



Recognition and Enrichment of Archival Documents

D5.5.

Mobile crowd sourcing tool

Delivery of HTR benefits via the web

Rory McNicholl, ULCC

Distribution:

<http://read.transkribus.eu/>

**READ
H2020 Project 674943**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 674943



Project ref no.	H2020 674943
Project acronym	READ
Project full title	Recognition and Enrichment of Archival Documents
Instrument	H2020-EINFRA-2015-1
Thematic Priority	EINFRA-9-2015 - e-Infrastructures for virtual research environments (VRE)
Start date / duration	01 January 2016 / 42 Months

Distribution	Public
Contractual date of delivery	1/01/2017
Actual date of delivery	31/12/2016
Date of last update	28/12/2016
Deliverable number	D5.5
Deliverable title	Mobile Crowd sourcing tool: Delivery of HTR benefits via the web
Type	Report on demonstrator
Status & version	1
Contributing WP(s)	5
Responsible beneficiary	ULCC
Other contributors	UIBK, NAF
Internal reviewers	Günter Mühlberger, Tobias, Hodel
Author(s)	Rory McNicholl
EC project officer	
Keywords	Handwritten Text Recognition, Web interface, crowd sourcing, mobile

Table of Contents

Executive Summary	3
1. Introduction.....	3
1.1. Web UI foundation.....	3
2 Related applications	4
2.1 Description of Applications	4
3 Mobile crowd sourcing tools.....	10

Executive Summary

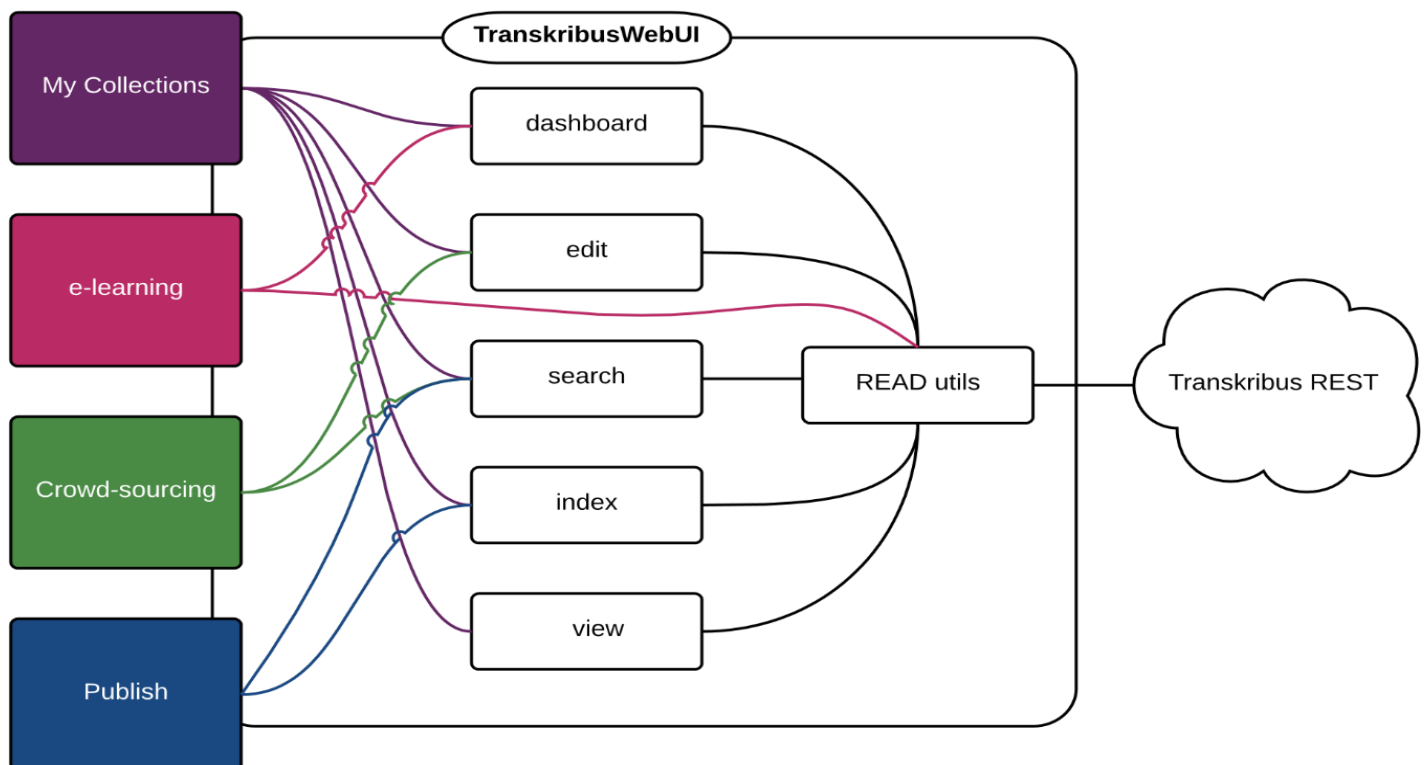
The majority of this document gives an overview of the foundational elements of the web user interfaces in general. Description of this necessary initial phase leads to an outline of the mobile crowd-sourcing tool currently under development.

1. Introduction

Web User Interfaces is a general term used to describe the presentation of the transkribus functionality to people via a web browser as distinct from the Expert tool desktop application.

1.1. Web UI foundation

The approach taken in the first year of the project was to provide a solid shared foundation for all web UI elements. This has been achieved with the use of the [python-django framework](#) which is designed to allow the rapid development of websites (projects) from re-usable functional components (applications). The following diagram shows the various elements and functions of the web UI and how they make use of a shared communication utilities to communicate with the [Transkribus RESTful web service](#).



READ web user interfaces overview

2 Related applications

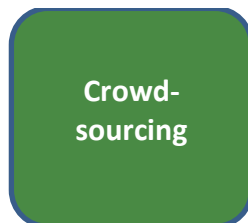
The four central elements of the web UI are described below. These can be considered as separate interfaces with distinct purposes and audiences. They each make use of different functions of the overall web UI. This approach has been mapped to the Django concept of projects and applications, where applications are standalone programs that perform a given function and can be combined with other applications in a given project. So the four elements below represent a Django project that delivers a website with a distinct purpose and audience and further mentions of “projects” is in reference to these rather than the project as a whole.



An interface for the collection owners and editors to view statistics on their collections and the use thereof. Also provides an editing interfaces for transcription, proof reading and some administrative actions.



Present fragments of image/texts with missing words for the purpose of transcription practice. Provide performance feedback to the user and potential for threshold for admission to crowd-sourcing projects.



An interface that allows registered users access to selected collections to use editing interfaces for transcription and metadata tagging.



Present to the general public the “finished product” from crowd-sourcing or institutional enhancement projects - fully transcribed tagged documents.

[READ web user interface projects](#)

Each of the web UI projects has been designed to work well on a range of devices including mobiles. This has been achieved by using the responsive bootstrap framework throughout and testing the interface with [Google’s “Mobile Friendly Websites” tool](#) and in more depth with [mobiReady](#).

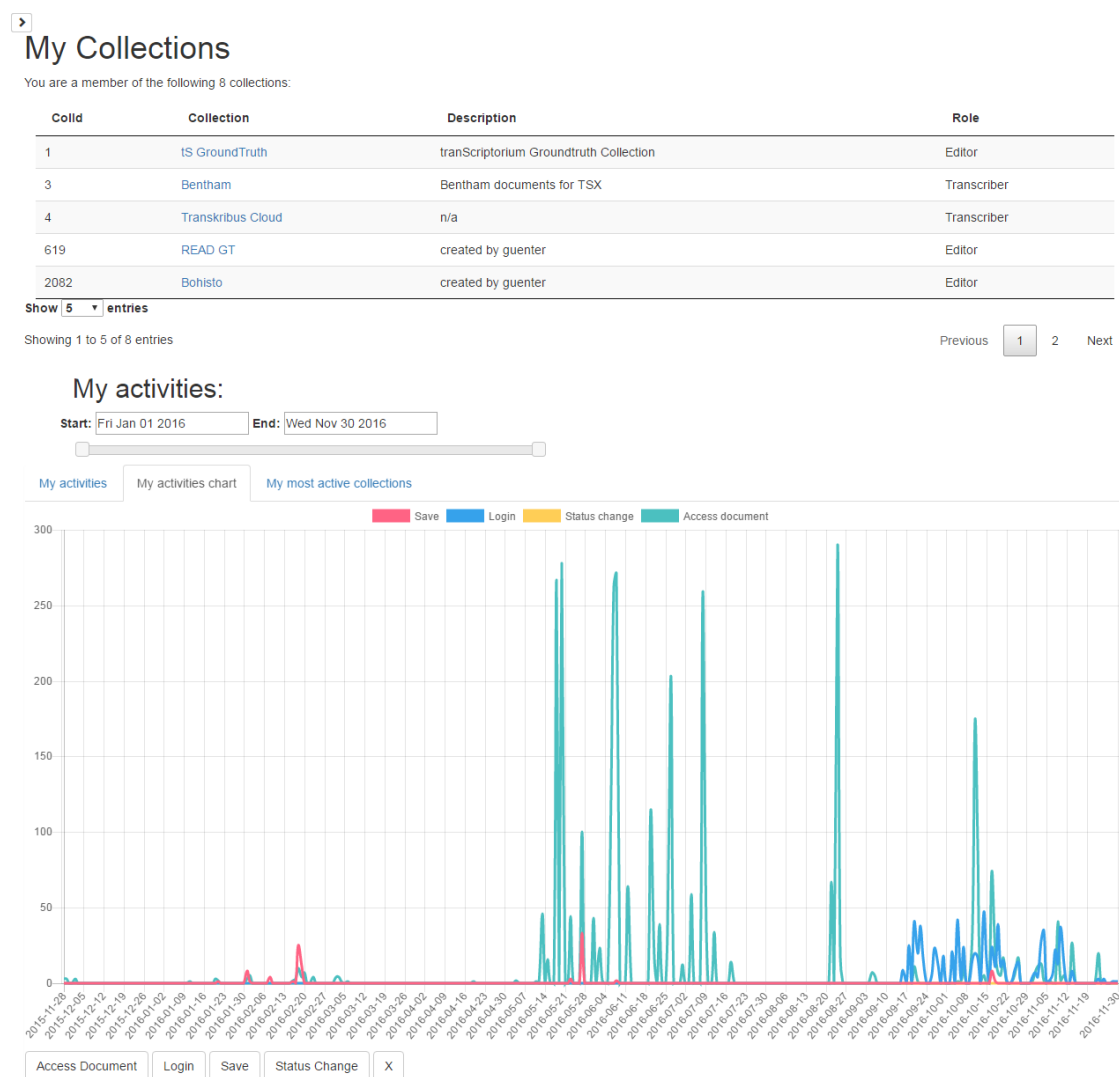
2.1 Description of Applications

Applications in this sense refer to the Django applications that perform a given task and represent reusable building blocks of the projects. The applications are designed to respond

to context such as the type of user accessing them and other parameters to provide the user experience warranted by that project. So an administrator accessing the dashboard via the “My Collections” project will have a difference experience compared to a public user accessing via a crowd-sourcing project that is constrained to a single collection.

Dashboard

[The dashboard application](#) brings together and refactors data on users, activity and page status to provide an overview for collection owners and editors. For individual users the dashboard will provide a profile page collating data for that specific user within a collection and the data on their interactions with the documents.



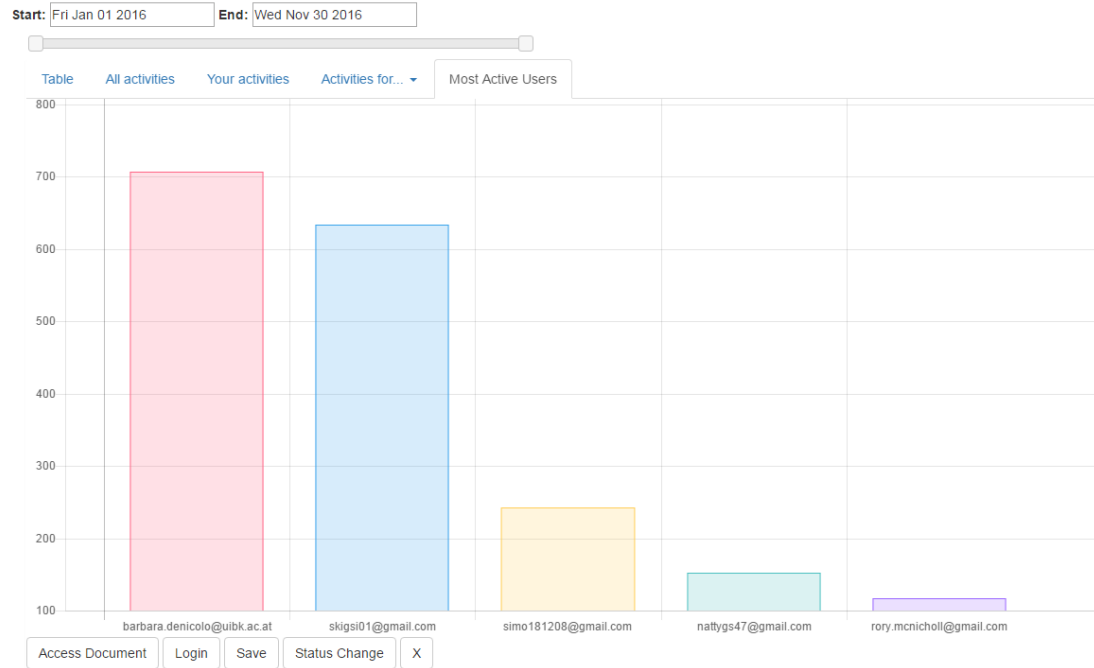


HS_115

Last activity HS_115

: User , , operation () on page ID of document ID

Activities for HS_115



Show 12



Showing 0 to 12 of 574

Previous 1 2 3 4 5 ... 48 Next

Edit

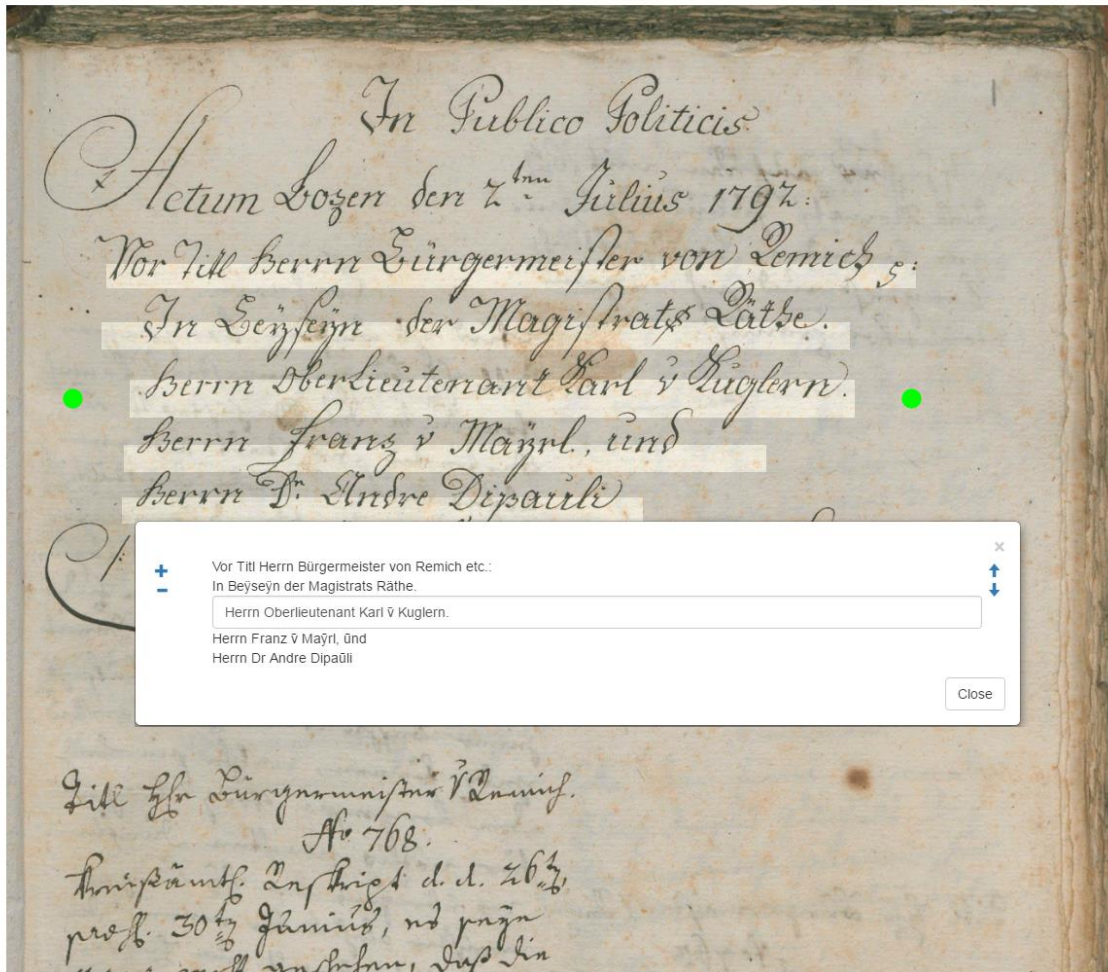
The Edit application presents the page images for transcription and enhancement. Different views are available so that the most appropriate can be used for a given task - such as transcription, tagging metadata proofreading and correction. There is no prescription of view to task, this can be the choice of the collection owner at deployment.



Correction Interface, Line by Line (working title)



Save changes



In-line/modal transcription editing, aka line by line

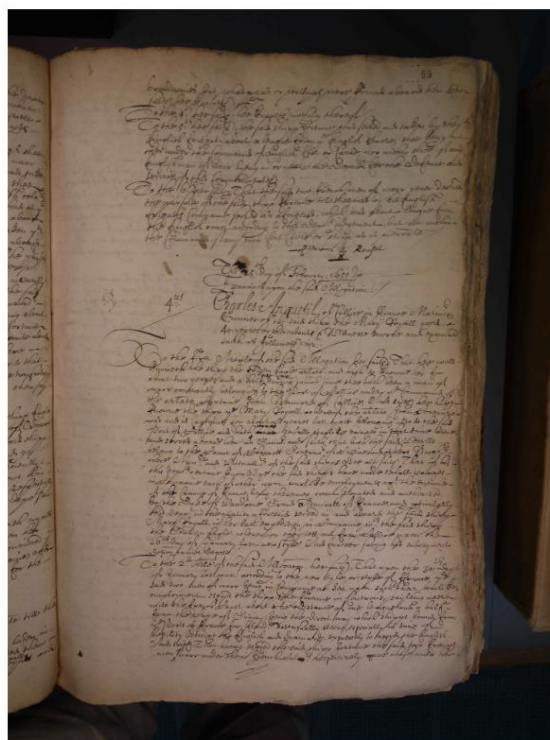
Proofreading Interface (working title)

● Original version

☒ Your version

Save changes

99
brandewines, but what paps or writings were found aboard her heed
saith hee knoweth not./
To the 4.th hee saith, hee knoweth nothing thereof./
To the 5.th hee saith, the said shipp fortune was seized and taken by the
said
English frigat abt a league from the English shoare, shee being note
then under the Command of English fort or Castle, nor within shot of any
Englismen of warr lying in or neere the downes for the defense And
service of this Commonweath./
To the 6.th hee saith, That the said two french men of warr never deserted
the pursue of the said shipp ffortune, till shee was by the English
frigatts Companie seized as aforesaid, which was about a League from
the English coast according to this depots judgement, but not within
the Command of any such fort Castle or shipp as is interate./
Piere, 4 Rouse!
The 25.th day of January 1655 This is a changed piece of text
Examined upon the said Allegation :-/
Rp. 4.: us
Charles Anquestil of Callice in ffrance Mariner
&.&.&.&.&.&.;&.
Gunner of the said shipp the Mary Royall aged xxx
40; yeares or thereabouts a Wittnesse sworne and examined
saith as followeth vizt.-/
To the first Article of the said Allegation hee saith, That hee well
knoweth the shipp the golden Eagle arlate, and hath so knowne her for
last past
about two yeares and a half during which time shee hath been a man of
warr constantly belonging to the Port of Callice under the Command of
the arlate Capitaine John Bennard of Callice, And saith hee hath
knowne the ship the Mary Royall as hereof the arlate John Coverue
was and is Captaine for about 8. yeares last past belongin also to the said
xxx
Port of Callice and hath xxx sewall warlike voages in that time benee
and served aboard her as Gunner, and saith that both the said Vessells
belong to the Count of Charrott Governour of Callice And others Burgheers
there, Arm.eers and Victual.eers of the said shipp. hee also saith That of his
this depots certaine knowledge the said ships have made sewall voages
in Companie each of other upon warlike employments ag.t the enemies
of the Crowne of france, being thereunto commissionated and authorized
by the Duke of Vandome Grand Admirall of ffrance, and particularly
this depot: in the quality aforesaid served in and aboard the said shipp
Mary Royall in her last expedition in Companie w.th the said shipp
the Goulden Eagle, whereupon they sett out from Callice upon the
26.th day of January Last new style, And further saving his subsequent
depon cannot depose./
To the 2.d Arle of the said Allegation hee saith, That upon the 30.th day
of January last past according to the new style or style of france, the
said two men of warr being in Companie at Sea upon such their warlike
employment espyed the ship the fortune in Controvys sayeing neere
unto the french Coast about the Distance of one league and a half
from the Cape of Grinlay, being the direct way which shippis bound from
the Ports of France for fllanders doe usually steere, especially in times of
hostility betwixt the English and Spaniards, expressly to avoyd the English
And saith, That having espied the said shipp fortune the said two french
men of warr under their french Colours respectively gave chase unto her



READ - A horizon 2020 e-infrastructure project



Figure "Side-by-side" transcription editing

Search

The search function of the transkribus web service based on [apache solr](#) became available earlier this year. Presently full-text searches for documents is possible via this API with metadata searches are due to follow. Discovery of documents and the text within documents, as well as faceted search and metadata indexing of content will be added to most of the web interface elements in the coming year (eg my collections, crowd-sourcing and publish/view).

Index

A section of the web-UI currently known as the library Renders of indexes for collections, documents, pages, transcripts, also regions and lines. The library is a demonstrator of the indexing capabilities that will be used to some degree in several of the web-UI elements.

This document has 574 pages:

ID	Page number	Image file name	Transcripts
113223	1	HS_115_1_object_128269.jpg	15
113230	2	HS_115_2_object_128270.jpg	14
113237	3	HS_115_3_object_128271.jpg	24
113244	4	HS_115_4_object_128272.jpg	15
113251	5	HS_115_5_object_128273.jpg	18
113259	6	HS_115_6_object_128274.jpg	22
113267	7	HS_115_7_object_128275.jpg	15
113275	8	HS_115_8_object_128276.jpg	15
113283	9	HS_115_9_object_128277.jpg	10
113291	10	HS_115_10_object_128278.jpg	14
113299	11	HS_115_11_object_128279.jpg	15
113306	12	HS_115_12_object_128280.jpg	15
113314	13	HS_115_13_object_128281.jpg	12
113322	14	HS_115_14_object_128282.jpg	7
113330	15	HS_115_15_object_128283.jpg	14
113338	16	HS_115_16_object_128284.jpg	9
113346	17	HS_115_17_object_128285.jpg	8
113354	18	HS_115_18_object_128286.jpg	11
113362	19	HS_115_19_object_128287.jpg	11
113370	20	HS_115_20_object_128288.jpg	10
113378	21	HS_115_21_object_128289.jpg	9
113386	22	HS_115_22_object_128290.jpg	8
113394	23	HS_115_23_object_128291.jpg	11
113402	24	HS_115_24_object_128292.jpg	7
113410	25	HS_115_25_object_128293.jpg	7
113418	26	HS_115_26_object_128294.jpg	5
113426	27	HS_115_27_object_128295.jpg	9
113434	28	HS_115_28_object_128296.jpg	6
113442	29	HS_115_29_object_128297.jpg	6
113450	30	HS_115_30_object_128298.jpg	2

View

A view of transcript and/or image with visual indications of enhancements on image and transcript. Views will show region and line data where appropriate on image and transcript. Styling of transcript text as appropriate and highlighting of metadata in text with linking via metadata searches between pages, documents and where appropriate between collections. This application will be the focus of the publish element of the web-UIs.

3 Mobile crowd sourcing tools

Although all projects have been designed with mobile in mind certain projects have a particular focus on delivering transkribus functionality to hand held devices and as such have been conceived as simple, small and involve tasks that can be undertaken over small periods of time. One such of these takes some of the smaller tasks suitable for crowd sourcing and delivers these in a stripped down, simple way.

Building on the applications already deployed (mainly edit and read utilities) and currently in development is a tool to present words that have been marked in a transcript with the TEI unclear tag <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-unclear.html>. These words will be presented for correction to general users within a minimal (but expandable) context.