University of Basel

Infrastructures and tools for research data management

Coffee lecture series

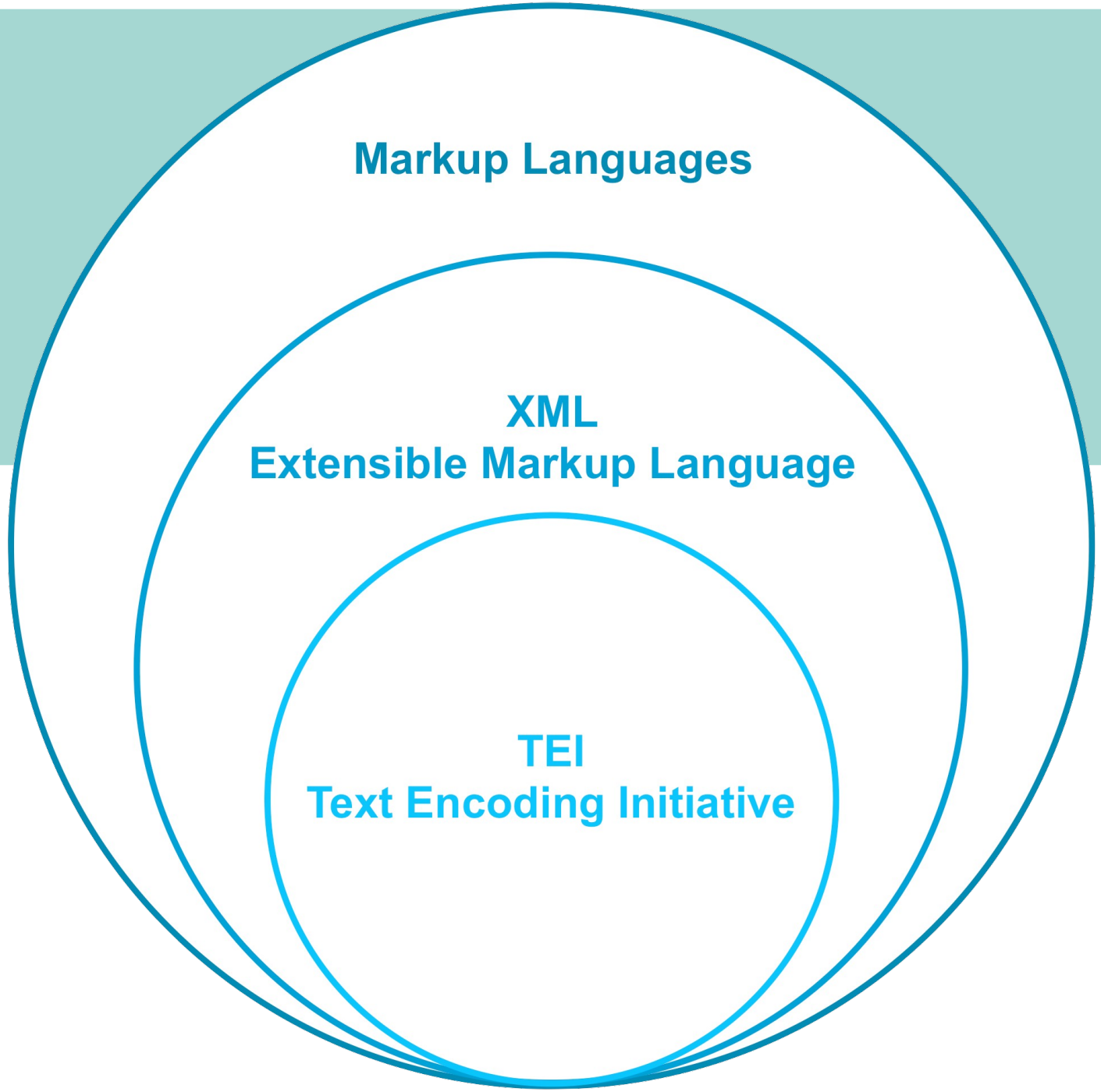
Research Data Management Network, fall semester 2022

University of Basel

Text Encoding Initiative (TEI)

An international standard to add intelligent markup to digital resources

University of Basel



16/18 Roman

Sample Manuscripts throughout the Ages

12/14 Helv

by Jeff Beck

10/12 Roman

The funny thing about sample manuscripts is that they never really say anything interesting. From the Byzantine period through the Post-Modern Age, the text usually just repeats the same thing over and over again. Over and over again.

The funny thing about sample manuscripts is that they never really say anything interesting. From the Byzantine period through the Post-Modern Age, the text usually just repeats the same thing over and over again. Over and over again.

The funny thing about sample manuscripts is that they never really say anything interesting. From the Byzantine period through the Post-Modern Age, the text usually just repeats the same thing over and over again. Over and over again.

The funny thing about sample manuscripts is that they never really say anything interesting. From the Byzantine period through the Post-Modern Age, the text usually just repeats the same thing over and over again. Over and over again.

https://www.ncbi.nlm.nih.gov/staff/beck/xml/markup/I-A.html

University of Basel

Make **explicit** (to a machine)
what is implicit (to a person)

Representing parts and features
of a digital resource
in a **formalised** way
→ machine processable

University of Basel

Procedural MARKUP (HTML)

```
<bold>Jane Austen</bold> wrote <italic>Price and Prejudice</italic>
```

Descriptive MARKUP (TEI)

```
<name>Jane Austen</name> wrote <title>Price and Prejudice</title>

name {font-weight: bold;}
title {font-style: italic;}
```

University of Basel

Descriptive markup allows for the **separation of form and content**

- multiple forms possible
- easier reuse (in different contexts)
- easier preservation

- Print
  - Body of text
  - Indices
  - ...
- Digital
  - Linear text in HTML
  - Linear text in PDF
  - Decomposed text
  - Indices
  - ...

From same XML source

Ciula and Pierazzo, "Encoding Texts", in *Medieval/Modern Manuscript Studies in the Digital Age*, London, 2014

University of Basel

The TEI was established in 1987 to develop, maintain, and promulgate hardware- and software-independent methods for encoding humanities data in electronic form.

https://tei-c.org/about/history

University of Basel

Since 1999, the TEI is led by a **consortium** that maintains and develops **guidelines** for text encoding.

P5: Guidelines for Electronic Text Encoding and Interchange

Version 4.5.0. Last updated on 25th October 2022, revision 3e98e619e

[English] [Deutsch] [Español] [Italiano] [Français] [日本語] [한국어] [中文]

Front Matter

Title

i. Releases of the TEI Guidelines

ii. Dedication

iii. Preface and Acknowledgments

⊞ iv. About These Guidelines

⊞ v. A Gentle Introduction to XML

⊞ vi. Languages and Character Sets

Back Matter

⊞ Appendix A Model Classes

⊞ Appendix B Attribute Classes

⊞ Appendix C Elements

⊞ Appendix D Attributes

⊞ Appendix E Datatypes and Other Macros

⊞ Appendix F Bibliography

⊞ Appendix G Deprecations

⊞ Appendix H Prefatory Notes

Appendix I Colophon

Text Body

⊞ 1 The TEI Infrastructure

⊞ 2 The TEI Header

⊞ 3 Elements Available in All TEI Documents

⊞ 4 Default Text Structure

⊞ 5 Characters, Glyphs, and Writing Modes

⊞ 6 Verse

⊞ 7 Performance Texts

⊞ 8 Transcriptions of Speech

⊞ 9 Dictionaries

⊞ 10 Manuscript Description

⊞ 11 Representation of Primary Sources

⊞ 12 Critical Apparatus

⊞ 13 Names, Dates, People, and Places

⊞ 14 Tables, Formulæ, Graphics, and Notated Music

⊞ 15 Language Corpora

⊞ 16 Linking, Segmentation, and Alignment

⊞ 17 Simple Analytic Mechanisms

⊞ 18 Feature Structures

⊞ 19 Graphs, Networks, and Trees

⊞ 20 Non-hierarchical Structures

⊞ 21 Certainty, Precision, and Responsibility

⊞ 22 Documentation Elements

⊞ 23 Using the TEI

TEI sourcecode

- Getting and Using the TEI Sources.
- TEI GitHub Repository
- Bug Reports, Feature Requests, etc.

University of Basel

Each project should define its own **schema**

The schema represents in a structured and formalised way your **understanding** and **interpretation** of the object of study (scientific choices).

The TEI **modular** system (modules and classes) helps in customizing the schema.

JOURNAL OF THE TEXT ENCODING INITIATIVE

The *Journal of the Text Encoding Initiative* is the official journal of the Text Encoding Initiative Consortium. It publishes the proceedings of the annual *TEI Conference and Members' Meeting* and special thematic issues: state-of-the-art reports on electronic textual editing, current trends in TEI encoding, and new use cases for TEI. It furthermore provides a forum for articles on the discussion of the interface between the TEI and other communities, and more generally of the role of technological standards in the digital humanities, including digital scholarly editing, linguistic analysis, corpora creation, and newer areas such as mass digitization, semantic web research, and editing within virtual worlds.

Search

INDEX

- Authors
- Keywords

OPEN ISSUES

- Issue 14 | 2021
  Selected Papers from the 2019 TEI Conference

FULL TEXT ISSUES

- Issue 11 | July 2019 - June 2020
  Selected Papers from the 2016 TEI Conference
- Issue 12 | July 2019 - May 2020
  Selected Papers from the 2017 TEI Conference
- Issue 13 | May 2020 - November 2022
  Selected Papers from the 2018 TEI Conference
- Issue 10 | December 2016 - July 2019
  Selected Papers from the 2015 TEI Conference
- Issue 9 | September

CURRENT OPEN ISSUE

ISSUE 14 | 2021 (OPEN ISSUE)

Selected Papers from the 2019 TEI Conference

Edited by **Georg Vogeler**

3 LATEST TEXTS

Michał Kozak, Alejandro Rodríguez, Alejandro Benito-Santos, Roberto Therón, Michelle Doran, Amelie Dorn, Jennifer Edmond, Cezary Mazurek and Eveline Wandl-Vogt

Analyzing and Visualizing Uncertain Knowledge: The Use of TEI Annotations in the PROVIDEDH Open Science Platform [Full text]

13 September 2022

Tanja Wissik

Encoding Interruptions in Parliamentary Data: From Applause to Interjections and Laughter [Full text]

30 June 2022

University of Basel

TEI in action

examples

University of Basel

Collect and describe historical sources

Example of digital catalogues of manuscripts:

e-codices
Manuscriptorium
MANUS

Collezioni

Collezioni svizzere

Luogo, Biblioteca / Collezione	Documenti
Tutte le biblioteche e collezioni	**1290**
Aarau, Aargauer Kantonsbibliothek	15
Aarau, Staatsarchiv Aargau	6
Basel, Universitätsbibliothek	37

Novità

Nuovi manoscritti:

19.03.2015

Newsletters precedenti:

- Numero 19, marzo 2015
- Numero 18, gennaio 2015
- Numero 17, dicembre 2014
- Numero 16, dicembre 2014

Fac-similé: Une page | Deux pages | Miniature

Détails | Annotations | Bibliographie additionnelle

Pays de conservation:	Suisse
Lieu:	Cologny
Bibliothèque / Collection:	Fondation Martin Bodmer
Cote:	Cod. Bodmer 130
Résumé du manuscrit:	Pétrarque, Triumphi
Caractéristiques:	Parchemin · I + 185 + III ff. · 20.0 x 12.0 cm · Italie (Padoue) · vers 1500
Langue:	Italien, Latin
Titre du manuscrit:	Réalisé dans les premières années du XVIème siècle, alors que l'imprimerie a déjà affirmé son savoir-faire, le CB 130 témoigne d'une maîtrise souveraine de la calligraphie et de l'art pictural. Copié par Bartolomeo Sanvito, qui a exécuté quatre autres manuscrits du *Canzoniere* et des *Triumphi* de Pétrarque, il offre une écriture sobre et équilibrée, enrichie d'enluminures raffinées. Trois peintures sur feuille de parchemin marquent le début des parties du livre.
Description standard:	Allegretti Paola, Catalogo dei codici italiani, Cod. Bodmer 130, in "Corona Nova. Bulletin de la Bibliotheca Bodmeriana", II (2003), pp. 66-76. Voir la description standard
DOI (Digital Object Identifier):	10.5076/e-codices-cb-0130
Lien permanent:	http://www.e-codices.unifr.ch/fr/list/one/fmb/cb-0130
IIIF Manifest URL:	http://www.e-codices.unifr.ch/metadata/iiif/fmb-cb-0130/manifest.json
Comment citer:	Cologny, Fondation Martin Bodmer, Cod. Bodmer 130: Pétrarque, Triumphi (http://www.e-codices.unifr.ch/fr/list/one/fmb/cb-0130).
En ligne depuis:	25.03.2009
Droits:	Images: CC BY NC (Concernant tous les autres droits, voir chaque description de manuscrits et nos conditions d 'utilisation)

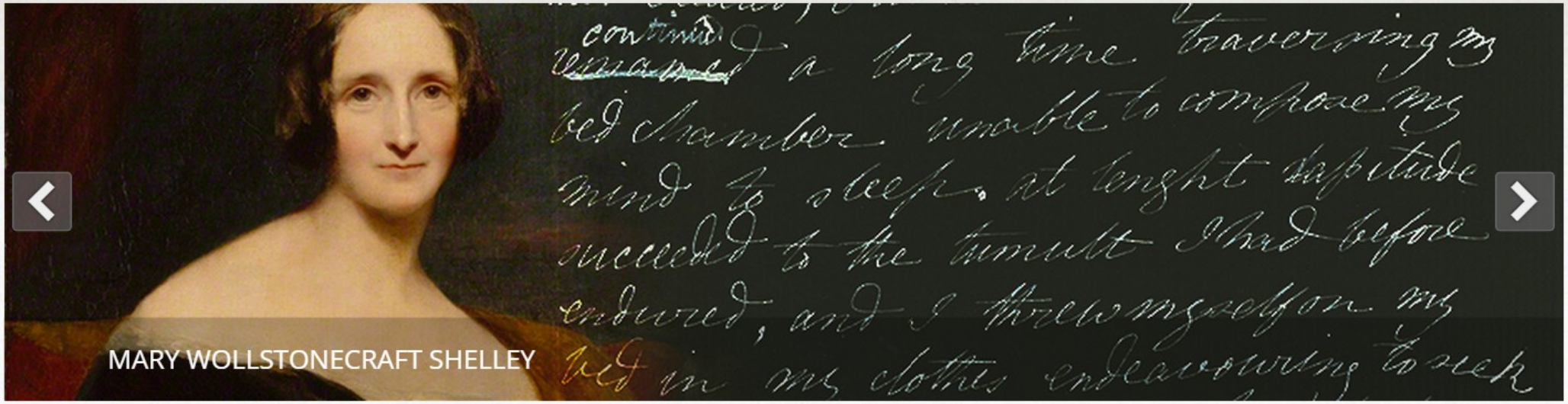
University of Basel

Scholarly editing (critical, genetic, diplomatic, etc.)

Catalogues of scholarly editions:

Patrick Sahle, A catalogue Digital Scholarly Editions
Greta Franzini, Catalogue Digital Editions

THE Shelley-Godwin ARCHIVE BETA



About the Archive

The Shelley-Godwin Archive will provide the digitized manuscripts of Percy Bysshe Shelley, Mary Wollstonecraft Shelley, William Godwin, and Mary Wollstonecraft, bringing together online for the first time ever the widely dispersed handwritten legacy of this uniquely gifted family of writers. The result of a partnership between the New York Public Library and the Maryland

S-GA in the News

October 30, 2013

The New York Times Arts Beat

'Frankenstein' Manuscript Comes Alive in Online Shelley Archive

by Jennifer Schuessler

Draft Notebook A

Author(s) : Mary Shelley
Date Written : [August or September]-[?December] 1816
Title/Literary Work : Frankenstein

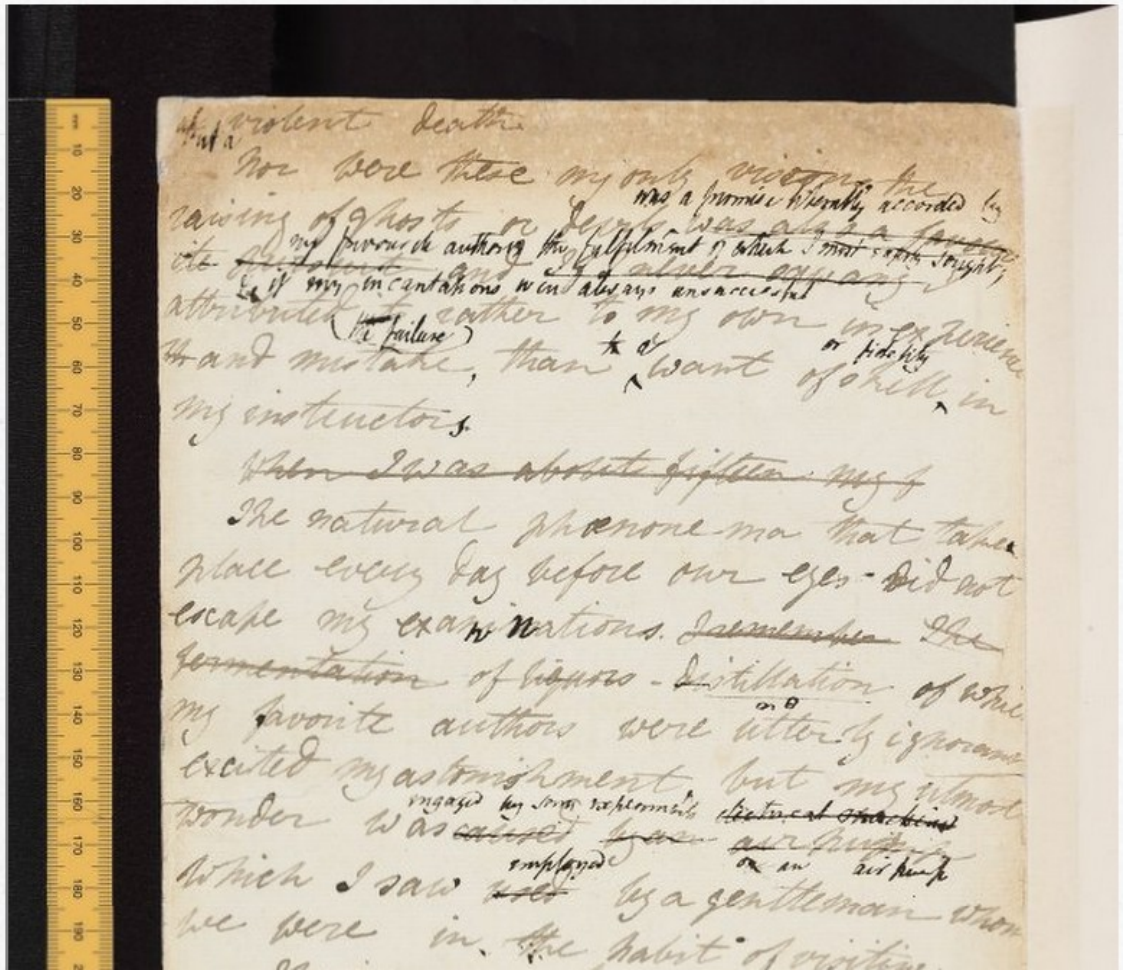
View : Frankenstein, Draft Notebook A
State : draft
Institution : Mary Shelley

Hand(s) : Mary Shelley, Percy Shelley
Shelfmark : MS. Abinger c. 56
Folio : 2v

Transcription Status:
Metadata Status:

Search

LIMIT VIEW:
All | Mary Shelley | Percy Shelley



b violent death.

ut a

Nor were these my only visions, the

was a promise liberally accorded by

raising of ghosts or devils ~~was also a favour~~

^

my favourite authors; the fulfilment of which I most eagerly sought;

~~ite pursuit and If I never saw any~~

& if my incantations were always unsuccessful

attributed ~~it~~ rather to my own inexperience

the failure

to a *or fidelity*

~~th~~ and mistake, than want of skill in

^ ^

my instructor *s* .

~~When I was about fifteen my f~~

The natural phænonema that takes place every day before our eyes did not escape my exa ~~n~~ *m* i ~~m~~ *n* ations. ~~I remember~~ ~~The fermentation of liquors~~ – ~~di~~ stillation of which

2v

Draft Notebook A

Author(s) : Mary Shelley
Date Written : [August or September]-[?December] 1816
Title/Literary Work : Frankenstein

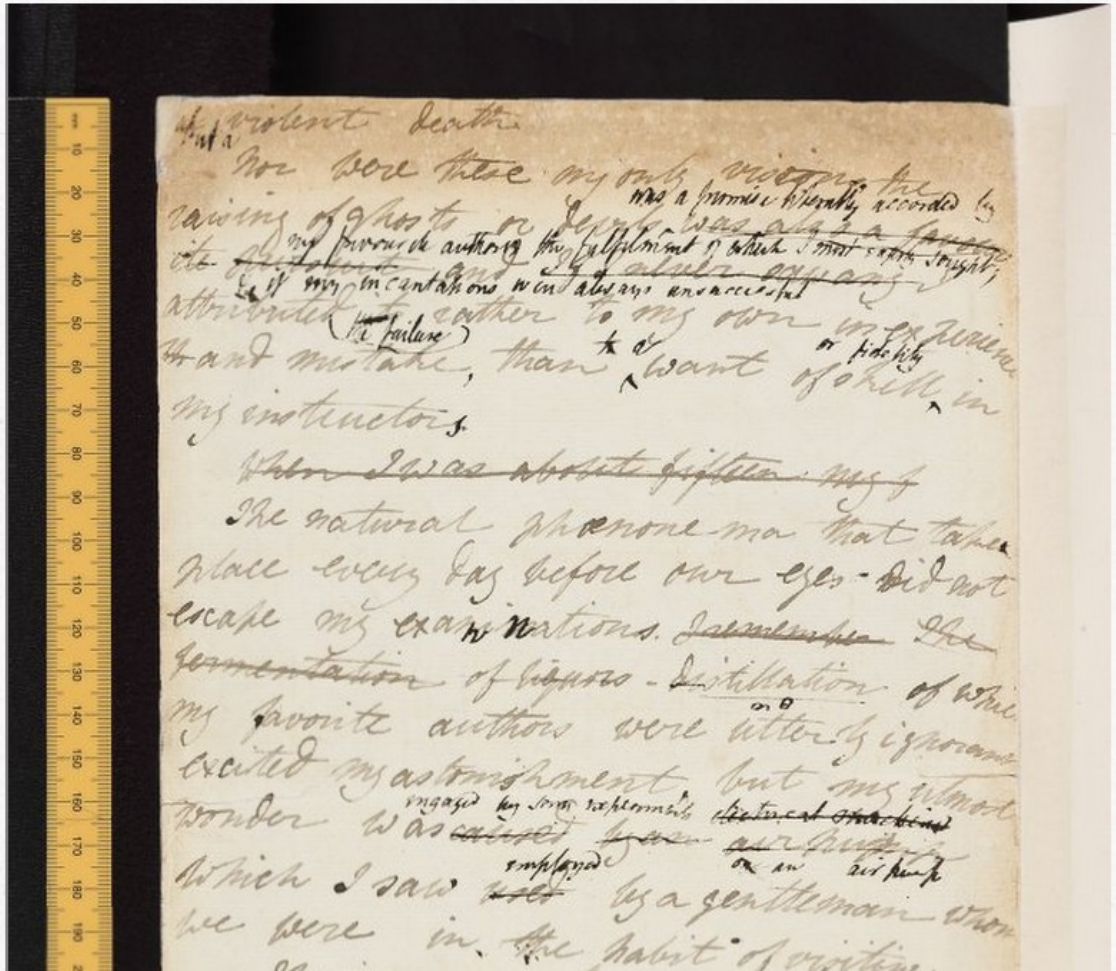
View : Frankenstein, Draft Notebook A
State : draft
Institution : Mary Shelley

Hand(s) : Mary Shelley, Percy Shelley
Shelfmark : MS. Abinger c. 56
Folio : 2v

Transcription Status:
Metadata Status:

Search

LIMIT VIEW:
All | Mary Shelley | Percy Shelley



b violent death.
ut a

Nor were these my only visions, the

was a promise liberally accorded by

raising of ghosts or devils ~~was also a favour~~

^

my favourite authors; the fulfilment of which I most eagerly sought;

~~ite pursuit and If I never saw any~~

& if my incantations were always unsuccessful

attributed ~~it~~ rather to my own inexperience

the failure

to a *or fidelity*

~~th~~ and mistake, than want of skill in

^ ^

my instructor *s* .

~~When I was about fifteen my f~~

The natural phænonema that takes
place every day before our eyes did not
escape my exa ~~n~~ *m* i ~~m~~ *n* ations. ~~I remember~~ ~~The fermentation of liquors~~ – ~~di~~ stillation of which

2v

Draft Notebook A

Author(s) : Mary Shelley
Date Written : [August or September]-[?December] 1816
Title/Literary Work : Frankenstein

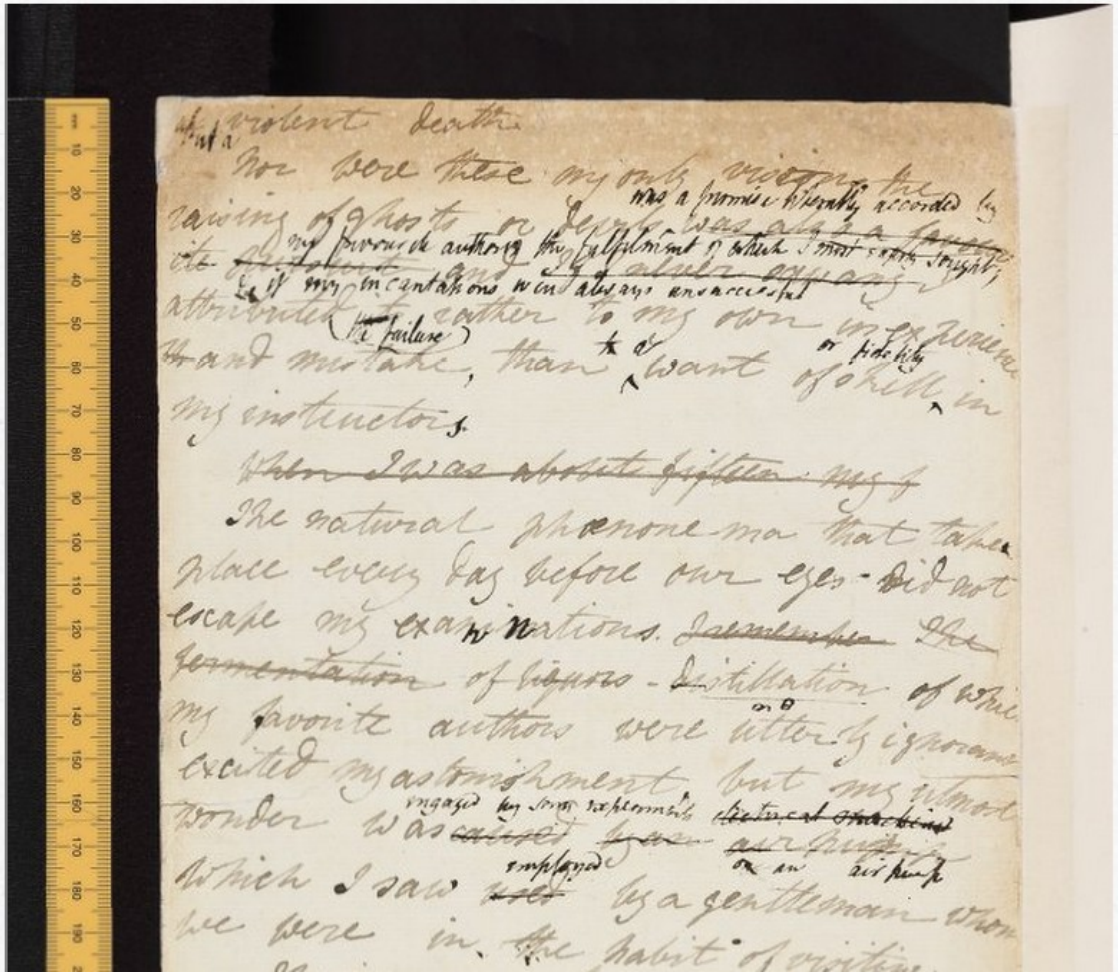
View : Frankenstein, Draft Notebook A
State : draft
Institution : Mary Shelley

Hand(s) : Mary Shelley, Percy Shelley
Shelfmark : MS. Abinger c. 56
Folio : 2v

Transcription Status:
Metadata Status:

Search

LIMIT VIEW:
All | Mary Shelley | Percy Shelley



b violent death.
ut a

Nor were these my only visions, the

was a promise liberally accorded by

raising of ghosts or devils ~~was also a favour~~

^

my favourite authors; the fulfilment of which I most eagerly sought;
~~ite pursuit and If I never saw any~~

& if my incantations were always unsuccessful

attributed ~~it~~ rather to my own inexperience

the failure

to a *or fidelity*

~~th~~ and mistake, than want of skill in

^ ^

my instructor *s* .

~~When I was about fifteen my f~~

The natural phænonema that takes
place every day before our eyes did not
escape my exa ~~n~~ *m* i ~~m~~ *n* ations. ~~I remember~~ ~~The~~
~~fermentation of liquors~~ – ~~di~~ stillation of which

Sammlung Schweizerischer Rechtsquellen online

Dokumente

Kanton		Abteilung
	ZH	I. Abteilung: Die Rechtsquellen des Kantons Zürich 897
	BE	II. Abteilung: Die Rechtsquellen des Kantons Bern
	LU	III. Abteilung: Die Rechtsquellen des Kantons Luzern
	UR	IV. Abteilung: Die Rechtsquellen des Kantons Uri
	SZ	V. Abteilung: Die Rechtsquellen des Kantons Schwyz
	OW/NW	VI. Abteilung: Die Rechtsquellen des Kantons Unterwalden
	GL	VII. Abteilung: Die Rechtsquellen des Kantons Glarus
	ZG	VIII. Abteilung: Die Rechtsquellen des Kantons Zug
	FR	IX^e partie : Les sources du droit du canton de Fribourg IX. Abteilung: Die Rechtsquellen des Kantons Freiburg 2713
	SO	X. Abteilung: Die Rechtsquellen des Kantons Solothurn
	BS/BL	XI. Abteilung: Die Rechtsquellen der Kantone Basel

Suche

Volltext

Wonach suchen?

Textgattung:

- [x] durchsuche Bearbeitungstext
  - [x] Titel
  - [x] Signatur
  - [x] Regest
  - [x] Kommentar
  - [x] Anmerkungen
  - [x] Siegel
  - [x] Literatur
- [x] durchsuche Editionstext

Felder und Facetten:

Zeitraum:

1050 –
1888

Sprache:

University of Basel

Linguistic annotation

For example:
Word-level annotation
Dictionaries
Spoken/multimodal corpora

TEITOK - a Tokenized TEI environment

TEITOK is a web-based platform for viewing, creating, and editing corpora with both rich textual mark-up and linguistic annotation, initially developed at the Centro de Linguística da Universidade de Lisboa, later at CELGA-ILTEC, and currently maintained at the ÚFAL institute of Charles University, Prague.

The system has a modular design with numerous modules making serving a wide range of different corpus types. Below are some examples of some of those, and the type of corpora TEITOK can deal with. More modules are added frequently, and it is possible to add custom modules as well.

The source is maintained at GitLab and some conversion tools are maintained on GitHub.

COPLE2

• Home
• XML Files
• Search
• Login

Powered by <TEI:TOK>
Maarten Janssen, 2014-

R&D Unit funded by

FCT Fundação para a Ciência e a Tecnologia

en093CVITF

Native language	English
Other foreign languages	Afrikaans,French
Proficiency	A1
Genre	personalLetter
Prompt	Answer to a postcard about the informant's life in Portugal
Topic	Daily life
Tokens	99

View options

Text: Transcription | Student form | Teacher form - Show: Colors | Images - Tags: POS tag (ort) | Lemma (ort)

Olá,

Tudo bem? Como estás com a [fig1] sua (C) (MF) (CE) familia? Tudo bem aqui. Obrigado para (MF) este postal.

Esta optímo em Lisboa mas eu sinto falta a (MF) minha família e a (MF) (C) namorada. Lisboa é uma cidade lindíssimo (MF).

Sim, tenho muito amigos. Eles [...] são (C) muito (C) (MF) nacionalidades diferente (C) (MF).

Da segunda-feira á (CE) [...] sexta-feira, tenho aulas ás (MF) 8h até (MF) meio-dia. Depois (MF) aulas, tomo almoço na cantina da Universidade. Gosto da comida Português (CE) e o café é melhor.

```
<u who="#MJ" start="#T0" end="#T2">
  <seg type="intonation-phrase" subtype="falling">
    <w>I</w><vocal><desc>cough</desc></vocal><w>see</w><w>a</w><w>door</w>
  </seg>
  <anchor synch="#T1"/>
    <seg type="intonation-phrase" subtype="falling">
    <w>I</w><pause dur="PT0.3S"/><w>want</w><w>to</w><w>paint</w><w>it</w>
    <unclear><choice><seg><w>black</w></seg><seg><w>blue</w></seg></choice></unclear>
  </seg>
</u>

<u who="#MJ" start="#T0" end="#T2">
  <seg type="utterance" subtype="declarative">
    <w>I</w><vocal><desc>cough</desc></vocal><w>see</w><w>a</w><w>door</w>
  </seg>
  <anchor synch="#T1"/>
  <seg type="utterance" subtype="declarative">
    <w>I</w><pause dur="PT0.3S"/><w>want</w><w>to</w><w>paint</w><w>it</w>
    <unclear><choice><seg><w>black</w></seg><seg><w>blue</w></seg></choice></unclear>
  </seg>
</u>
```

Hedeland, Hanna, and Thomas Schmidt. 2022. ‘The TEI-Based ISO Standard ‘Transcription of Spoken Language’as an Exchange Format within CLARIN and Beyond’. In CLARIN Annual Conference, 34–45. https://doi.org/10.3384/ecp1894.

University of Basel

And also ...

Parliamentary data
Scientific publications (journal, proceedings, etc.)
TEI for distant reading
...

https://polmine.github.io/

About

Purpose

There is an unprecedented availability of digitized, politically relevant text. This opens up new horizons for social science research. Turning text into corpora and acquiring abilities to work productively with vast amounts of textual data will stimulate research on old and new research questions in the social sciences. Providing the data and the code to exploit the opportunities of digitalization for our discipline is the purpose of the PolMine Project.

Code

The formula "code is theory" drives what we develop. Valid research findings require to combine qualitative and quantitative analytical steps seamlessly in an interactive workflow. We see text as linguistic data. These ideas are implemented using the statistical programming language R. The R package polmineR is our core package for text analysis and is complemented by packages for corpus preparation.

Corpora

Our focus is to turn texts issued by public institutions into language resources for research. A digital public archive of democracy is the ultimate vision. GermaParl, a corpus of parliamentary protocols is our flagship corpus. We strive for a sustainable research data management that includes a fully reproducible data preparation workflow, work with standardized, TEI-compatible data formats, and involve users for quality management.

Tutorials

Research

Who

More on TEI for parliamentary data:

https://www.clarin.eu/event/2019/parlaformat-workshop

University of Basel

From TEI …

EpiDoc, Charters Encoding Initiative (CEI), Music Encoding Initiative (MEI) for specific communities

University of Basel

Thank you!

For more information, visit https://rise.unibas.ch and contact us at rise@unibas.ch