

# Documenting Workflows for HTR to TEI Conversions for Cultural Institutions: The Evolving Hands Project



James Cummings, Alexandra Healey, Diane Jakacki, Ian Johnson,  
Carrie Pirmann, Valentina Flex, Evie Jeffrey

<https://doi.org/10.5281/zenodo.8135692>

<https://research.ncl.ac.uk/evolvinghands/>

# “Evolving Hands: Building Workflows and Scalable Practices for Handwriting Recognition and Text Encoding in Cultural Institutions”

Funded by an NEH/AHRC

“New Directions for Digital Scholarship in Cultural Institutions” grant

James Cummings (Newcastle) and Diane Jakacki (Bucknell), principal investigators  
(with the help of many others)



NATIONAL ENDOWMENT  
FOR THE HUMANITIES



Arts and  
Humanities  
Research Council

# Evolving Hands Case Studies:

## Newcastle University: Gertrude Bell Diaries

- Exploring use of Transkribus Web for tagging elements
- Modified up-conversion of page2tei XSLT
- Comparison Oxygen vs LEAF-Writer for enhancing data



## Bucknell University: Various Projects

- Scholarly production at scale on multiple projects as a pipeline
- Unusual printed texts such as Suzette project and Heresies zine
- Integration with LEAF-VRE



## Joint Project: Newcastle and Bucknell

- Legacy Printed Records of Early English Drama Volumes for REED London and Newcastle-Upon-Tyne
- Structured printed text with in-text expanded *abbreviations*
- Bespoke cypher conversion



# Evolving Hands Newcastle University case study

Archive of Gertrude Bell (1868 - 1926)

Explorer / Archaeologist / Political Agent

Vast corpus of digitised content available online since 1998  
[gertrudebell.ncl.ac.uk](https://gertrudebell.ncl.ac.uk). Recently revamped website. Wanted to explore innovative ways to bring unseen material to the public and enhance transcripts.

Selected Transkribus platform as the most popular user facing tool for creating transcripts of historic documents (2021)

Pay as you go, selected with goal of keeping resulting workflows simple and accessible

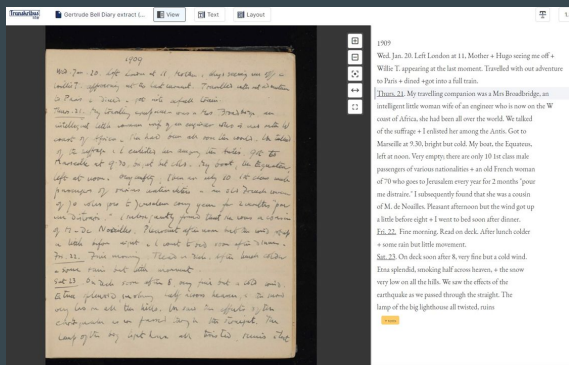


# Handwritten Text Recognition

1 Transcribe

2 Train

3 Refine



# Text encoding

abbr	expansion			
add	place			
sic	correction			
unclear	alternative	reason		
date	when			
orgName				
persName	firstname	dateOfDeath	dateOfBirth	lastname
placeName	country	placeName		

1 Schema

2 Environment

3 Tagging

# Next steps - finalising data and publication

```
<body>
  <pb facs='#facs_1' n='1'/>
  <p facs='#facs_1_r1'>
    <lg>
      <l facs='#facs_1_r1l1'>1909 </l>
      <l facs='#facs_1_r1l2'><date when='1909-1-20' month='1' year='1909' day='20'>
Wed. Jan. 20.</date> Left <placeName country='Eng' ref='Getty Geographic'
placeName='London'><country>Eng</co
firstname='Florence' dateOfDeath='1
'Bell'><forename>Florence</forename
</birth><death>1931</death>Mother<
```

Please select album

Navigation sidebar for 'Trevelyan Family Album Volume 1'.

- About this item
- LEFT
  - Trevelyan Family Album Volume 1
- RIGHT
  - Trevelyan Family Album Volume 1
- RESOURCE
  - Trevelyan Family Album Volume 1
  - Bibliographic citation CPT/PA/1
  - Collection Trevelyan (Charles Phillips) Archive
  - Creator Trevelyan Charles Phillips; Trevelyan Mary Katharine
  - Date created 1894; 1895; 1896; 1897; 1898; 1899; 1900; 1901; 1902; 1903
  - Description Album of photographs and ephemera compiled



# Evolving Hands: Bucknell Case Studies

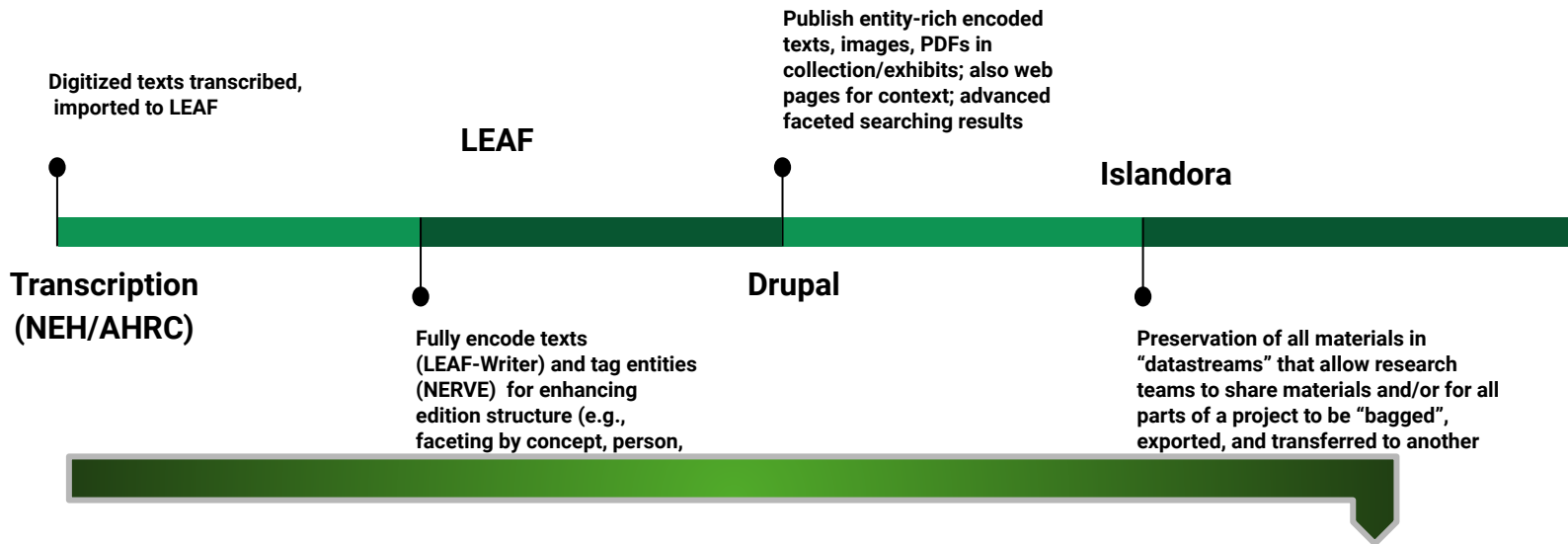
Research question:

- Diane Jakacki  
(Digital Scholarship Coordinator/  
Affiliate Faculty in CDHU)
- Carrie Pirmann  
(Social Sciences Subject Librarian)

How can we best integrate HTR and  
OCR into digital scholarly production  
pipelines for multiple / multiple types  
of research projects?

---

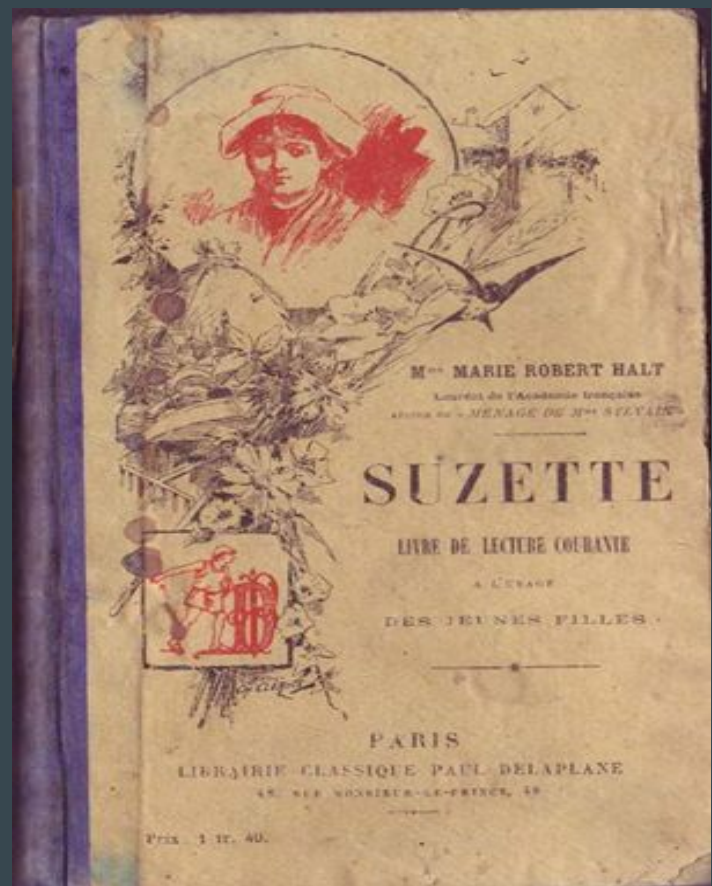
# The LEAF Production Pipeline



**LEAF**  
Linked Editing  
Academic Framework







# The Suzette Project (Prof. John Westbrook, PI)

12 SUZETTE.

toutes couleurs regardaient M<sup>me</sup> Valon en train de feuilleter un livre.

M<sup>me</sup> Valon était l'institutrice nouvelle de Fragicourt, arrivée depuis un mois à peine avec son mari, le maître de l'école des garçons. Elle leva son visage encore jeune, promena autour de la classe un regard gravement souriant et commença la lecture attendue :

LA FAMILLE AUX TEMPS PRÉHISTORIQUES.

C'était un lieu sauvage. De grands arbres surplombaient le terrain; des rochers, des ébénistes d'énormes, suspendus aux talus ravins\*, aux branches, aux défilures des rochers, formant de leurs innombrables et souples haies\* un inextricable\* rideau.

Au profond de la forêt s'étendaient des chemins sinistres, les terribles du loup, du grand ours des cavernes. Là et là un roucoulement de tourterelle, un pépiement\* dépassaient rompent, comme un murmure d'espérance, la sombre désolation de ces lieux.

Tout à coup, dans le sifflet à peine frayé, les branches sèches créquaient.

Vêtus de peaux de bêtes, les cheveux et la barbe incultes et longs, le corps ploqué sous le poids d'une bêche morte, sa main musculeuse appuyée sur un épié\*, un homme s'avancait



Loup.

Ce quadrupède est vorace et d'une force grande; il habite nos éplaves et on le rencontre fréquemment en France.



Oye.

Ce quadrupède, également vorace, est un grand chat; on le trouve principalement en Sibirie et au Canada.

LE PETIT HÉROÏSME. 25

pillons; Charlot court à ses côtés, et probablement monsieur François grimpe de branche en branche à la poursuite des oiseaux.

Mais ici on balaye, on range, on frotte; et parfois, il voit un verre par terre ou le balai qui s'allonge comme un balourd\* et accroche une pile d'assiettes.

— Ah! il y a des jours où tout va de travers!

Questionnaire. — Monsieur est content? — Que faisiez-vous ce matin, les amis de Suzette? — Je me suis levé à six heures. — Où est votre chambre? — Dans la chambre de monsieur. — Comment passez-vous votre temps, cette semaine? — Je suis allé à la messe, j'ai écrit quelques lettres. — Quel métier faites-vous? — Je suis instituteur.

DÉVELOPPEMENT DES SUJETS PROPOSÉS POUR EXERCICES.

Morale. — Montre, en citant un exemple, que vous comprenez ce que c'est de sacrifier son plaisir à son devoir.

Maman est occupée à un travail qu'elle ne peut abandonner un instant; moi, je joue dans le jardin avec mes amis, et voici que j'entends mon frère crier dans son bureau: je pourrais continuer à jouer, mais combien le pauvre petit se tourmenterait, et combien maman serait chagrin de ne pouvoir venir à son secours! Alors je sacrifie mon plaisir à mon devoir, et prenant le bébé dans mes bras, je m'efforce de soulager son mal et de l'égayer par mes sourires ou de douces paroles.

Il en est de même lorsque, pour préparer mes leçons du lendemain, je refuse de prendre part à une partie de plaisir organisée par mes compagnons.

Histoire. — Quelle est votre patrie? — Pourquoi l'aimez-vous? — Que font les enfants lorsque, pendant la guerre, ils entrent en France? — Qui défend la patrie? — Connaissez-vous des femmes qui ont exposé ou sacrifié leur vie pour la patrie?

Un patriote est la France, cette belle et vaste contrée habitée par une nation honnête, laborieuse et vaillante. Je l'aime, parce que sur son sol habitent ceux qui sont l'orgueil de nos affections; parce que c'est là que nos ancêtres ont vécu et que reposent leurs restes vénérés; parce qu'il y a rien dans le monde de grand et de généreux auquel elle n'ait participé. C'est de plus un magnifique pays, fertile, pittoresque et situé de la manière la plus heureuse sur le continent européen.

Lorsque les ennemis envahissent la France, et nous avons connu cette honte en 1814, 1815, 1870, ils portent partout la dévastation et la ruine; ils agissent en maîtres, insultent les gens, et tuent sans pitié ceux qui leur résistent. Il n'est pas de situation plus affreuse.

Le défenseur de la patrie est couronné à l'étranger; on s'honore, tout homme vaillant devient soldat.

Souvent des hommes courageux ont exposé et sacrifié leur vie pour la patrie; parmi elles, je citerai l'héroïque Jeanne d'Arc, que les Anglais brûlèrent vive à Rouen; les femmes de Beauvais, en 1412; celles de Marseille, en 1576. En 1871, un jeune receveur des postes et des télégraphes se fit tuer par les Prussiens.

SUZETTE (Maltresse).



LEAF WRITER

Entites

- TEI
- teiHeader
- text
- body
- div
- div
- pb
- pb
- pb
- pb
- persName
- pb
- objectName
- objectName
- objectName
- objectName
- placeName
- p
- p
- p

Tags

Enfance\_ch01.xml

Selection Image Viewer Validation

1/2

L'ENFANCE DE SUZETTE JE DÉDIE CE LIVRE À MA CHÈRE NIECE MARGUERITE. MARIE ROBERT HALT.

1. — Jolie maison.

Mme Dumay, la maman, achevait de ranger la grande pièce, en ce moment éclairée d'un gai rayon de soleil.

2.

4 L'ENFANCE DE SUZETTE. Les chandeliers, sur la cheminée, semblaient d'or, et la batterie de cuisine, accrochée au mur, paraissait être d'argent, tant tout cela reluisait. Le poêle, ciré à la mine de plomb, reluisait aussi.

Les rideaux, aux vitres, étaient d'un blanc de neige. Sur l'appui des fenêtres il y avait des vases de fleurs. Pas un grain de poussière sur les meubles. Oh! la gentille maison! la plus gentille de Fragicourt; qu'elle était gaie et propre! Oh! la bonne mère que celle à qui l'on devait toute cette minutieuse propreté!

De temps en temps elle se penchait pour regarder, par la fenêtre, dans le jardin où l'on entendait des voix joyeuses d'enfants.

Puis elle appela :

— « Allons, petits, voilà l'heure! » [1]

[1] CAUSERIE : Grammaire. — D'où vient le mot chandelier? Dites les homonymes du mot poêle.

Ménage. — De quoi se compose la batterie de cuisine? — Qu'est-ce que la mine de plomb? — A quel usage sert-elle? — Sous quelle forme les écoliers emploient-ils la mine de plomb? — Pour quoi les chandeliers et la batterie de cuisine reluisent-ils comme de l'or et de l'argent?

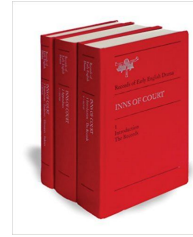
XML:LAND RDF (NO OVERLAP) JSON-LD CUSTOM SCHEMA

LEAF-Writer dev Powered by Tiny

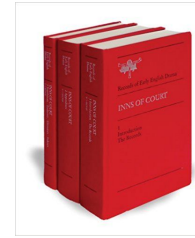


# Evolving Hands Case Study: Records of English Drama collections

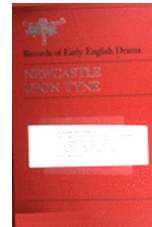
- ❖ 23 Collections
- ❖ 34 Volumes
- ❖ 20,000+ pages of legacy records (now published online)
- ❖ Edited English, Latin, &c.
- ❖ Embedded expanded abbreviations



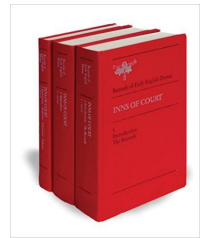
Inns of Court



Ecclesiastical London



Newcastle-Upon-Tyne



Civic London to 1558



f 625 (*Treasurer's payments*)

...

® Allocaciones  
in festo Natalis  
domini

Dictus nunc Computans petit Allocari de x li. solutis  
Thome Hamond senescallo hospicij tempore Natalis  
domini pro expensis sociorum hospicij predicti vt  
patet per warrantum Et de vj li. xiiij s. iiij d. solutis  
eidem Thome pro musica hospicij tempore predicto

5

Summa xvj li. xiiij s. iiij d.

...

10

***Middle Temple Parliament Book*** MT Arch: MT.1/MPA/No. 4  
f 81v (*25 November*)

...

ffestus Natalis

Ordinatum est ad hoc parliamentum quod festum Natalis  
Domini solempniter (non grande) celebrabitur, Et quod  
Communie continuabuntur vsque ad proximum Terminum,  
Et allocatur hic Commorantibus vna biga Carbonum, Et  
quadraginta solidi pro salario Musicorum ante hac consueto  
Et preteria allocatur pro Communijs

15

iiij li.

...

20

***Middle Temple Treasurers' Receipts*** MT Arch: MT.2/TRB/1a  
f 13 (*2 November 1614–24 November 1615*) (*Pensions, fines,  
rents, and ameracements*)

...

25

To mr ffrancis Barker for collecting of iiij li. towardes

```
<!--es fb-->
<xsl:variable name="es">
  <xsl:apply-templates select="$er" mode="replaceStrings">
    <xsl:with-param name="search" select="'fb'"/>
    <xsl:with-param name="replace" select="'es'"/>
  </xsl:apply-templates>
</xsl:variable>
<!--ie ef-->
<xsl:variable name="ie">
  <xsl:apply-templates select="$es" mode="replaceStrings">
    <xsl:with-param name="search" select="'ef'"/>
    <xsl:with-param name="replace" select="'ie'"/>
  </xsl:apply-templates>
</xsl:variable>
<!--is eb-->
<xsl:variable name="is">
  <xsl:apply-templates select="$ie" mode="replaceStrings">
    <xsl:with-param name="search" select="'eb'"/>
    <xsl:with-param name="replace" select="'is'"/>
  </xsl:apply-templates>
</xsl:variable>
<!--ium eYP-->
<xsl:variable name="ium">
  <xsl:apply-templates select="$is" mode="replaceStrings">
    <xsl:with-param name="search" select="'eYP'"/>
    <xsl:with-param name="replace" select="'ium'"/>
  </xsl:apply-templates>
</xsl:variable>
```

# Planned Outputs

- **How-Tos:** multiple documented recipes for small cultural institutions for HTR to enriched TEI conversion of handwritten and structured print
- **Transformation Software:** all XSLT Stylesheets to be openly released
- **Process Documentation:** records of processes to be openly released
- **Case Study Datasets:** data generated from the case studies to be openly released
- **Training Materials:** the training materials for introductory workshops to be openly released
- **Conference Papers:** updates on project process at conferences (like DH2023!)
- **Journal Article:** we plan more formal review of the project in a journal article

# Thank you.

@JamesCummings  
@DianeJakacki  
Alexandra Healey

<https://doi.org/10.5281/zenodo.8135692>

<https://research.ncl.ac.uk/evolvinghands/>

James Cummings, Diane Jakacki,  
Ian Johnson, Carrie Pirmann,  
Alexandra Healey, Valentina Flex,  
Evie Jeffrey



**Newcastle**  
University

**Bucknell**  
UNIVERSITY



**Arts and  
Humanities  
Research Council**



**NATIONAL  
ENDOWMENT  
FOR THE  
HUMANITIES**