

# Small TEI Projects on a Large Scale: The TEI Archiving, Publishing, and Access Project

Julia Flanders

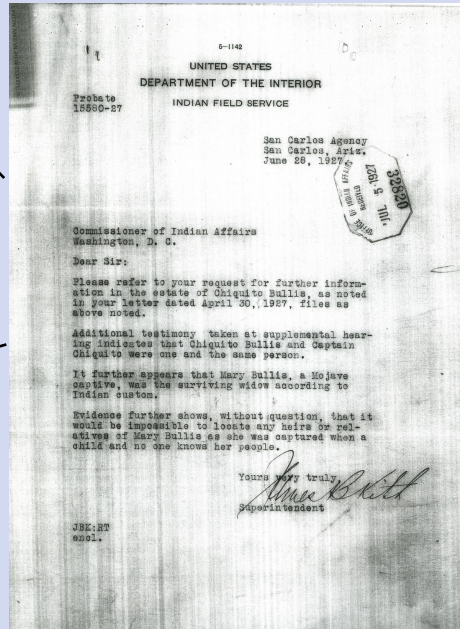
Women Writers Project

Center for Digital Scholarship

Brown University



**Archival representation:**  
Physical document with  
dimensions, damaged edges,  
coloring

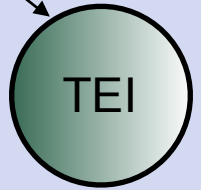
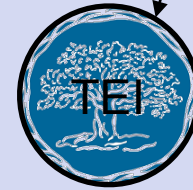
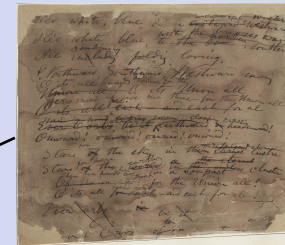
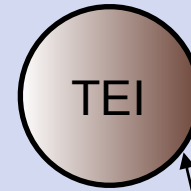
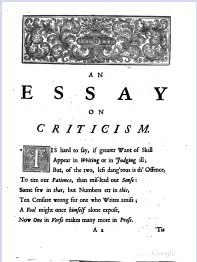
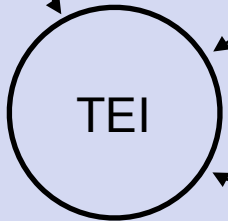
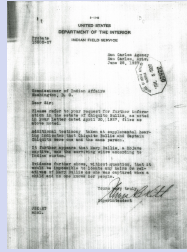
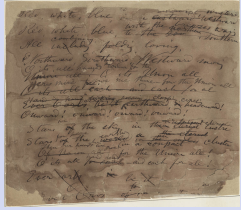


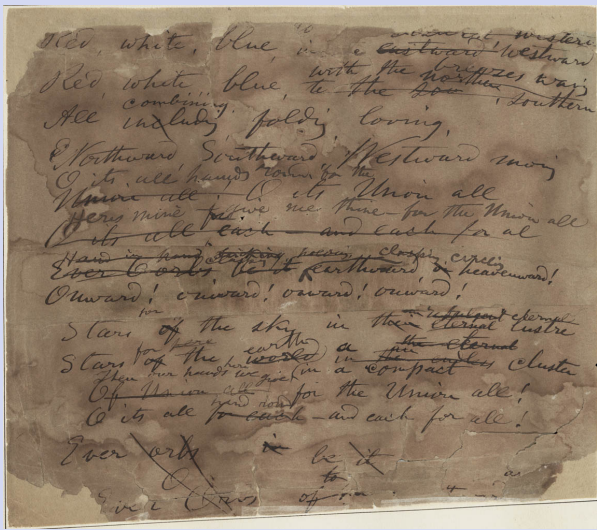
**Social-historical representation:**  
References to people, places,  
events, organizations

**Linguistic representation:**  
Parts of speech, phrasing,  
clauses

**Digital library representation:**  
Metadata record with details of authorship,  
date of creation, provenance, genre

**Genre-oriented representation:**  
Document structure, rhetoric





**Genre-oriented representation:**

Document structure, rhetoric

Interest in representing the author's final intentions

Interest in representing the revision process

Emphasis on thematics

Emphasis on formal structure

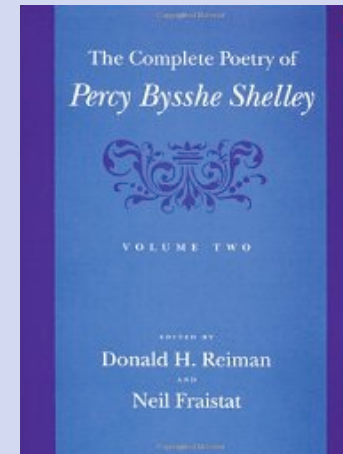
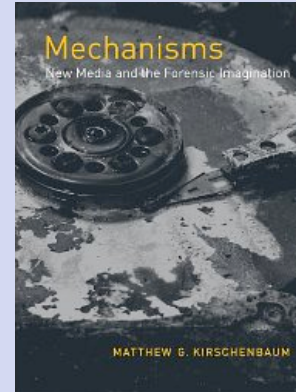
Different understanding of formal structure

One poem or several...?

Is this even a poem at all...?



```
<TEI>
  <teiHeader/>
  <text>
    <body>
      <p/>
    </body>
  </text>
</TEI>
```



<http://www.lib.uchicago.edu/e/scrc/search/>



<http://collation.folger.edu/2011/11/the-books-on-our-shelf/>



- Repository service
- Publication service
- API
- User data management interface



Strongly collaborative, community-driven:

Brown University, Wheaton College, Hamilton College, Mt. Holyoke College, Dickinson College, University of Virginia, Willamette University, University of Puget Sound...

Your institution here!



## **User interface, TAPAS schemas, TAPAS publication stylesheets, access control**

The Drupal front end of TAPAS provides the interface layer through which users can upload and manage their TAPAS data, configure publication options, and view published TAPAS projects.

In theory, Islandora provides an interface in Drupal for interacting with data stored in Fedora

## **TEI, image files, user-contributed schemas, user-contributed stylesheets, TAPAS metadata**

Each resource is stored in Fedora as a separate repository object, represented by a FOXML record. Relationships between objects are represented in RDF.



# Convergence and Divergence

- **Casual convergence** establishes a baseline of minimal function
- **Intentional convergence** establishes an realm of higher function
- **Intentional divergence** requires individualized support

# Challenges, Questions

How to balance the value of collective uniformity against the value of individual scholarly insight

How to encourage conformance (by offering functionality) without causing people to code to the tool

- How to deal with semantics: what can we know about the meaning expressed by markup?



	General	Verse	Drama	Letters	Manuscripts
Basic	<ul style="list-style-type: none"> <li>• major divisions</li> <li>• headings</li> <li>• paragraphs</li> <li>• renditional highlighting</li> <li>• figures</li> </ul>	<ul style="list-style-type: none"> <li>• poetic lines</li> <li>• line groups</li> </ul>	<ul style="list-style-type: none"> <li>• speeches</li> <li>• speakers</li> <li>• stage directions</li> <li>• act-scene boundaries</li> </ul>	<ul style="list-style-type: none"> <li>• datelines, signatures</li> <li>• basic paragraphing</li> </ul>	<ul style="list-style-type: none"> <li>• additions, deletions</li> <li>• gaps, illegible text</li> </ul>
Better	<ul style="list-style-type: none"> <li>• genre metadata</li> <li>• identification of names</li> <li>• quotations</li> <li>• direct speech</li> </ul>	<ul style="list-style-type: none"> <li>• identification of part-lines</li> <li>• identification of genre</li> </ul>	<ul style="list-style-type: none"> <li>• speaker identification</li> <li>• distinction between verse and prose</li> </ul>	<ul style="list-style-type: none"> <li>• openers, closers</li> </ul>	<ul style="list-style-type: none"> <li>• handwriting identification</li> </ul>
Optimal	<ul style="list-style-type: none"> <li>• named entity identification</li> </ul>	<ul style="list-style-type: none"> <li>• consistent marking of poem boundaries</li> <li>• rhyme</li> <li>• named entity identification</li> </ul>	<ul style="list-style-type: none"> <li>• named entity identification</li> </ul>	<ul style="list-style-type: none"> <li>• standardized letter metadata</li> <li>• named entity identification</li> </ul>	<ul style="list-style-type: none"> <li>• named entity identification</li> </ul>

# Probably out of scope...

- Individual renditional schemes
- Individual controlled vocabularies
- Special manuscript features
- Positionality and layout details
- Advanced domain-specific markup (e.g. genetic manuscript encoding)

# Basic Encoding

```
<div type="scene">
  <head>Scene 2</head>
  <stage>Lacrimilla at her dressing table, assisted by her maid Primula.</stage>
  <sp> <speaker>Lacrimilla</speaker>
    <p>La, how quiet the house is tonight! Surely some nonsense is afoot below stairs.</p>
  </sp>
  <sp><speaker>Primula</speaker>
    <p>Yes, Miss. To be sure, Miss.</p>
  </sp>
  <sp><speaker>Lacrimilla</speaker>
    <p>I wonder whether Floristan will be at the play tonight? La, how I shall quiz him! And how discomfited he will be!</p>
  </sp>
  <stage>She hums a tune.</stage>
  <sp><speaker>Primula</speaker>
    <p>Miss, the second measure...surely it should be in dotted crotchets?</p>
  </sp>
  <sp><speaker>Lacrimilla</speaker>
    <p>Hah? How so?</p>
  </sp>
  <sp><speaker>Primula</speaker>
    <p>That tune, Miss. It is Countess Almaviva's last aria from Figaro, is it not? The second measure is in dotted crotchets, not in triplets. If I may say so, Miss.</p>
  </sp>
  <sp><speaker>Lacrimilla</speaker>
    <p>Don't be an obsequious dolt, Primula. It's terribly vulgar. You're quite right. I should have looked a perfect fool in front of Floristan! What a narrow escape!</p>
  </sp>
  <stage>Primula curtsies and goes to the door. The curtain falls.</stage>
</div>
```

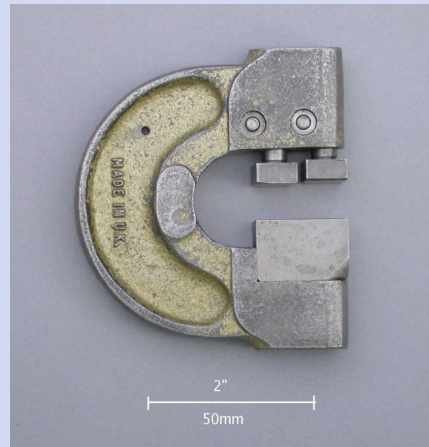
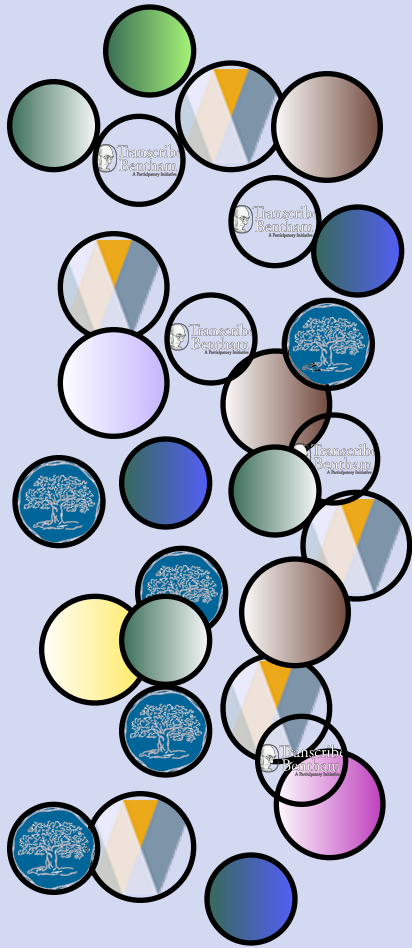
# Better Encoding

```
<div type="scene">
  <head>Scene 2</head>
  <stage>Lacrimilla at her dressing table, assisted by her maid Primula.</stage>
  <sp who="#lac"> <speaker>Lacrimilla</speaker>
    <p>La, how quiet the house is tonight! Surely some nonsense is afoot below stairs.</p>
  </sp>
  <sp who="#prim"><speaker>Primula</speaker>
    <p>Yes, Miss. To be sure, Miss.</p>
  </sp>
  <sp who="#lac" >speaker>Lacrimilla</speaker>
    <p>I wonder whether Floristan will be at the play tonight? La, how I shall quiz him! And how discomfited he will be!</p>
  </sp>
  <stage>She hums a tune.</stage>
  <sp who="#prim" >speaker>Primula</speaker>
    <p> Miss, the second measure...surely it should be in dotted crotchets?</p>
  </sp>
  <sp who="#lac" >speaker>Lacrimilla</speaker>
    <p>Hah? How so?</p>
  </sp>
  <sp who="#prim" >speaker>Primula</speaker>
    <p>That tune, Miss. It is Countess Almaviva's last aria from Figaro, is it not? The second measure is in dotted crotchets, not in triplets. If I may say so, Miss.</p>
  </sp>
  <sp who="#lac" >speaker>Lacrimilla</speaker>
    <p>Don't be an obsequious dolt, Primula. It's terribly vulgar. You're quite right. I should have looked a perfect fool in front of Floristan! What a narrow escape!</p>
  </sp>
  <stage>Primula curtsies and goes to the door. The curtain falls.</stage>
</div>
```

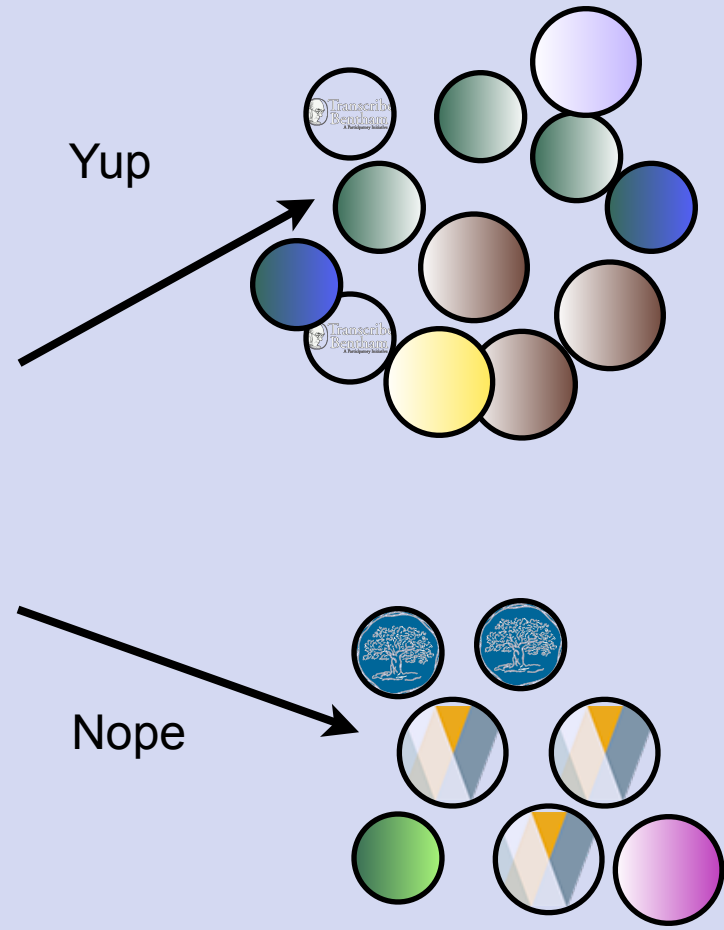
# Optimal Encoding

```
<div type="scene">
  <head>Scene 2</head>
  <stage><persName ref="#lac">Lacrimilla</persName> at her dressing table, assisted by her maid
    <persName ref="#prim">Primula</persName>.</stage>
  <sp who="#lac"> <speaker>Lacrimilla</speaker>
    <p>La, how quiet the house is tonight! Surely some nonsense is afoot below stairs.</p>
  </sp>
  <sp who="#prim"><speaker>Primula</speaker>
    <p>Yes, Miss. To be sure, Miss.</p>
  </sp>
  <sp who="#lac" ><speaker>Lacrimilla</speaker>
    <p>I wonder whether <persName ref="#flo">Floristan</persName> will be at the play tonight? La, how I shall quiz him!
      And how discomfited he will be!</p>
  </sp>
  <stage>She hums a tune.</stage>
  <sp who="#prim" ><speaker>Primula</speaker>
    <p> Miss, the second measure...surely it should be in dotted crotchets?</p>
  </sp>
  <sp who="#lac" ><speaker>Lacrimilla</speaker>
    <p>Hah? How so?</p>
  </sp>
  <sp who="#prim" ><speaker>Primula</speaker>
    <p>That tune, Miss. It is <persName ref="#almaviva">Countess Almaviva's</persName> last aria from
      <title>Figaro</title>, is it not? The second measure is in dotted crotchets, not in triplets. If I may say so, Miss.</p>
  </sp>
  <sp who="#lac" ><speaker>Lacrimilla</speaker>
    <p>Don't be an obsequious dolt, <persName ref="#prim">Primula</persName>. It's terribly vulgar. You're quite right. I
      should have looked a perfect fool in front of <persName ref="#flo">Floristan</persName>! What a narrow escape!
    </p>
  </sp>
  <stage> <persName ref="#prim">Primula</persName> curtsies and goes to the door. The curtain falls.</stage>
</div>
```

# The role of constraint



Is it TEI?

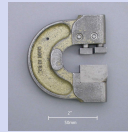
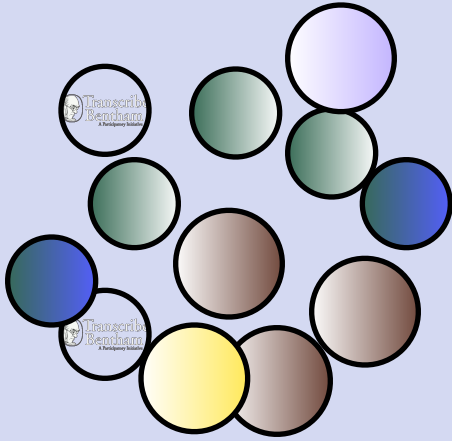


Yup

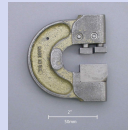
Nope



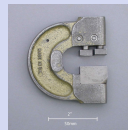
# The role of constraint



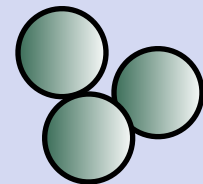
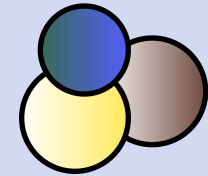
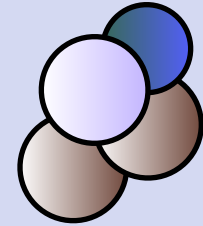
Will it do X...?



Does it contain Y...?



Is it missing Z...?



# Challenges and Questions

- What is the useful horizon of our curatorial responsibility? What happens next?
- What information do we need to gather at the point of submission to be able to fulfill that responsibility?
- What is the right balance between autonomy and convergence?
- What can TAPAS do for readers and users as well as contributors?



Thank you!

<http://www.tapasproject.org>  
[info@tapasproject.org](mailto:info@tapasproject.org)

