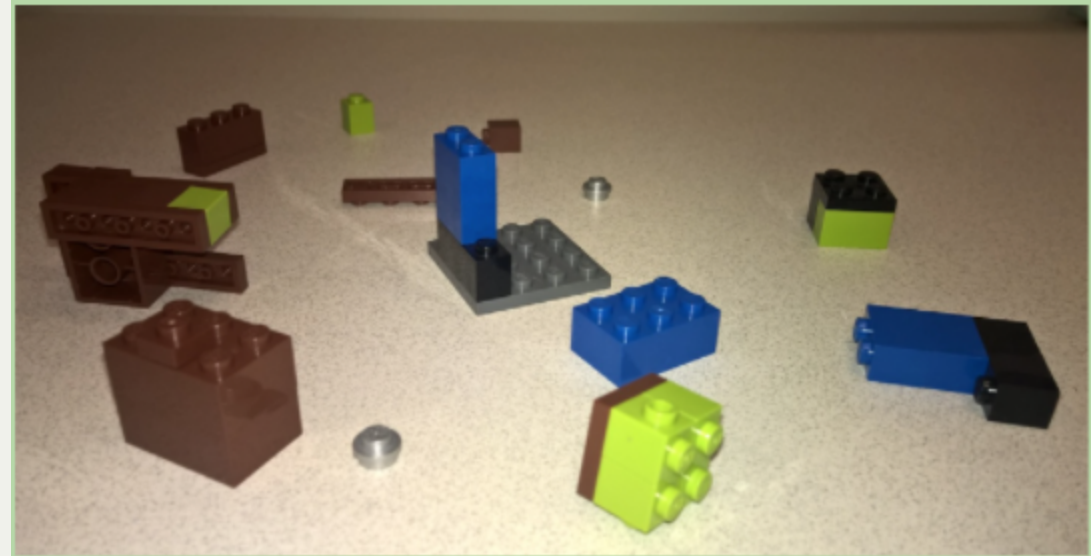
# Variorum project

- **Manuscript:**  
(Notebooks: Abinger c56, c57, c58)
- **"Thomas copy" Edition** (1818 edition with hand annotations by Mary Shelley)
- **1818 Edition** (3 volumes)
- **1823 Edition** (2 volumes)
- **1831 Edition** (1/2 of a volume)



Illustrations by Bernie Wrightson.  
*Frankenstein*. Mary Wollstonecraft Shelley. Marvel Comics 1983.  
images from Dark Horse Comics 2008 reprint.

# Building a Digital Variorum




Elisa Beshero-Bondar  
@epyllia





# Motivating Questions

- Can we make an edition that conveniently compares the manuscripts to the print publications?
- Can we make a comprehensive collation to show changes to the novel over time, from 1816 to 1831?
  - How many versions? (5 and a bit?)
  - Which editorial interventions persist from 1816 to 1831?
    - MWS in the "Thomas" copy: how much of this persists into 1831?
    - PBS's additions: which/how many of these persist to 1831?
    - What parts of the novel were most mutable?


## Our Project Genealogy: Critical and Diplomatic Editions Leading to the Pgh Frankenstein Project

 **1974** **Rieger:** inline collation of "Thomas" w/ 1818, 1831 variants in endnotes


 **~mid-1990s** **Curran and Lynch: *PA Electronic Edition (PAEE)***, collation of 1818 and 1831: HTML

 **1996** **Crook** crit. ed of 1818, variants of "Thomas", 1823, and 1831 in endnotes (P&C MWS collected works)


**C. Robinson, *The Frankenstein Notebooks*** (Garland): print facsimile of 1816 ms drafts

 **2007** ***Romantic Circles*** TEI conversion of **PAEE**; separates the texts of 1818 and 1831; collation via Juxta

 **2013** ***Shelley-Godwin Archive*** publishes diplomatic edition of 1816 ms drafts

 **2017** ***Pittsburgh Bicentennial Frankenstein Project*** begins: assembly/proof-correcting of PAEE files; OCR/proof-correcting 1823; "bridge" TEI edition of S-GA notebook files; automated collation; incorporating "Thomas" copy text

Legend:

 print edition

 digital edition

## Speaker notes

"We stand on the shoulders of giants." As Rikk mentioned, we're building on a lineage of Frankenstein edition work that illuminates the novel's gestation and transformation.

Rieger: INLINE collation: wants to show MWS's process, as she handwrote alterations on a copy of the 1818 Frankenstein. As you read Rieger, you see a system of markup distinguishing MWS's hand from the print, showing her insertions and additions. It's important for FOREGROUNDING the act of comparing texts in the line of the reader's sight--precedent for Curran's PA Electronic Edition.

1990s: BURST of activity! Frankenstein was a major EARLY experiment to build a scholarly apparatus in hypertext in the early years of the World Wide Web. The web and print editions improve the scope of comparison--beyond just 1818 vs 1831. Charlie Robinson's facsimile edition of the Abinger manuscript notebooks illuminates the writing process and hands at work over the novel. Much more is now available to the scholar to document CHANGE OVER TIME.

# The dream of the 90s...

Hypertext / Hypercard books and the *PAEE*

## *Frankenstein; or, the Modern Prometheus*

By [Mary Wollstonecraft Shelley](#)

<a href="#">Table of Chapters</a>	<a href="#">Table of Frames</a>
<a href="#">Contents</a>	<a href="#">Critical Commentary</a>
<b>Master Index</b>  <a href="#">Biographies</a> <a href="#">Characters</a> <a href="#">Chronology</a> <a href="#">Contexts</a> <a href="#">Critical Commentary</a> <a href="#">Illustrations</a> <a href="#">Geography</a> <a href="#">Maps</a> <a href="#">Works Included in this Edition</a>	

- Accessing (reading, writing, editing) texts in nonlinear ways
- Multiplying and individualizing points of access

# The dream of the '90s:

## *Frankenstein's* inspiration for hypertext experiment

“Hypertext’s next step . . .

### P A T C H W O R K G I R L

is spectacular in every sense (from Rayban to Debordian to Cirque du Soleil). It is embodied writing in the company of Helene Cixous, Carolyn Guyer, Jeanette Winterson, and its progenitor and co-author (the other Shelley) Mary.

Jackson weaves aspects of the Frankenstein tale, graphic novels, feminist meditation, and pure storytelling into a riveting work of multiple fiction with a fugue-like poetry and an artist’s eye distinctly its (and her) own.

This is a work of dream and desire and defying boundaries, an electronic collage, a theatre of windows, and a cyborg song of communion and reunion.”

— MICHAEL JOYCE —



by S H E L L E Y J A C K S O N

ISBN 1-884511-23-6

For Macintosh or Windows, \$19.95

A self-described “student of the art of digression,” Shelley Jackson holds an AB in studio art from Stanford University and an MFA in creative writing from Brown University. Her work has appeared in *Degenerative Prose* and various journals, including *Conjunctions*. She has written and illustrated a children’s book, *The Old Woman and the Wave*, to be published by Orchard Books.



- Roughly contemporary with the *PAEE* + mid '90s scholarly edition efforts
- What if the female creature survived and had a chance to create her own story with lots of options?
- Experimental nonlinear navigation...hundreds of hypercards...plot your own course

## Speaker notes

I've included this slide to emphasize how the 1990s marked a moment of experiments with Frankenstein as a body of text--The building of a Creature and the assemblage of a textual body become continuous activities--the reader participates. It isn't just that we theorized with Stanley Fish about readers assembling texts for themselves; the experimenters with Frankenstein like Shelley Jackson and Stuart Curran invited readers to engage actively in remixing, juxtaposing, and exploring options.

# PAEE: Hypertext Collation Experiment

1818



Chapter



Contents Index



Frame



1831

When I was about fifteen years old, we had retired to our house near Belrive, when we witnessed a most violent and terrible thunder-storm. It advanced from behind the mountains of Jura; and the thunder burst at once with frightful loudness from various quarters of the heavens. I remained, while the storm lasted, watching its progress with curiosity and delight. As I stood at the door, on a sudden I beheld a stream of fire issue from an old and beautiful oak, which stood about twenty yards from our house; and so soon as the dazzling light vanished, the oak had disappeared, and nothing remained but a blasted stump. When we visited it the next morning, we found the tree shattered in a singular manner. It was not splintered by the shock, but entirely reduced to thin ribbands of wood. I never beheld anything so utterly destroyed.

The catastrophe of this tree excited my extreme astonishment; and I eagerly inquired of my father the nature and origin of thunder and lightning. He replied, "Electricity;" describing at the same time the various effects of that power. He constructed a small electrical machine, and exhibited a few experiments; he made also a kite, with a wire

When I was about fifteen years old, we had retired to our house near Belrive, when we witnessed a most violent and terrible thunder-storm. It advanced from behind the mountains of Jura; and the thunder burst at once with frightful loudness from various quarters of the heavens. I remained, while the storm lasted, watching its progress with curiosity and delight. As I stood at the door, on a sudden I beheld a stream of fire issue from an old and beautiful oak, which stood about twenty yards from our house; and so soon as the dazzling light vanished, the oak had disappeared, and nothing remained but a blasted stump. When we visited it the next morning, we found the tree shattered in a singular manner. It was not splintered by the shock, but entirely reduced to thin ribands of wood. I never beheld anything so utterly destroyed.

Before this I was not unacquainted with the more obvious laws of electricity. On this occasion a man of great research in natural philosophy was with us, and, excited by this catastrophe, he entered on the explanation of a theory which he had formed on the subject of electricity and galvanism, which was at once new and astonishing to me. All

When I was about fifteen years old, we had retired to our house near Belrive, when we witnessed a most violent and terrible thunder-storm. It advanced from behind the mountains of Jura; and the thunder burst at once with frightful loudness from various quarters of the heavens. I remained, while the storm lasted, watching its progress with curiosity and delight. As I stood at the door, on a sudden I beheld a stream of fire issue from an old and beautiful oak, which stood about twenty yards from our house; and so soon as the dazzling light vanished, the oak had disappeared, and nothing remained but a blasted stump. When we visited it the next morning, we found the tree shattered in a singular manner. It was not splintered by the shock, but entirely reduced to thin **[ribbands]** <ribbands> of wood. I never beheld anything so utterly destroyed.

**[The catastrophe of this tree excited my extreme astonishment; and I eagerly inquired of my father the nature and origin of thunder and lightning. He replied, "Electricity;" describing at the same time the various effects of that power. He constructed a small electrical machine, and exhibited a few experiments; he made also a kite, with a wire and string, which drew down that fluid from the clouds.**

**This last stroke completed the overthrow of Cornelius Agrippa, Albertus Magnus, and Paracelsus, who had so long reigned the lords of my imagination. But by some fatality I did not feel inclined to commence the study of any modern system; and this disinclination was influenced by the following circumstance.]**

hundreds of small html files, juxtaposed in frames

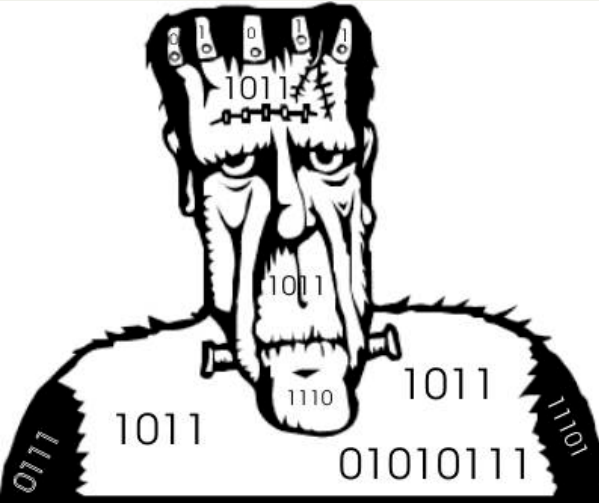
Speaker notes

Discuss HOW this works as collation and what's so exciting about it.



# The dream of the '90s is alive...

## (in Pittsburgh)



source: [I programmer](#) article on "Frankenstein" malware

### Digital Collation for a "Variorum" interface

- select a text from what version the reader chooses:
  - [1816 MS](#) | [1818](#) | ["Thomas"](#) | [1823](#) | [1831](#)
- compare that text to what version the reader chooses
- view the "molten" portions of the novel in context with the stable portions
- navigate multiple texts in context with one another
- make the critical apparatus a vantage point: see how the novel changed over time without having to find the fine-print endnotes

### *The Creature of Collation?*

*We make newly formed text "bodies" from disparately formed source materials.*

# TEI: The Text Encoding Initiative

- a community-maintained standard
- 1987 @ Vassar: draft of Poughkeepsie Principles
  - “ *provide a standard format for data interchange in humanities research.*
  - *Guidelines for the Encoding and Interchange of Machine Readable Texts*: first drafted 1990; published on the web by 1999 (P3)
  - Standards for encoding texts co-evolve with standards for developing **human and machine-readable markup languages**
    - HTML (w3c) || (early) SGML and XML
- TEI XML tree structure:
  - meant to store a stable format not subject to commercial processing requirements
- possible to publish TEI directly or convert to HTML; PDF; TEX; other document formats.

Speaker notes

SGML = standard generalized markup language

XML = eXtensible markup language

## From PAEE to Pgh Variorum... values in common

1. Small pieces are optimal for collation.
2. There is no single "complete" edition.
3. Each output (plain text, XML, TEI collation) = **viable edition on its own.**
4. Interface invites the user to play: put the pieces together.

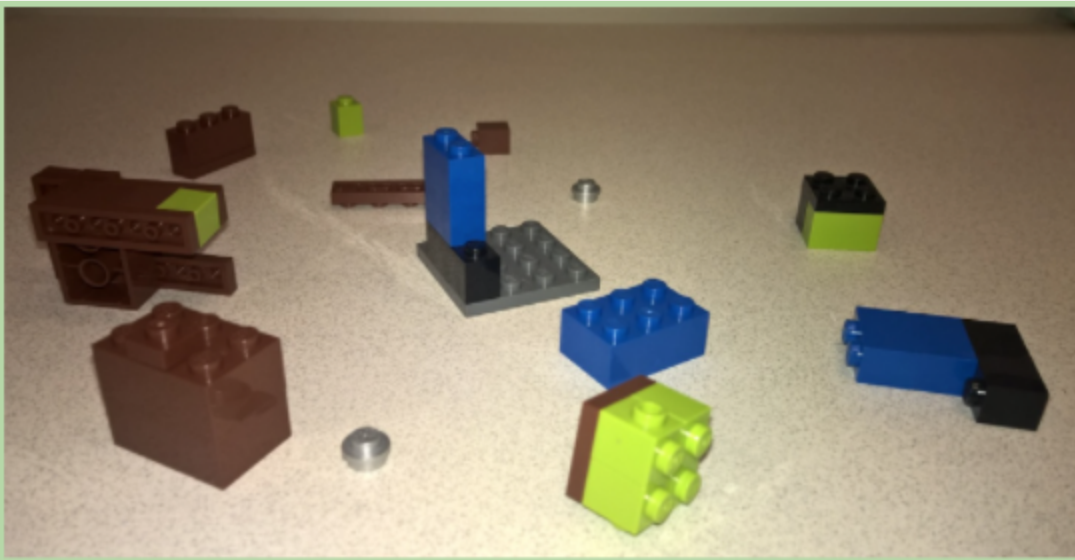


image source: a friend's Lego set

# A Bridge-Building Challenge

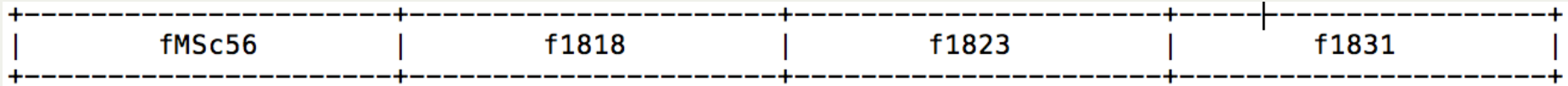


Pittsburgh's bridges (1963)

Source: [NewsCastic.com](http://NewsCastic.com)

- Reconcile multiple kinds of text encoding:
  - old '90s HTML (1818, 1831)
  - not-so-plain OCR-generated text (1823)
  - TEI XML for manuscripts: (S-GA diplomatic edition)
- Construct "Bridge" XML for collation
  - Markup-assisted machine collation (collateX):
  - "flattened" XML hierarchies for even collation units
  - ms metadata markup (e.g. "hands") to *ignore* in collation, but *preserve* in the output
  - pointers outward to manuscript editions (S-GA, Morgan Library)

# Collation "stitchery"



- Can be done by hand in TEI
- Automated: via **CollateX**
  - Algorithms for locating union and "delta" points in "streams" of text
  - Inputs in a variety of formats (XML/TEI, plain text, JSON)
- Output / Visualization options:
  - Text table (above); SVG flow chart; XML
  - JuxtaCommons on the web
  - **Develop a custom web interface (via XML output)**

-	CHAPTER	CHAPTER	CHAPTER
-	IV.	IV.	V.
<surface> <graphic>	</head> <p>	</head> <p>	</head> <p>
</graphic><lb/><milestone spanTo="#c56-0045.04" unit="tei:head"/>Chapter 7<hi rend="sup"><hi rend="underline">th</hi></hi><anchor xml:id="c56-0045.04"/><milestone unit="tei:p"/><lb/>It	I<hi rend="smallcaps">T</hi>	I<hi rend="smallcaps">T</hi>	I<hi rend="smallcaps">T</hi>
was on a dreary night of	was on a dreary night of	was on a dreary night of	was on a dreary night of
November	November,	November,	November,
<lb/>that I beheld <del rend="strikethrough"><add hand="#pbs" place="superlinear">the	that I beheld the	that I beheld the	that I beheld the
frame on whic</add></del>	accomplishment of	accomplishment of	accomplishment of
my	my	my	my
man comple<add place="intra-linear">te</add><add>ed</add><add hand="#pbs" place="intra-linear">,</add>.<del rend="strikethrough">And</del><lb/>with	toils. With	toils. With	toils. With
an anxiety that	an anxiety that	an anxiety that	an anxiety that

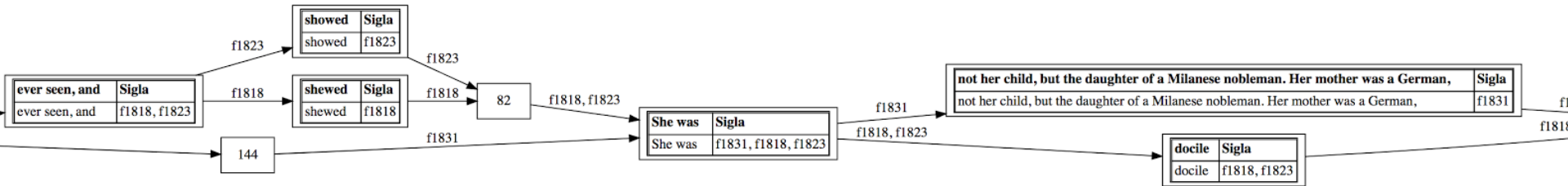
infinite pass  
 to form. His  
 beautiful and I had  
 handsome handsome. A  
 yellow sun skin  
 muscles and  
 of a lustous black, &  
 was flowing as  
 up but then  
 formed a m  
 his watery eye

-	CHAPTER	CHAPTER	CHAPTER
-	IV.	IV.	V.
<surface> <graphic>	</head> <p>	</head> <p>	</head> <p>
</graphic><lb/><mile stone spanTo="#c56-0 045.04" unit="te:he ad"/>Chapter 7<hi rend="sup"><hi rend= "underline">th</hi>< /hi><anchor xml:id=" c56-0045.04"/><miles tone unit="te:p"/>< lb/>It	I<hi rend="smallcaps >T</hi>	I<hi rend="smallcaps >T</hi>	I<hi rend="smallcaps >T</hi>
was on a dreary night of	was on a dreary night of	was on a dreary night of	was on a dreary night of
November	November,	November,	November,
<lb/>that I beheld <del rend="striker ough"><add hand="#pbs" place="s uperlinear">the	that I beheld the	that I beheld the	that I beheld the
frame on whic</add></del>	accomplishment of	accomplishment of	accomplishment of
my	my	my	my
man comple<add place ="intra-linear">te</a dd><add>ed</add><add hand="#pbs" place="i ntra-linear">,</add>. <del rend="striker ough">And</del><lb/> with	toils. With	toils. With	toils. With
an anxiety that almost	an anxiety that almost	an anxiety that almost	an anxiety that almost
amount <lb/>ed	amounted	amounted	amounted
to	to	to	to

image source: S-GA



# A running text stream...? Or an architecture of bridges?



(collateX SVG output)

# XML collation: flagging variants and Percy's hand

```
XPath 3.0 - //app[not(@type='invariant')]
collation_C08.xml x collation_C12.xml x collation_C16.xml x collation_C09.xml x collation_C13.xml x collation_C14.xml x collation_C10.xml x
root
1 type="invariant"><rdg wit="#f1818">the other far more dreadfully </rdg><rdg wit="#f1823">the other far more
1 dreadfully </rdg><rdg wit="#f1831">the other far more dreadfully </rdg><rdg wit="#fMSc56">the &lt;lb
1 n="c56-0092_main_3"/&gt;;other far more dreadfully </rdg></app><app><rdg wit="#f1818">murdered, </rdg><rdg
1 wit="#f1823">murdered, </rdg><rdg wit="#f1831">murdered, </rdg><rdg wit="#fMSc56">murdered </rdg></app><app
1 type="invariant"><rdg wit="#f1818">with every </rdg><rdg wit="#f1823">with every </rdg><rdg wit="#f1831">with
1 every </rdg><rdg wit="#fMSc56">with &lt;lb n="c56-0092_main_4"/&gt;;every </rdg></app><app><rdg
1 wit="#f1818">aggravation of </rdg><rdg wit="#f1823">aggravation of </rdg><rdg wit="#f1831">aggravation of
1 </rdg><rdg wit="#fMSc56">agravation &lt;del rend="strikethrough"&gt;ignominy&lt;/del&gt; </rdg></app><app
1 type="invariant"><rdg wit="#f1818">infamy that could make </rdg><rdg wit="#f1823">infamy that could make
1 </rdg><rdg wit="#f1831">infamy that could make </rdg><rdg wit="#fMSc56">infamy that could &lt;lb
1 n="c56-0092_main_5"/&gt;;make </rdg></app><app><rdg wit="#f1818">the </rdg><rdg wit="#f1823">the </rdg><rdg
1 wit="#f1831">the </rdg><rdg wit="#fMSc56">that </rdg></app><app type="invariant"><rdg wit="#f1818">murder
1 </rdg><rdg wit="#f1823">murder </rdg><rdg wit="#f1831">murder </rdg><rdg wit="#fMSc56">murder
1 </rdg></app><app><rdg wit="#f1818">memorable </rdg><rdg wit="#f1823">memorable </rdg><rdg
1 wit="#f1831">memorable </rdg><rdg wit="#fMSc56">&lt;del rend="strikethrough"&gt;more
1 terrible&lt;/del&gt;&lt;del&gt;&lt;add hand="#pbs" place="superlinear"&gt;memorable </rdg></app><app
1 type="invariant"><rdg wit="#f1818">in horror. Justine also was a girl of </rdg><rdg wit="#f1823">in horror.
1 Justine also was a girl of </rdg><rdg wit="#f1831">in horror. Justine also was a girl of </rdg><rdg
1 wit="#fMSc56">in horror.&lt;/add&gt; Justine &lt;lb n="c56-0092_main_6"/&gt;;also was a girl of
1 </rdg></app><app><rdg wit="#f1818">merit, </rdg><rdg wit="#f1823">merit, </rdg><rdg wit="#f1831">merit,
```

# Stylometry and Digital Frankenstein

Matthew Lavin  
University of Pittsburgh  
@mjlavin80

# Research Questions

1. How does *Frankenstein* change stylistically across different expressions/manifestations?
2. How can those changes be attributed and/or characterized?
  - In whose authorial voice is *Frankenstein*? Direct analysis of Percy and Mary
3. Do stylistic changes affect how *Frankenstein* reads in relation to cultural categories like genre, “modernness,” linguistic register generally, and scientific vocabulary? If so, how?

# Outline of Exploratory Measures

In notebooks, term counts/relative frequencies of:

- Mary's hand initial
- Mary's hand strikethrough
- Percy's hand suggested vs. adopted
- Mary's hand revised (sometimes Mary ver1, ver2, final, etc.)

Across our three print editions:

- Term counts/relative frequencies of each text
- Term frequencies weighted against frequency across all documents (TF-IDF)

	index	word	1818_1823
0	5814	shewed	7
1	2662	showed	7
2	352	daemon	6
3	7175	fulfil	6
4	1365	dæmon	6
5	6397	fulfill	6
6	5404	when	5
7	711	but	5
8	5591	my	5
9	1311	an	5
10	3570	shown	4
11	6767	shewn	4
12	3033	and	4
13	506	that	4
14	2604	may	4
15	6575	he	4

## Term Counts

Absolute Values of  
Term Count  
Differences  
across Editions,  
1818 to 1823 (left) and  
1823 to 1831 (right)

	index	word	1823_1831
0	5919	the	220
1	6391	of	169
2	6071	to	142
3	5591	my	103
4	3033	and	98
5	6034	a	93
6	3743	i	75
7	3284	in	69
8	654	me	66
9	3765	was	42
10	2413	their	41
11	1744	his	37
12	7267	as	35
13	5579	on	34
14	5813	her	33
15	4390	it	33

# Relative Term Frequencies

	index	word	1818_1823
0	4348	shewed	0.001555
1	3129	fulfill	0.001333
2	4149	daemon	0.001333
3	5356	showed	0.000920
4	6164	shewn	0.000889
5	2772	dæmon	0.000790
6	6925	fulfil	0.000789
7	2654	an	0.000685
8	7283	shown	0.000677
9	6918	phænomena	0.000667
10	6394	aye	0.000667
11	3507	when	0.000613
12	1390	but	0.000570
13	5255	may	0.000512
14	3844	avelânche	0.000508
15	1731	phenomena	0.000507

Absolute Value of  
Weighted Term  
Frequency Differences  
across Editions (tf-idf),  
1818 to 1823 (left) and  
1823 to 1831 (right)

	index	word	1823_1831
0	212	i	0.007377
1	1007	that	0.006288
2	2988	of	0.006038
3	4810	to	0.005899
4	6147	and	0.005206
5	3142	we	0.004865
6	7169	their	0.004240
7	5279	not	0.004156
8	3848	this	0.003893
9	4744	a	0.003732
10	1284	me	0.003429
11	6263	is	0.003332
12	4531	the	0.003270
13	1861	you	0.002936
14	3912	my	0.002796
15	5947	father	0.002652

# Collational Alignment

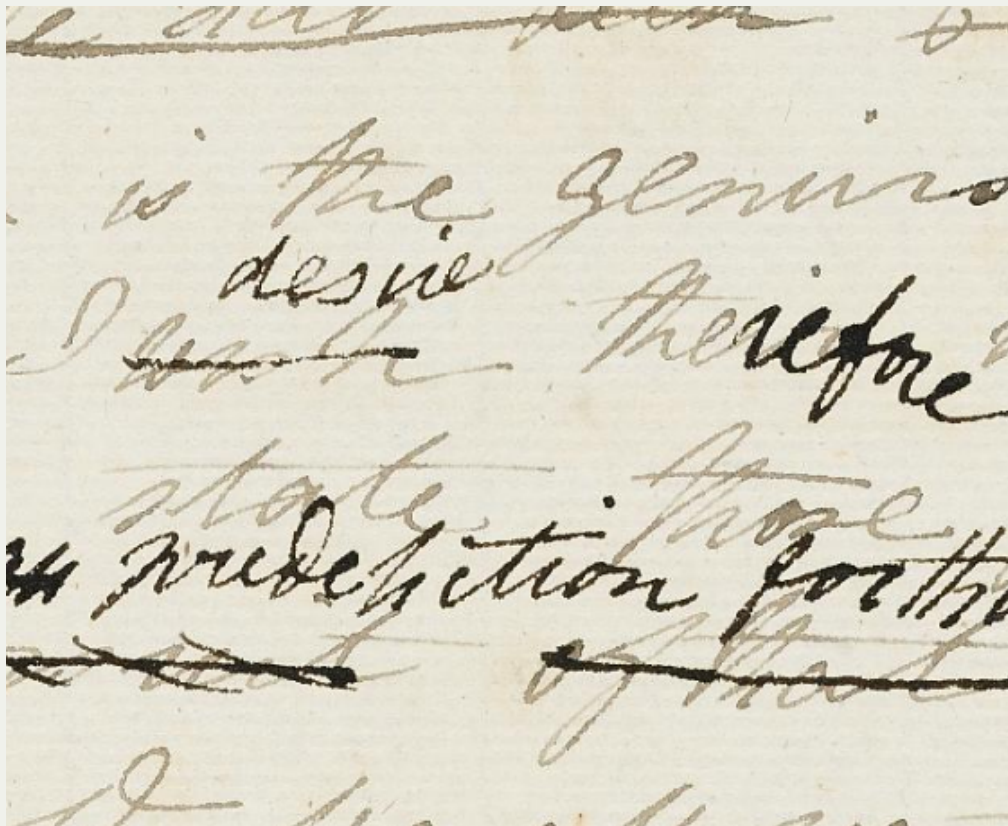
n	34609	one	n	34556	.
n	34657	them	n	34609	some
	34658	indeed	n	34657	them ,
n	34893	peacefully		34658	indeed ,
n	34899	alteration	n	34893	peaceably
n	34904	the	n	34899	alteration ,
n	35021	day - light,	n	35020	daylight,
n	35029	as	n	35238	degenerati o n-of
n	35240	degenerati o n-of	n	35480	fellowcreatures
n	35482	fellow - creatures	n	35499	advantages;
n	35501	acquisitions;	n	35528	few ?
n	35530	few .	n	35559	end u ed
n	35561	end ow ed	n	35607	theirs.
n	35609	their ' s.	n	35668	for
n	35670	forever		35669	ever
n	35678	or	n	35677	n or
n	35688	" Of	n	35687	"
n	35857	the		35688	Of
n	35896	wrapt	n	35895	wrapped
n	35929	" But	n	35928	"
n	36072	"SOME		35929	But
			n	36072	"Some

## Types of Changes:

- Spelling normalizations
- Punctuation
- Word insertions, substitutions, deletions
- Word to phrase or phrase to word
- Reordering



# Shelley-Godwin Notebooks



## Chapt. 2

Those events which materially influence our future destinies are often caused by slight or trivial occurrences.

*derive thier origin from a statement of the* Strange as the simple fact

may appear my fate had been *Chemist* Natural philosophy has is the genius that has

regulated my fate I **wish** *therefore* in this account of my early years to state those facts which

*first acquired predilection for that science.* led to my love pursuit of that study. When I was eleven years old we all went on a party

*the baths near Thonon.* of pleasure to Thonon and were confined there ~~to~~ obliged by the rain and ~~t~~The inclemency of the weather obliged us to remain a day confined to the inn. In this house I chanced

Image courtesy of [shelleygodwinarchive.org](http://shelleygodwinarchive.org)

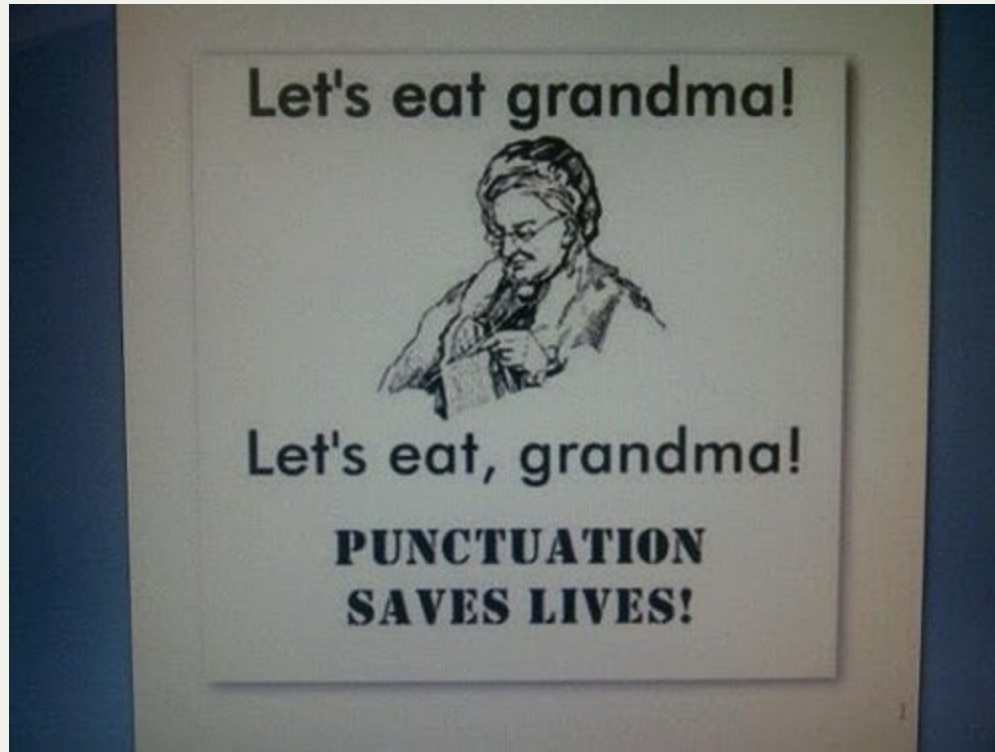


Image courtesy of [dailywritingtips.com](http://dailywritingtips.com)

Punctuation Matters ... but not for all measures

# Dynamics of DH Collaboration

The workflows and analytical paradigms of “machine learning DH” and “scholarly editing DH” are not factory fitted to one another, but they can be adapted to work in tandem. The gains are more valuable than the cost of the retrofit.

# Dynamics of DH Collaboration

How can a single, carefully curated edition or set of editions be worked into a “macroanalysis” model where many uncorrected, dirty OCR texts are being compared to one another?

What kinds of questions can we ask with hand-corrected editions that we cannot ask with HTRC corpora?

# Open Data and Reproducibility

I have argued elsewhere that openness invites open discussion and collaboration. It doesn't guarantee that these things will happen, but closed data practices all but guarantee that these practices will be difficult or impossible.

# Next Steps: Questions/Methods

How can we characterize changes by trends established in analysis of each person's hand?

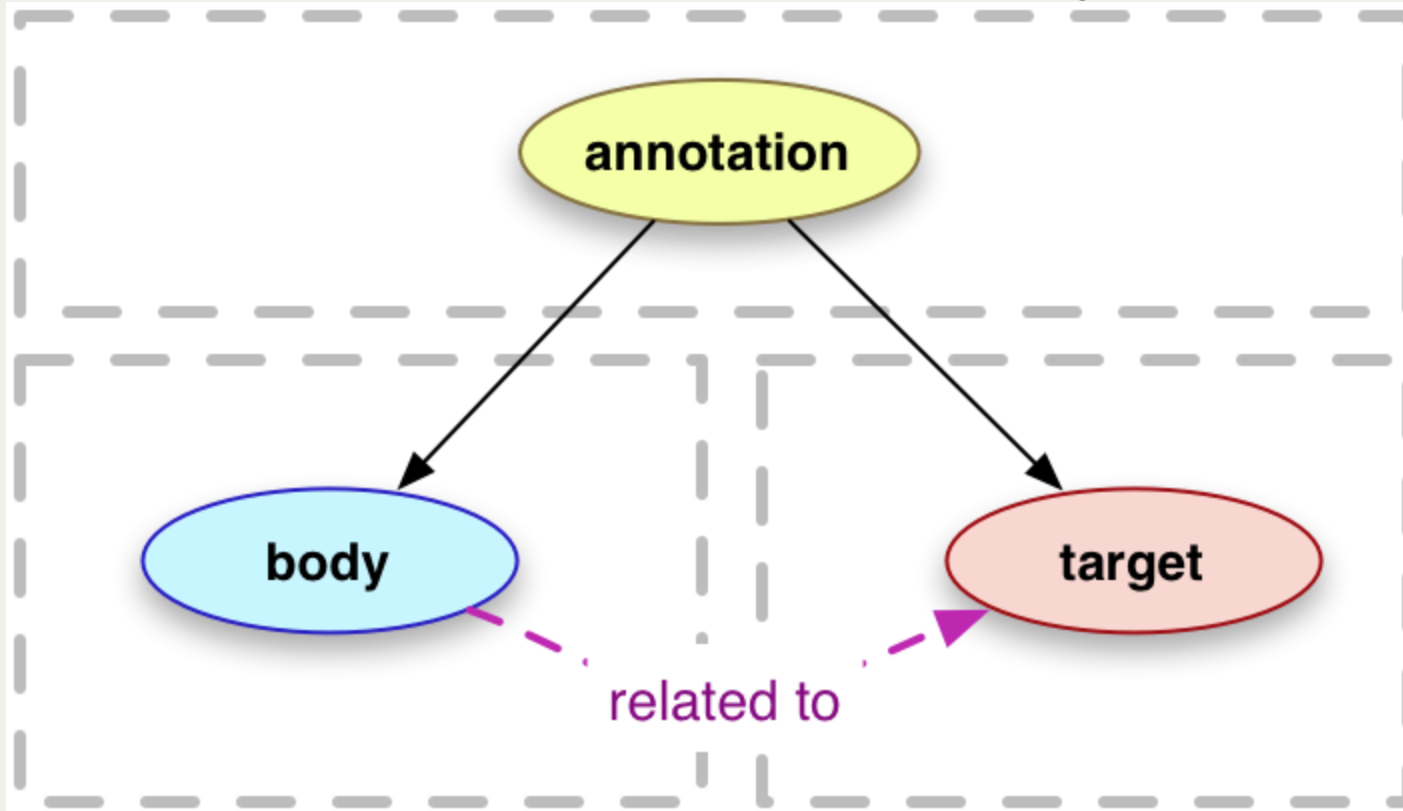
How can we think about changes as moving closer or further away from a genre baseline?

How do index quantifications like "how modern" or "how scientific" each version of the text is? How do we account for "modern" and "scientific" as rapidly changing ideas?

## Speaker notes

As Scott Weingart's work has shown, the computational determination of "who wrote what" has become less central to DH inquiry. I believe strongly, however, that these approaches will have a second life as we turn increasingly toward the use of authorship attribution techniques to study authorship as an historical and social construction.

Jon Klancher  
@jklancher



Source: [Web Annotation Data Model](#)  
(w3c Recommendation of 23 Feb 2017)



# Annotated Print Editions of Frankenstein

1993 Leonard Wolf, ed., *The Essential Frankenstein: The Definitive, Annotated Edition of Mary Shelley's Classic Novel* (New York: Plume). (1st edition as *The Annotated Frankenstein*, 1977)

2012 Susan J. Wolfson and Ronald L. Levaio, *The Annotated Frankenstein* (Cambridge, MA: Belknap Press of Harvard University).

2017 Leslie S. Klinger, ed., *The New Annotated Frankenstein* (New York: Liveright/Norton).

2017 David G. Guston, ed., *Frankenstein: Annotated for Scientists, Engineers, and Creators of All Kinds* (Cambridge, MA: MIT Press).

My father had a sister, whom he tenderly loved, and who had married early in life an Italian gentleman. Soon after her marriage, she had accompanied her husband into [his]<sup>11</sup> native country, and for some years my father had very little communication with her. About the time I mentioned she died; and a few months afterwards he received a letter from her husband, acquainting him with his intention of marrying an Italian lady, and requesting my father to take charge of the infant Elizabeth, the only child of his deceased sister.<sup>12</sup> “It is my wish,” he said, “that you should consider her as your own daughter, and educate her thus. Her mother’s fortune is secured to her, the documents of which I will commit to your keeping. Reflect upon this proposition; and decide whether you would prefer educating your niece yourself to her being brought up by a stepmother.”<sup>13</sup>

My father did not hesitate, and immediately went to Italy, that he might accompany the little Elizabeth to her future home.<sup>14</sup> I

11 The misprint of “her” in 1818 is corrected to “his” in 1823 (Ch. I, 44).

12 Percy’s sister and mother had the name “Elizabeth.” Mary layers several family names into her narrative. The father’s disposal of a child, here a seemingly incidental plot device, is related to the multiple versions of parental care and carelessness in the novel. It is striking to see a father cast off his only child within a few months of her losing her mother. Would the gentleman have discarded a son?

13 A “stepmother” invokes a fairy-tale lore (and its social basis) of coolness to a stepchild. Mary’s stepmother (the second Mrs. Godwin) was less than loving, and her father would send his daughter, at age thirteen, to live in Scotland.

14 In 1831, Mary Shelley revised Elizabeth’s relation from a close blood tie (first cousin) to an adopted orphan—sharpening the contrast of the beautiful orphan’s good fortune in finding a loving family and the misfortunes of the abject ugly Creature (even with a connection to the Frankensteins).

Susan J. Wolfson and Ronald L. Levao, *The Annotated Frankenstein* (Cambridge, MA: Belknap Press, 2012)

was the eldest, and the destined successor to all his labours and utility. No creature could have more tender parents than mine. My improvement and health were their constant care, especially as I remained for several years their only child. But before I continue my narrative, I must record an incident which took place when I was four years of age.

My father had a sister, whom he tenderly loved, and who had married early in life an Italian gentleman. Soon after her marriage, she had accompanied her husband into her<sup>10</sup> native country, and for some years my father had very little communication with her. About the time I mentioned she died; and a few months afterwards he received a letter from her husband, acquainting him with his intention of marrying an Italian lady, and requesting my father to take charge of the infant Elizabeth,<sup>11</sup> the only child of his deceased sister. "It is my wish," he said, "that you should consider her as your own daughter, and educate her thus. Her mother's fortune is secured to her, the documents of which I will commit to your keeping. Reflect upon this proposition; and decide whether you would prefer educating your niece yourself to her being brought up by a stepmother."<sup>12</sup>

My father did not hesitate, and immediately went to Italy, that he might accompany the little Elizabeth to her future home. I have often heard my mother say, that she was at that time the most beautiful child she had ever seen, and shewed signs even then of a gentle and affectionate disposition. These indications, and a desire to bind as closely as possible the ties of domestic love, determined my mother to consider Elizabeth as my future wife; a design which she never found reason to repent.

From this time Elizabeth Lavenza became my playfellow, and, as we grew older, my friend. She was docile and good tempered, yet gay and playful as a summer insect. Although

There was a considerable difference between the ages of my parents, but this circumstance seemed to unite them only closer in bonds of devoted affection. There was a sense of justice in my father's upright mind which rendered it necessary that he should approve highly to love strongly. Perhaps during former years he had suffered from the late-discovered unworthiness of one beloved and so was disposed to set a greater value on tried worth. There was a show of gratitude and worship in his attachment to my mother, differing wholly from the doting fondness of age, for it was inspired by reverence for her virtues and a desire to be the means of, in some degree, recompensing her for the sorrows she had endured, but which gave inexpressible grace to his behaviour to her. Everything was made to yield to her wishes and her convenience. He strove to shelter her, as a fair exotic is sheltered by the gardener, from every rougher wind and to surround her with all that could tend to excite pleasurable emotion in her soft and benevolent mind. Her health, and even the tranquillity of her hitherto constant spirit, had been shaken by what she had gone through. During the two years that had elapsed previous to their marriage my father had gradually relinquished all his public functions; and immediately after their union they sought the pleasant climate of Italy, and the change of scene and interest attendant on a tour through that land of wonders, as a restorative for her weakened frame.

them in their rambles. I remained for several years their only child. Much as they were attached to each other, they seemed to draw inexhaustible stores of

This annotation is the verbatim 1831 altered text.

2017 Leslie S. Klinger, ed., *The New Annotated Frankenstein* (New York: Liveright/Norton).

more grace than she did to constraint and caprice. Her imagination was luxuriant, yet her capability of application was great. Her person was the image of her mind; her hazel eyes,

## Wolfson annotation:

Published in 1791, in the wake of the French Revolution (Volney was part of the Revolutionary government), *Les Ruines; ou Meditation sur les revolutions des empires* appeared in English as *Ruins, or Meditations on the Revolutions of Empires*, in 1792.

## Klinger annotation:

More properly, *The Ruins, Or, Meditation on the Revolutions of Empires; and the Laws of Nature*, by Constantin-François Chasseboeuf, who took the name Volney, published in 1791 in French. It was translated in 1802 into English. The book is described by Frankenstein scholar Pamela Clemit as a “powerful Enlightenment critique of ancient and modern governments as tyrannical and supported by religious fraud” (“Frankenstein, Matilda, and the Legacies of Godwin and Wollstonecraft,” in *The Cambridge Companion to Mary Shelley*, ed. Esther Schor [Cambridge: Cambridge University Press, 2003] 35.)

*In light of the date of translation*, the book in question must have been the French edition, and Safie and the creature learned French....

## Our Frankenstein Variorum annotation:

Of the books the Creature hears read aloud in the forest, Volney's *The Ruins; or, A Survey of the Revolutions of Empires* (1792) was the most closely associated with Europe's radical Enlightenment. (It was first published in French as *Les Ruines: ou Meditation sur les revolutions des empires* in 1791.) The Creature learns an illuminating critique of imperialism and exploitation from Volney, even as he also absorbs some of the Enlightenment's own prejudices ("slothful Asiatics"). The effect on the Creature is to give him a sense of the social or structural and not only a personal framework for understanding virtue and suffering. On Volney's role in the novel, see also Ian Balfour, "Allegories of Origins: Frankenstein after the Enlightenment," *SEL: Studies in English Literature 1500-1900* 56.4 (2016): 777-98.

hypothes.is: all tags so far...

poetry 15 family 9 geography 7  
religion 6 natural philosophy 5  
science 5 John Milton 4  
domestic affections 4 sympathy 4  
Romanticism 3 alchemy 3 arctic 3  
empire 3 imagination 3 mother 3  
north pole 3 Humphry Davy 2  
Lyrical Ballads 2 Russia 2  
Samuel Taylor Coleridge 2  
Shakespeare 2 brother 2 education 2  
friendship 2 guilt 2 incest 2  
murder 2 novel 2  
political philosophy 2 romance 2

religion 6 arctic 2 poetry 2  
Adam 1 Albertus Magnus 1 Hell 1  
John Milton 1 Milton 1 Paracelsus 1  
Russia 1 St. Petersburg 1  
alchemy 1 creator 1 geography 1  
imagination 1 north pole 1  
paradise 1 science 1

geography 7 Russia 2  
Geneva Summer 1  
Northwest Passage 1 St. Petersburg 1  
Switzerland 1 Tambora 1 arctic 1  
climate 1 education 1 england 1  
enlightenment 1 political philosophy 1  
religion 1 voyages 1

using the [hypothes.is](https://hypothes.is) tool for digital annotation with tags

# Annotations that Tunnel through the Texts

(not only pointing outside)

travel/expedition: Walton

domestic affection  
(Walton - Margaret Seville)

law / judicial system  
(Justine)

travel/expedition: Victor



travel/expedition: Clerval

travel/expedition:  
Creature

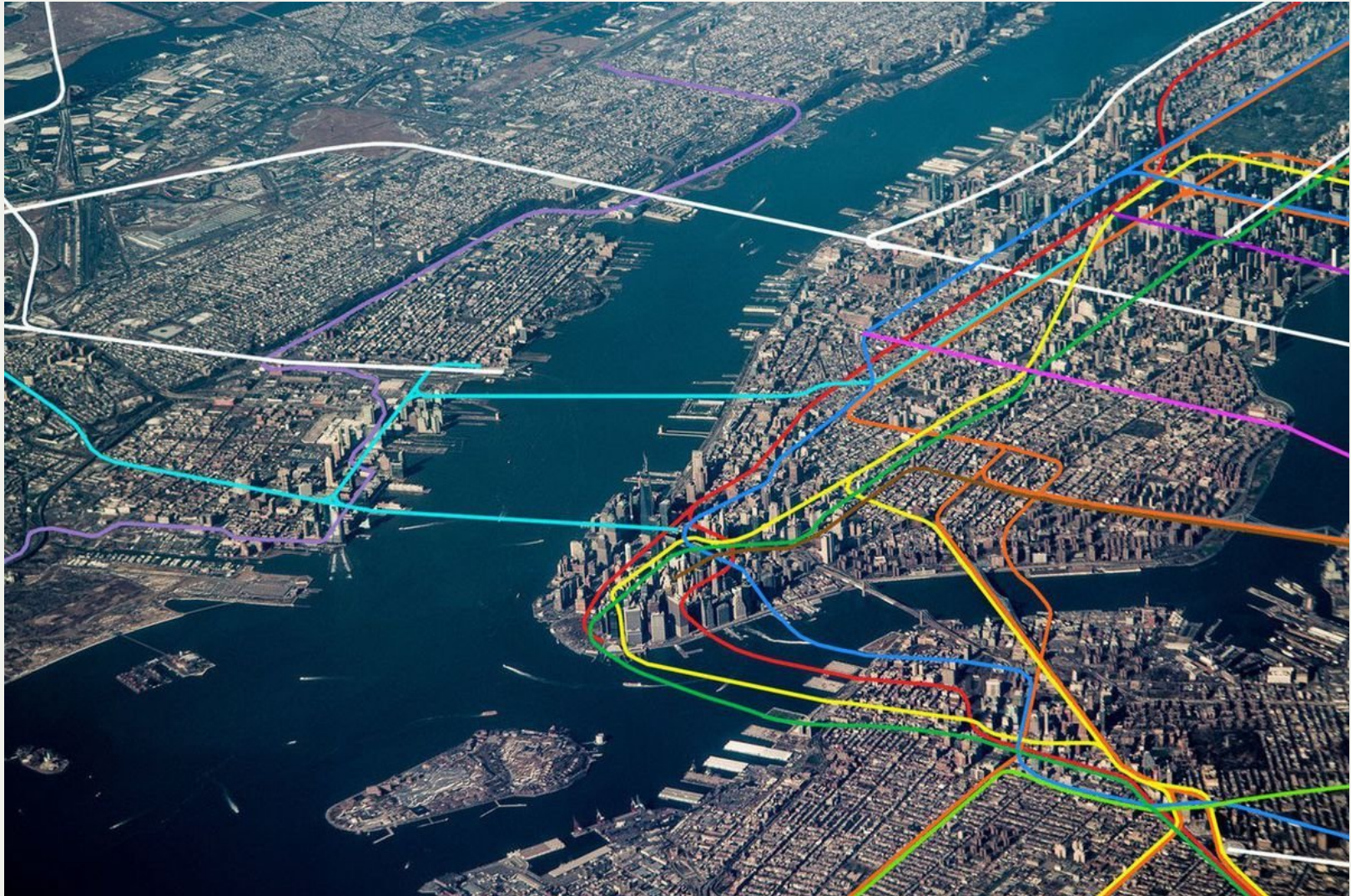
domestic affection  
(Frankenstein family)

domestic affection  
(DeLaceys and Safie)

law / judicial system  
Victor/Kirwin

law / judicial system  
Felix DeLacey

# Annotations in the Variorum Interface



Source: 6Sqft: "See New York City's subway lines superimposed over an aerial photo of the city" (2015)

## Speaker notes

### Annotations and Intertextuality

"tunneling" through the texts

distinct from external pointers to context

affected by collation "bridges"

goal: available in each reading view

id markers signal textual locations:

book | chapter | paragraph

collation alignment shows corresponding locations in other versions

Portable hypotheses annotations: "pinnable" by text string and id position markers



# Frankenstein's invitation/challenge:

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <!--METADATA -->
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <!--DOCUMENT-->
    </body>
  </text>
</TEI>
```

## *The work continues...*

- Collation
- Annotation
- Stylometry
- Visualization and Variorum Interface

- Build Digitally:
  - Experiment with human and machine reading
- Build a Strong, Sustainable Bridge:
  - update Romantic Circles edition
  - interlinks to Shelley-Godwin Archive
  - Notebooks: point to ms pages
  - Morgan Library "Thomas copy"
- Centralize the Critical Apparatus
  - a tool for scholars
  - a metanarrative?
  - a remixing of the **reading process** for all who care about *Frankenstein*
  - make all the texts available to all the readers