

# META(DATA)MORPHOSIS

---

Ashley M. Clark, XML Applications Developer  
Sarah Connell, Project Manager  
Northeastern University Women Writers Project

XML In, Web Out  
August 1, 2016

# The Women Writers Project and Women Writers Online

- 390 texts written, translated by, or attributed to women
- Primarily print texts that were first published between 1526 and 1850
- TEI-encoded & published through the WWO interface

The screenshot displays the Women Writers Online (WWO) interface. At the top, the title "Women Writers Online" is visible. Below it, there is a search bar with the placeholder text "Search..." and an "OK" button. To the left of the search bar, there is a "Filter" section with two categories: "Genre" and "Date". Under "Genre", there are links for "non-fiction [175]", "verse [141]", "drama [67]", "fiction [43]", "(3)", and "prose [2]". Under "Date", there are links for "16th c. [29]", "17th c. [164]", "18th c. [102]", and "19th c. [86]". To the right of the search bar, there is a "Results: 381 texts" section. Below this, a list of search results is displayed, including titles like "[unknown] — The Bridling, Saddling, and Riding of a Rich Churl in Hampshire, 1595", "[unknown] — The Wonderful Discoverie of the Witchcrafts of Margaret and Phillip Flower, 1619", "[unknown] — Swetnam, the Woman-Hater, Arraigned by Women, 1620", "[unknown] — Eliza's Babes, 1652", "[unknown] — The Female Wits, 1704", "[unknown] — The Fortunate Transport, 1750", "[unknown] — England's Tears: A Poem, 1774", "[unknown] — Changing Scenes, 1825", "[unknown] — Memoir of Mrs. Chloe Spear, 1832", "Adams, Hannah — A Summary History of New-England, 1799", and "Adams, Hannah — The History of the Jews, 1812".

# The Cultures of Reception initiative

- Goal: support collaborative research into the transatlantic reception and readership of texts by women
- Begun at Brown University in 2010: first phases included selecting, sourcing, & gathering data on the texts to be transcribed; setting up a transcription interface; and a substantial amount of the encoding work
- In 2013, the WWP moved to Northeastern University; since then, priorities have been:
  - continuing with transcription, and
  - preparing for publication—data cleanup and creating a publication interface

# The *Critical Review* on Maria Edgeworth

ART. 26.—*Leonora*; by Miss Edgeworth. 8vo. 2 Vols. Johnson. 1806.

THE professors of modern philosophy have been already hunted down by moral writers with such vigour that we trust very few of the race remain ; but while a single animal of this description exists, the efforts towards a complete extermination must not be relaxed :—there is now less glory in the enterprize, but the attempt is in itself always meritorious.

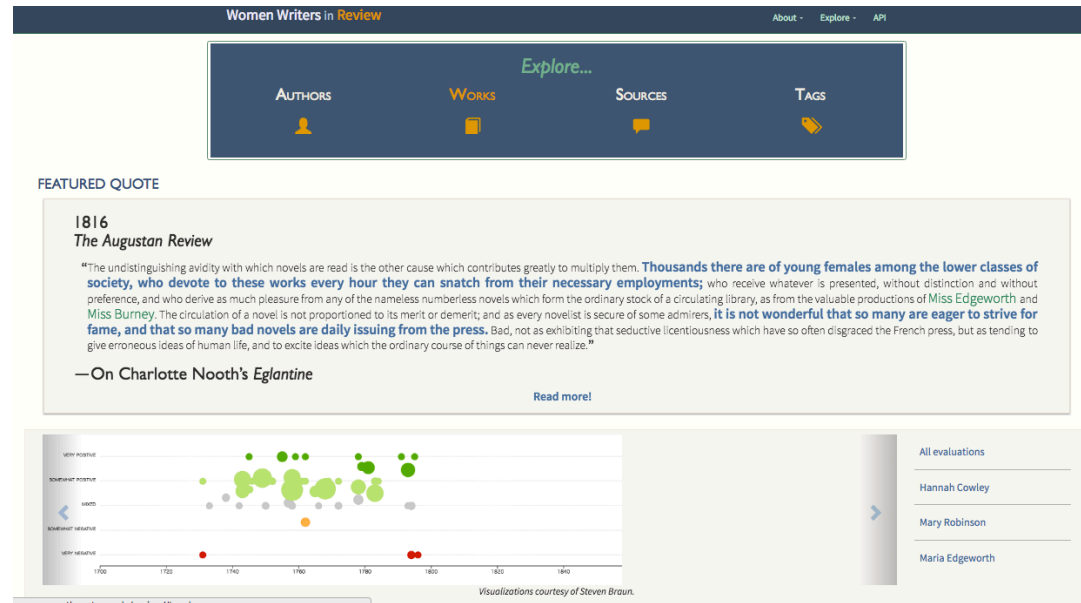
This novel is written in a series of letters. Leonora is a virtuous woman, and attributing the reports which she hears of Olivia's conduct to the mischievous spirit of scandal and to the malignity of envy, invites her to her house as an asylum from the persecutions of the malicious. Olivia is a professor of the modern philosophy, and has no other conceptions of the rules of right and wrong, than of rules for the game of whist, which may be very useful in the game of life, but which may be broken through or complied with in any particular emergency. She comes ripe from France, a determined foe to all those restraints which confine tide-less blooded females within

Source: *HathiTrust*

<http://hdl.handle.net/2027/njp.32101076402963>

# Cultures of Reception and Women Writers in Review

- 690 texts responding to works written or translated by women
- Literary and theatrical reviews, publication notices, textual extracts &c.
- Published between 1770 and 1830
- TEI-encoded & published through the Women Writers in Review interface



# Two TEI projects focused on women's texts

## **Women Writers Online**

- First published 1999
- Substantial tagset (around 165 elements)
- Designed as a reading interface
- Encoded texts are very generically & chronologically diverse

## **Women Writers in Review**

- First published 2016 (very soon!)
- More limited tagset
- Interface foregrounds discovery & exploration
- Encoded texts tend to be brief, represent a more constrained set of genres & publication dates

# Publication challenges and goals

## **Challenges:**

- Substantially increased need for context to make each “review” useful (i.e., not just the publication details for the review itself, but also those for the authors and texts being discussed)
- Relative obscurity of transcribed materials (both in the likelihood that readers would look for individual reviews and in their authorship and titles)

## **Goals:**

- Linking between authors and texts in both interfaces
- Making the reviews discoverable and navigable: easy to explore and useful for research

# The transcription interface

## Transcription

The complete transcription  
of the reception item or  
review.

### Transcription

```
<head>Art. 26.—<italic><title>Leonora</title>; by <persName>Miss Edgeworth</persName>. 8vo.  
2 Vols.</italic> Johnson. 1806.</head>  
<p>The professors of modern philosophy have been already hunted down by moral writers with  
such vigour that we trust very few of the race remain; but while a single animal of this description  
exists, the efforts towards a complete extermination must not be relaxed:—there is now less glory  
in the enterprize, but the attempt is in itself always meritorious.</p>  
<p>This novel is written in a series of letters. <persName>Leonora</persName> is a virtuous  
woman, and attributing the reports which she hears of <persName>Olivia's</persName> conduct  
to the mischievous spirit of scandal and to the malignity of envy, invites her to her house as an  
asylum from the persecutions of the malicious. <persName>Olivia</persName> is a professor of  
the modern philosophy, and has no other conceptions of the rules of right and wrong, that of  
rules for the game of whist, which may be very useful in the game of life, but which may be  
broken through or complied with in any particular emergence. She comes ripe from
```

The transcription of the reception item's content.

### Excerpt or partial transcription?

☐ Check if the transcription is an excerpt from a longer reception item

### Quotations from other texts?

☐ Check if the transcription contains quotes from works other than the primary text



# The transcription interface: tagging themes

## Transcription Details

More information about the transcription of the reception item or review.

### General evaluation

Somewhat positive

### Themes

- |  |   |
|--|---|
| <input type="checkbox"/> Class or socioeconomic status | <input type="checkbox"/> Gender                     |
| <input checked="" type="checkbox"/> Nation or empire   | <input type="checkbox"/> Place, space, or geography |
| <input type="checkbox"/> Race                          |   |

### Tags

Christianity; religion; morality

Tags or general keywords that describe or identify significant aspects of the reception item. Separate multiple keywords with a semi-colon.

Update


# Transcription architecture

- Designed to get encoders of all skill levels working with little training
- CouchDB backend
  - NoSQL, JSON-based database
  - Accessible via web interface or HTTP request
- BackboneJS frontend
  - Connected to CouchDB API
  - Contains an index of records in the Couch database
  - Allows users to edit and update JSON records

# JSON records in Couch DB

Overview > reception > c03a5d2e8f72f27f185b13fe0a048325

collection	"reception"
created	1335968168000
creator	"jmelson"
docDate	"1806-02"
eval	"part_positive"
isWellformed	true
modified	1465824508000
pages	"215-16"
rcvdAuthor	"Edgeworth, Maria"
rcvdDocDate	"1806"
rcvdTitle	"Leonora"
source	"The Critical Review; or, Annals of Literature"
subtype	"literary review"
title	"Art. 26.4Leonora; by Miss Edgeworth. 8vo. 2 Vols. Johnson. 1806."
transcription	"<head>Art. 26.—<i><title>Leonora</title>; by <persName>Miss Edgeworth</persName>. 8vo. 2 Vols. </i></i> Johnson. 1806.</head><p>The professors of modern philosophy have been already hunted down by moral writers with such vigour that we trust very few of the race remain; but while a single animal of this description exists, the efforts towards a complete extermination must not be relaxed:—there is now less glory in the enterprize, but the attempt is in itself always meritorious.</p><p>This novel is written in a series of letters. <persName>Leonora</persName> is a virtuous woman, and attributing the reports which she hears of <persName>Olivia's</persName> conduct to the mischievous spirit of scandal and to the malignity of envy, invites her to her house as an asylum from the persecutions of the malicious. <persName>Olivia</persName> is a professor of the modern philosophy, and has no other conceptions of the rules of right and wrong, that of rules for the game of whist, which may be very useful in the game of life, but which may be broken through or complied with in any particular emergence. She comes ripe from <placeName>France</placeName>, a determined foe to all those restraints which confine <i>tide-less blooded</i> females within the pale of virtue and decorum. and."

  
**CouchDB**  
relax

Tools

- Overview
- Configuration
- Replicator
- Status
- Test Suite

Recent Databases

- reception

[Signup or Login](#)  
Futon on Apache CouchDB 1.0.4

# Pre-publication decisions: XML

- Transcriptions in CouchDB aren't XML but JSON strings. Well-formedness and validity aren't guaranteed.
- The Women Writers Project has invested much of its architecture and expertise into XML technologies, especially TEI encoding and XSLT.
- To make publication of transcriptions easier, we should convert the reviews from JSON records to TEI-encoded XML files.

# Pre-publication decisions: Metadata

- There are over 600 JSON records with transcriptions.
  - A single index would give a human user information overload.
  - A human user would find more value in browsing reviews in manageable sets, by categories.
- Each record includes a variety of contextual information:
  - Information on the record itself (such as date last edited),
  - Information on the source of the review, often a journal or magazine,
  - Information on the reviewed author, and
  - Information on the reviewed work (including the probable edition reviewed).

# The Inspector

- Most of the existing metadata was serviceable, but prone to inconsistency.
- Normalizing the metadata manually would be tiresome for a human.

Our “robots”—XQueries and XSLT report-makers—could not easily identify distinct entities (misspelled names, pseudonyms, etc.).

- Instead, we decided on a composite, cyborg approach.

# The Inspector: Report maker

```

• reception.xql × • index.html × • couch.xql ×
271 {~
272 : Return a truncated report on each ill-formed transcription in CouchDB.
273 :
274 : @return XML
275 :}
276 declare function cor:getIllformedTranscriptions() {
277   let $docSeq := cor:getAll("_design/reception/_view/transcription", ())
278   return
279     for $record in $docSeq//item[@type eq "object"]
280     return
281       if ( $record//pair[@name eq "isWellformed"]/@type eq 'boolean' and $record//pair[@name eq "isWellformed"]/text() eq 'true' ) then
282         ()
283       else
284         let $isWellformed := cor:isWellformed($record//pair[@name eq "transcription"])
285         return
286           if ( $isWellformed ) then
287             ()
288           else
289             <corItem>
290               <source>{ $record//pair[@name eq "source"]/text() }</source>
291               <date when="{ $record//pair[@name eq "docDate"]/text() }"/>
292               <rcvdAuthor>{ $record//pair[@name eq "rcvdAuthor"]/text() }</rcvdAuthor>
293               <rcvdTitle>{ $record//pair[@name eq "rcvdTitle"]/text() }</rcvdTitle>
294               <id>{ $record/pair[@name eq "id"]/text() }</id>
295               <rev>{ $record/pair[@name eq "key"]/text() }</rev>
296             </corItem>
297 };
298
299 {~

```

# The Inspector: Ill-formedness index

inspector

Log in to eXist

## Reports

illformed-cor

## Tadpole Generator

Cultures of Reception

## illformed-cor













CouchDB username

CouchDB password

Log into CouchDB

Regenerate report

155 Reception transcriptions are ill-formed. 1 have been edited.

Received Author	Received Title (Trimmed)	Source	Date		
Adams, Hannah	A View of Religions	The American Review, and Literary Journal	1802-03		
Aikin, Lucy	Epistles on Women	The Monthly Review; or Literary Journal	1811		
Bannerman, Anne	Epistle from the Marquis de Lafayette to General Washington	The Critical Review; or, Annals of Literature	1800-09		
Barbauld, Anna Laetitia (Aikin)	A Legacy for Young Ladies	The Monthly Review; or Literary Journal	1826		
Barbauld, Anna Laetitia (Aikin)	The Works of Anna Laetitia Barbauld	The Monthly Review; or Literary Journal	1825		
Barbauld, Anna Laetitia (Aikin)	The Works of Anna Laetitia Barbauld	The North American Review	1826-10		



# Ill-formedness report: Editor

inspector

Log in to eXist

## Reports

illformed-cor

## Tadpole Generator

Cultures of Reception

## illformed-cor

CouchDB username

CouchDB password

Log into CouchDB

Submit correction ↗

```
1 <head>ART VII. <title><title>A View of Religions, in two Parts. Part</title> i. <title>Containing an Alphabetical
  Compendium of the various Religious Denominations which have appeared in the World, from the beginning of the
  Christian Aera to the present Day. Part</title> ii. <title>Containing a brief Account of the different Schemes of
  Religion now embraced among Mankind. The whole collected from the best Authors, ancient and modern.</title> By</title>
  <persName>Hannah Adams</persName>. <title>The third Edition, with large Additions.</title> 8vo. pp. 504.
  <placeName>Boston</placeName>. <persName>Manning</persName> <title>and</title> <persName>Loring</persName>. 1801
  .</head>
2
3 <p>The first edition of this work was published about ten or twelve years ago. A second was soon called for, and made its
  appearance, with considerable improvements. The demand for a third, and much larger impression, within so short a
  period, is one, among many proofs, of the growing taste for reading, and the increased sale of books, which it is
  pleasing to observe in our country.</p>
4
5 <p>Of the compiler of the present volume, our readers will recollect that we spoke in terms of respect, on a former
  occasion,* as the author of a <title>Summary History of <placeName>New-England</placeName></title>, <title>
  &#38;c. If she appeared to advantage in narrating civil and political events, and took an honourable station among
  American historians, she is not less worthy of praise, in pourtraying the different christian denominations, and in
  describing the various departments of the religious world.-Religion is a subject with which every man is bound to be
  acquainted. This is a duty which he owes to himself as a rational being, to say nothing of higher and more interesting
  grounds of obligation. Nor is it enough for a liberal mind to make himself acquainted with that system of religious
  belief which is agreeable to scripture: he will also desire to know what other systems have been taught; by what
  arguments those creeds which differ from the truth have been defended; and to explore, as far as possible, the various
  forms and sources of error.-Such minute inquiries are not, indeed, incumbent on every individual; many have not
  leisure, and many more have not talents adequate to the task of making them, but they are desirable to all who have
  leisure and talents; and we are persuaded, in opposition to the opinion of some, that such investigations, when
  thoroughly and candidly made, instead of unhinging the mind, or diminishing the firmness and stability of christian
  faith, will confer additional light, decision and steadiness on both. <title>The truth as it is in Jesus</title>
  will appear more excellent and valuable when contrasted with delusive notions; and the mind will acquire new
  satisfaction and confidence on returning from the regions of falsehood, covered with briars and thorns, to the goodly
  fields which revelation sets before us.</p>
6
```

# The Inspector: Ill-formedness report

• reception.xql × • index.html × • couch.xql × • illformed-cor.xml ×

set corItem rcvdAuthor

```
1 <set xml:id="illformed-cor" toHTML="modules/illformed-cor.xq">
2   <corItem>
3     <source>The Augustan Review; a Monthly Production</source>
4     <date when="1816-05"/>
5     <rcvdAuthor>Morgan, Lady Sydney (Owenson)</rcvdAuthor>
6     <rcvdTitle>O'Donnell: A National Tale</rcvdTitle>
7     <id>c03a5d2e8f72f27f185b13fe0a0af25d</id>
8     <rev>10-afe5b4d8ffcca75e172e2edfb3e96940</rev>
9   </corItem>
10  <corItem>
11    <source>The New Annual Register, or General Repository of History, Politics, and Literature</source>
12    <date when="1792"/>
13    <rcvdAuthor>Smith, Charlotte (Turner)</rcvdAuthor>
14    <rcvdTitle>Desmond</rcvdTitle>
15    <id>c03a5d2e8f72f27f185b13fe0a112950</id>
16    <rev>10-fae7730c430034f40d87e6416b8aa09a</rev>
17  </corItem>
18  <corItem>
19    <source>The Critical Review; or, Annals of Literature</source>
20    <date when="1817-05"/>
21    <rcvdAuthor>Edgeworth, Maria</rcvdAuthor>
22    <rcvdTitle>Comic Dramas, in Three Acts</rcvdTitle>
23    <id>c03a5d2e8f72f27f185b13fe0a060532</id>
24    <rev>11-108f31a61fda6434377e81f4264a0b35</rev>
25  </corItem>
26  <corItem>
27    <source>The Critical Review; or, Annals of Literature</source>
28    <date when="1797"/>
29    <rcvdAuthor>Edgeworth, Maria</rcvdAuthor>
30    <rcvdTitle>Letters for Literary Ladies. To Which Is Added, an Essay on the Noble Science of Self-Justification</rcvdTitle>
31    <id>bee857503f04ddd247104bf6b602af26</id>
```

# JSON to TEI

- XQuery turned JSON records into intermediate XML.
- XSLT transformed each record into TEI. The XSLT also did a pass where it tagged the new TEI for potential metadata problems, that we'd eventually fix in the new files.
- The new files were added to a Subversion repository as the new canonical records for the reviews.
- The combined metadata for each named entity was then drawn into its own record within a TEI file for that category of entity.

# Sample bibliography entry

```
<biblStruct xml:id="CR0030" corresp="#frbr.work">
  <monogr corresp="#frbr.expression">
    <author>
      <persName ref="persons.xml#medgewort.onp">Edgeworth, Maria</persName>
    </author>
    <title type="main" subtype="display">Comic Dramas, in Three Acts</title>
    <respStmt>
      <resp>contributor</resp>
      <persName>Edgeworth, Richard Lovell</persName>
    </respStmt>
    <note>12mo. pp. 286</note>
    <imprint xml:id="CR0030.1" corresp="#frbr.manifestation">
      <publisher>
        <name>Hunter, R.</name>
      </publisher>
      <pubPlace>
        <placeName><settlement>London</settlement>, <country>England</country></placeName>
      </pubPlace>
      <date when="1817"/>
    </imprint>
    <imprint xml:id="CR0030.2" corresp="#frbr.manifestation">
      <publisher>
        <name>Wells and Lilly</name>
      </publisher>
      <pubPlace>
        <placeName><settlement>Boston</settlement>, <region>Massachusetts</region>, <country>United States of America</country></placeName>
      </pubPlace>
      <date when="1817"/>
    </imprint>
  </monogr>
</biblStruct>
```

# Creating display titles

- Most reviews are anonymous or pseudonymous, and unlikely to be searched by their titles, so we needed to give each one labels that would be useful & intelligible, eg:

1817-09: *The Edinburgh Magazine* on Edgeworth's *Comic Dramas*

- Both sources and reviewed works tend toward the verbose:  
*A View of Religions, in two Parts. Part i. Containing an Alphabetical Compendium of the various Religious Denominations which have appeared in the World, from the beginning of the Christian Æra to the present Day. Part ii. Containing a brief Account of the different Schemes of Religion now embraced among Mankind. The whole collected from the best Authors, ancient and modern*

# For example

```
<biblStruct xml:id="satmag">
  <monogr>
    <title type="main">The Saturday Magazine: Being in Great Part a Compilation
      from the British Reviews, Magazines, and Scientific Journals</title>
    <title type="display">The Saturday Magazine</title>
    <imprint><pubPlace><placeName><settlement>Philadelphia</settlement>,
      <region>Pennsylvania</region>, <country>United States of America</country></placeName></pubPlace></imprint>
  </monogr>
</biblStruct>
```

1821-07-21: *The Saturday Magazine*  
on Edgeworth's *Early Lessons*

1821-10-06: *The Saturday Magazine*  
on Carter

1822-03-09: *The Saturday Magazine*  
on Aikin's *The Memoirs of the Court of King James the First*

1822-03-30: *The Saturday Magazine*  
on Bengier's *Memoirs of the Life of Anne Boleyn*

[The Saturday Magazine](#)  
Philadelphia, Pennsylvania, United States of America

[The Scots Magazine](#)  
Edinburgh, Scotland

[Select Reviews of Literature](#)  
Philadelphia, Pennsylvania, United States of America

[The Souvenir](#)  
Philadelphia, Pennsylvania, United States of America

[The Theatrical Censor](#)  
Philadelphia, Pennsylvania, United States of America

[The Theatrical Inquisitor](#)  
London, England

[The Time Piece](#)  
New York, New York, United States of America

[The Town and Country Magazine](#)  
London, England

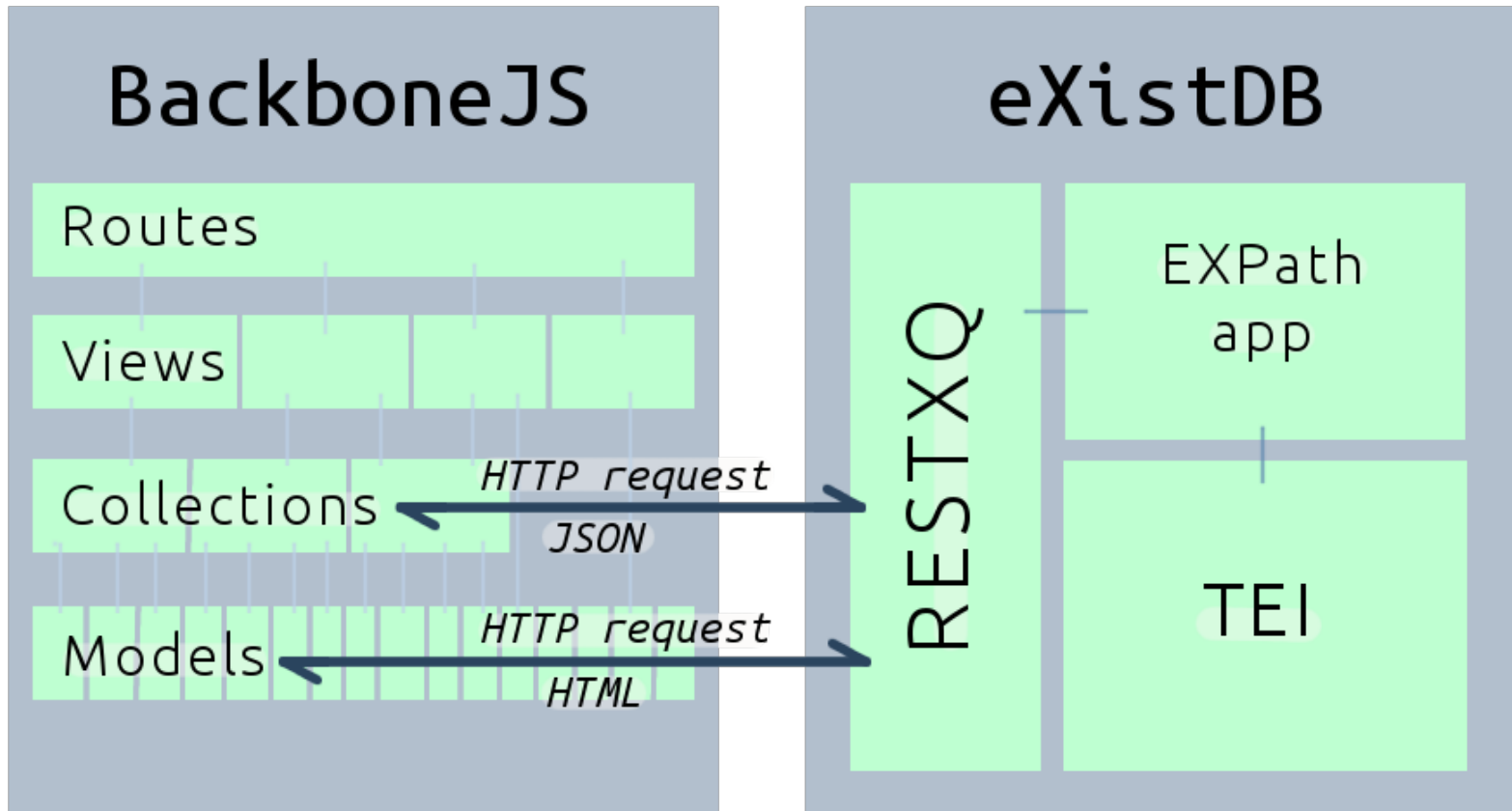
# Women Writers in Review: Goals

- Making these materials easy to browse and search
- Providing extensive linking among texts and authors
- Supporting discovery & exploration
- Offering a clean and readable display

Helping researchers ask questions like:

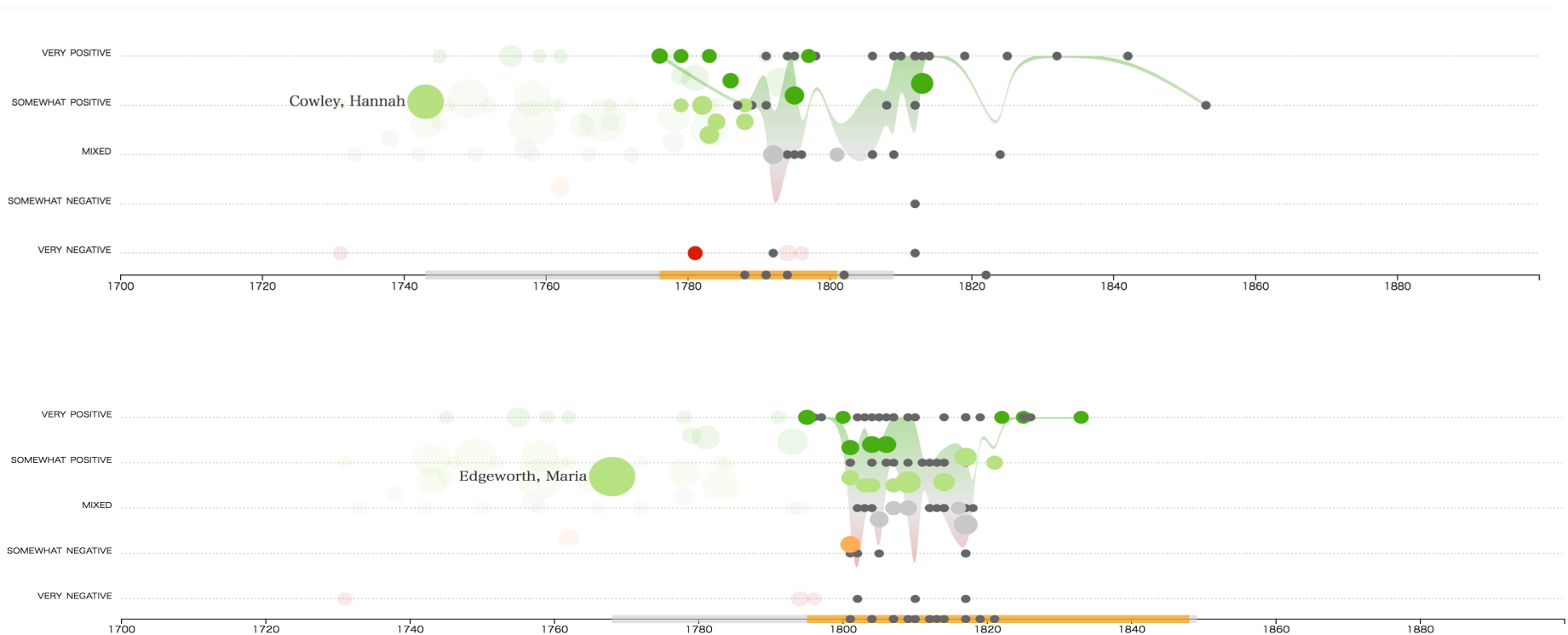
- How has Hannah Cowley been reviewed in British and American periodicals?
- What changes are evident in the *British Critic* over time?
- How did periodicals in this period discuss questions of women's authorship?

# Women Writers in Review: Components





# Visualizing evaluations over time



Prototype visualizations created by Steven Braun,  
Digital Scholarship Group, Northeastern University Libraries

# THANK YOU!

---

[wwp@northeastern.edu](mailto:wwp@northeastern.edu)

[@Nuwwp](#)

[wwp.northeastern.edu/review](http://wwp.northeastern.edu/review)

[wwp.northeastern.edu/blog](http://wwp.northeastern.edu/blog)