

Guide to Encoding

	See also the document outline to the left for clickable links to different subsections of this document.
---	--

Encoding inscriptions for the IIP project requires that you become familiar with several aspects of the process. This document describes the process of entering and editing inscriptions in machine readable form so they can be displayed, shared and analyzed.

There are three parts (like Gaul) to an inscription encoded using EpiDoc.

1. **Metadata:** information about the source, the material support and provenance is called metadata. It is encoded in the `<teiHeader>` section.
2. **Facsimile:** Information about any image of the inscription is encoded in the `<facsimile>` section.
3. **Text:** The text of the inscription, translation and commentary are encoded in the `<text>` section.

Note that if you are an encoder, before getting started here you should have already proceeded through the below documentation:

- [Installing and Setting Up Oxygen XML](#)
- [Installing Keyboards](#) (if you are working with a non-Latin language)
- [What is an Inscription and How are they Published?](#)
- [Introduction to XML, TEI, and EpiDoc](#)
- [Overview of Encoding Workflow](#) (the section "If you are an encoder")

0. File Naming

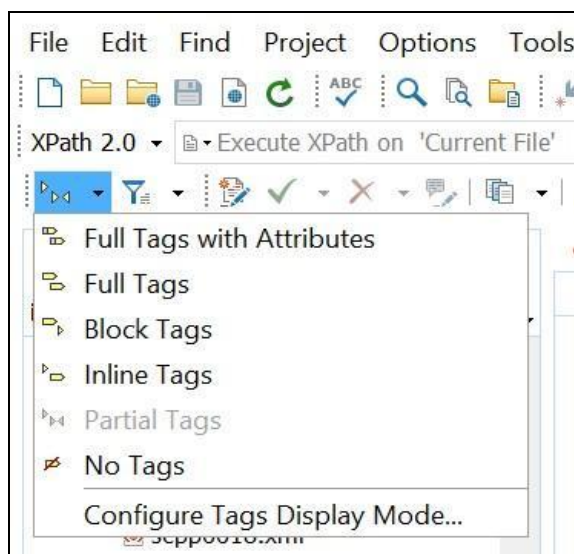
IIP uses a set of fixed file naming conventions - each inscription is named according to its place of origin, and then numbered in sequence. Filenames take the form `aaaa0000.xml` where "a" represents an alphabetic character and "0" represents a numeral. Please make sure that you are using the correct alphabetic prefix, and ask if you aren't sure. Examples are `zoor0234.xml`,

masa0035.xml. Occasionally there is an alphabetic character following the number if there is more than one text on the same support. For example, masa0095a.xml and masa0095b.xml. The filename, without the .xml suffix, is also the IIP identifier that is uniquely assigned to an inscription. More information about this is below in "Identifiers."

1. Metadata Section (in theteiHeader)

When you first open the IIP template in Author Mode, you will see some shaded areas and many red lines under the text and on the right side vertical bar. The lines indicate that the underlying XML is not (yet) correct. As you start adding information, the red lines will disappear.

At the top of the Author screen, you will see a variety of icons. You will be able to work most easily in Author Mode if you set the interface to display partial tags or no tags at all. To do this, click on the tags icon (the second icon from the left on the icon bar; with thick arrows), hold down, and select either "Partial Tags" or "No Tags" from the menu that appears.



Identifiers

- 1) Enter the **machine readable inscription name** in the box next to the title. The machine readable ID consists of **four lowercase letters** indicating the location of the inscription, **followed by 4 digits for the inscription number**. Pad with leading zeros if necessary. For example: zoor was used for Zoora, and 0002 would indicate inscription

number 2 – zoor0002 is the machine readable ID. It will also be the file name (i.e., you will be saving the file as: zoor0002.xml)

Inscriptions of Israel/Palestine: zoor0002

- 2) Enter the **Display ID** in the IIP Identifier box. The display ID is the more human readable version of the machine readable ID. **Capitalize the first letter, and put a space between the place and the number.** Ex: Zoor 0002 .

IIP Identifier

Display ID (Ex: Beth 0123): Zoor 0002

Summary Box

As we have seen, each inscription file has three major parts. The first part is information about the inscription, called metadata. In Text Mode, you can see that the metadata goes up to line 137. Most of the boxes in the Oxygen Author mode deal with the metadata.

Summary

Select Main Language: Greek Select Other Language:

Genre (Click blue text; check one or more boxes from list.): #funerary.epitaph

Religion (Click blue text; check one or more boxes from list.): Christian

Summary description of form (Place, Date. Object Description. Type of Inscription.):

Zoora, April 7, 402 CE. Tombstone. Funerary (Epitaph).

- Language

In the Summary Box, enter the language (or languages) of the inscription, using the pull-down menu. The present choices are Aramaic, Greek, Hebrew, and Latin. If you come across an inscription in another language, please bring it to our attention (we follow ISO 639-2 codes found here: [ISO 639 Code Tables | ISO 639-3](#))!

We have space for two languages in the template. If you have a bilingual inscription, the “Main Language” will be the language that appears first. Use “Other Language” for the language that appears second.

If you need to do something that the template does not accommodate (e.g., add a third language or delete a language if you made a mistake), go into Text mode in Oxygen; find the tag `<textLang>`, and make the adjustment there. You can, for example, simply add another `otherLangs=""` before the `/`, or delete the one that is there. There must always be a `mainLang`. Note that you will need to manually type in the code for some languages.

`xcl` = Armenian, `syc` = syriac, `geo` = Georgian, `xxx` = other.

```
<msContents>
  <textLang mainLang="" otherLangs=""/>
  <msItem class="" ana="">
    <p></p>
  </msItem>
</msContents>
```

- Genre

Here you will indicate what kind of text the inscription is, choosing from options available through the drop down menu. We are continuously working on developing a more comprehensive controlled vocabulary list for this field. See the current list in the [Controlled Vocabularies](#) documentation.

- Religion

We use the category of “religion” with ambivalence, because it is familiar and can be useful to scholars: we attribute a “religion” to inscriptions which display familiar religious symbols or other indications (e.g. language, text content, etc.) of belonging to either a Jewish or a Christian community. Instead of attempting to create a taxonomy of different Greek and Roman religious groups, we label those together as “Other.”

- Summary description

(in the `<p>` tag within `<msItem>` element, if you are working in Text Mode.) The Summary description contains essential information about the inscription. Note that the information in this description is also found in other parts of the xml file – in those places it is indexed and searched, not here.

The format of the summary description is as follows:

Place, Date. Object Description. Type of inscription.

Example: Zoora, April 4, 417 CE. Tombstone. Epitaph.

Please make sure to insert here the same values you choose from the relevant dropdown menus, that is: the value of "Place" will be the same as "City" within the "Provenance" box (below); Object Description will be the same as "Type of Object" in the "Physical Description" box (below); and the "Type of Inscription" will be the same as "Genre" in the "Summary" box (above).

Physical Description Box

The Physical description `<physDesc>` focuses on the inscription as an object, not on its text content. Much of the metadata is contained within this tag.

- Type of Object (`<objectDesc>`)

It is chosen from a pull-down menu. Note that this is not the type or genre of inscription, but of the object upon which it is inscribed (e.g. ossuary). These values are not a closed list; new values can be added as needed by the project manager(s). Let us know in case you encounter an object which does not appear within the current list. See the current list in the [Controlled Vocabularies](#) documentation.

Physical Description

Type of Object (Click blue text; check one or more boxes.): [Empty]

Material (Click blue text; check one or more boxes.): [Empty]

More detail if necessary:

Dimensions (Unit=cm)

Height (Top to Bottom):

Width (Left to Right Across):

Depth (Front to Back):

☐ Altar
☐ Amphora
☐ Amulet
☐ Arch
☐ Architectural Fragment
☐ Architrave
☐ Arcosolium
☐ Block
☐ Boundary Marker
☐ Bowl
☐ Building Stone
☐ Capital

- Material (<supportDesc> within <objectDesc>).

This pull-down menu operates much as the one above, with a two-level hierarchical controlled vocabulary. These values are not a closed list; new values can be added as needed by the project manager(s).

Physical Description

Type of Object (Click blue text; check one or more boxes.): [Empty]

Material (Click blue text; check one or more boxes.): [Empty]

More detail if necessary:

Dimensions (Unit=cm)

Height (Top to Bottom):

Width (Left to Right Across):

☐ Stone
☐ Marble
☐ Limestone
☐ Kurkar
☐ Granite
☐ Tufa
☐ Other Stone
☐ Unknown Stone
☐ Clay
☐ Metal
☐ Plaster
☐ Glass

- More Detail (<p> within <support>)

Any prose that is relevant to the type of object or its material. This field is not currently displayed. We do not use it often, but feel free to add relevant information if there is something exceptional about the object that it is helpful to share internally. Please comment only on the object and its material.

- Dimensions (<dimensions>)

All measurements for the three dimensions displayed (height, width and depth should be in **centimeters**. If no dimensions are indicated leave the field(s) blank. If the printed edition uses a different unit (e.g. millimeters), make sure to convert it into centimeters. If the object has dimensions that vary along a single side, or is uneven, enter the minimum and maximum separated by a dash.

Example: <height>22-27</height> **Example:** <width>18</width>

Dimensions (Unit=cm)

Height (Top to Bottom):

Width (Left to Right Across):

Depth (Front to Back):

- Condition (<condition>)

Indicate the condition of the inscription by selecting a value from the pull-down menu.

Condition:

More detail:

- Complete
- Complete but broken
- Complete and intact
- Single Fragment
- Multiple Fragments
- Contiguous Fragments
- Non-Contiguous Fragments

Complete

> The object is complete (generally)

Complete.broken

> We have all the parts, but the object is broken

Complete.intact	> The object is complete and intact
Fragment.single	> We have a single fragment
Fragments	> We have a number of fragments
Fragments.contig	> We have a number of fragments, and they are contiguous
Fragments.non_contig	> We have a number of fragments, and they are not contiguous

- More detail (<p> in <condition>)

Prose details relating to the condition of the inscription. This field is not currently displayed.

Condition: Single Fragment

More detail if necessary: Fragment of a gray marble plaque, broken on all sides. Only the fa

Layout of Writing on the Surface

- Columns and Written Lines (attributes in <layout> in <layoutDesc>)

Enter the number of columns and lines of the inscription.

- More detail (<p> in <layout>)

Prose details relating to the layout of the inscription. This field is not currently displayed.

Layout of Writing on the Surface

Columns: 1 Written Lines: 4

More detail if necessary: Traces of guidelines remain, and leaves divide words

- How was this written (<handnote> within <handDesc>)

Select from the pull-down menu indicating the means by which the inscription was made, e.g., chiseled or inset in a mosaic. Most inscriptions on stone or on ostraka are either inscribed or painted.

- Average dimensions of the letters (another `<handnote>` within `<handDesc>`)

Enter the average height of the letters, as **a range of cm values** in the "At Least" and "At Most" fields. If there is no variation, enter a single value in the "Quantity" field.

- More detail (`<p>` in `<handnote>`)

Prose description of any additional details relating to the writing. This field is not currently displayed.

Writing

How was this written: #engraved ▼

Average dimensions of the letters in cm. At least: 3 At most: 4 Quantity (alt. to above):

Remove Letter Dimensions

More detail if necessary: Irregular letters, with serifs, poorly cut with a point

Decorations (`<decoDesc>`)

As with `<handDesc>`, each decoration is contained within a `<decoNote>`. The controlled vocabulary for this field is in development. Describe each decoration in prose the best you can (this goes in the `<ab>` element within the `<decoNote>`). Then, also in prose, describe where the decoration on the object is (`<locus>` within the `<DecoNote>`). If there is more than one decoration, click the "New Decoration" button and enter the information. Repeat as necessary.

Decorations:	
Decoration	
Brief description:	Palm branches
Location of decoration:	Beginning (first line, upper left corner) and end (third line, bottom
New Decoration	

History and Provenance

This section describes the history of the inscription.

- Is this a fake?

Except in very rare circumstances, the value will remain "Genuine."

If you look at the XML, you will see an empty `<rs>` element if the inscription is genuine, and you will see `<rs>Fake</rs>` if the inscription is not genuine.

- Place and Date of Origin (`<origin>`)

These fields apply to the place and date where and when the inscription was first made displayed and made. Sometimes there is doubt about this, which is indicated primarily in the prose description.

- *Dates: NotBefore and NotAfter* (attributes in `<date>` element).
These values will not be visible to users, but will be used for finding and sorting. Please use 4 digit dates (300 will be entered as 0300). If the date is BCE, then it will have a minus sign. Most dates are approximate, so if something is in the first half of the second century BCE then you will enter `notBefore="-0200"` and `notAfter="-0150"`. If the date is exact, then both values will be the same.
- *Display Date* (prose in `<date>` element)
Enter the date as it should display in the Summary Box. **Example:** "First half of the Second century BCE".
- *Region, etc.* (`<placeName>`)

Region:

City:

Pleiades ID:

If there is no Pleiades ID:

This indicates the original site of the inscription, to the degree known, in a hierarchical fashion moving from broadest, region (<region>), to the city (<settlement>), to two progressively more specific locations (<geogName> and <geogFeat>) that indicate the site and the locus in case we have them.

The region is usually one of a few regions common in the IIP corpus, and so you should select the appropriate one from the drop-down. In unusual cases where a different region is necessary, you can switch into text mode and type the appropriate region into the XML in the text of the <region> element.

The spelling of place names generally follows those in Yoram Tsafrir, et al., *Tabula Imperii Romani: Iudaea/Palaestina* (Jerusalem: The Israel Academy of Sciences and Humanities, 1994).

We link our <settlement> to the Pleiades Gazetteer (<https://pleiades.stoa.org/>) or add geographical coordinates in case the Gazetteer lacks an ID for a particular place.

In order to insert the relevant Pleiades ID, go to the Gazetteer website and look for the place the inscription is from. If you search for Masada, for instance, you will see that the corresponding ID is 687968. Copy and paste the number at the end of the string within the Pleiades ID box, as follows:

<https://pleiades.stoa.org/places/687968>

If we do not have a Pleiades ID, click the 'Insert Coordinates' button to replace the Pleiades field with a field in which you can input the latitude and longitude instead.

City:

Coordinates:

If there is no Pleiades ID:

- *Current Location* (<placeName> in <provenance>)

In prose, whatever is known about the current location of the inscription. If the object is in its original location, use "in situ". If the inventory number is available, include it.

Current Location (if different from origin)

Ralli Museum, Caesarea, IAA Inv. no. 2005-772 (Once Archeologic) ▶ date ↵

Revision Log

It provides an internal history of who did what to the inscription. Each entry creates a new `<revisionDesc>` section. Add some brief information indicating what you changed, or that you created the file. You can add more lines to the revision log by clicking the "Add Revision" button.

Revision Log

When: 2007-03-05 Who persons.xml#Eunice_Hong

Type of change (supply additional info if needed): Initial Entry ▼

When: 2016-12-14 Who persons.xml#Elli_Mylonas

Type of change (supply additional info if needed): Jing Pleiades IDs to origin/placenames ▼

When: 2017-12-20 Who Gaia Lembi

Type of change (supply additional info if needed): Edited letters height ▼

When: 2019-01-29 Who persons.xml#Elli_Mylonas

Type of change (supply additional info if needed): ▶ to date element, with Periodo value. ▼

Add Revision

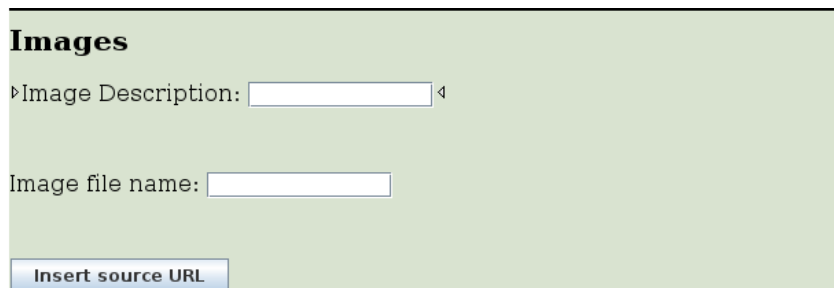
2. Facsimile Section

Images

We are in the process of acquiring images of the inscriptions in our corpus. Before they are associated with an inscription and added to the IIP website, images are stored in the "Images Admin" folder on Drive. Once it has been renamed and referenced in an IIP xml file, an image will be moved to the "image-git" folder on Dropbox.

If the "image-git" folder is getting too big and synchronizing it is unwieldy, it is possible to add a file directly to the GitHub repository by opening it in a browser window and dropping the file on the window. You will have to be logged into your GitHub account in order to do this.

To associate an image with an inscription, type the image description into the indicated field: this places the description into the `<desc>` element within the `<surface>` element. Paste the image file name into its appropriate field; in the XML, this edits the `url` attribute of the `<graphic>` element.

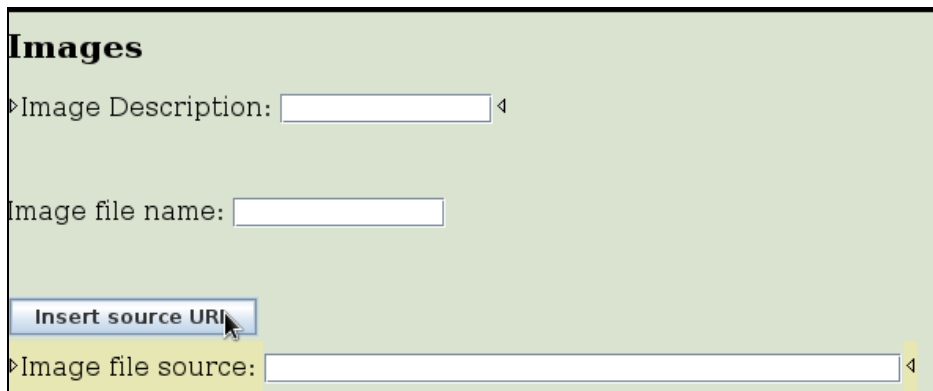


By default the form in author mode only allows for adding one image. If you would like to add another, it is best to do that in text mode. Here is an example for jaff0049, after switching to text mode:

```
<facsimile>
  <surface>
    <desc>Inscription front side. <persName role="Credit">University of Oslo, Institute of Classical Archaeology</persName></desc>
    <graphic url="jaff0049.jpg"/>
  </surface>
  <surface>
    <desc>Inscription back side. <persName role="Credit">University of Oslo, Institute of Classical Archaeology</persName></desc>
    <graphic url="jaff0049b.jpg"/>
  </surface>
</facsimile>
```

In this case, we have two images: one of the front side and one of the back. The `<surface>` tag contains both a brief description of the image (e.g. "Inscription front side") and the credits, marked within the `<persName role="Credit">` tag, as well as the reference to the .jpg file (`<graphic url="xyz.jpg">`). In most cases, it is sufficient for the `<desc>` to simply include the image credit.

If the image for an inscription happens to be sourced from a public domain image available on the internet, click "Insert source URL" to get a field to record the URL of the original image file.



The screenshot shows a form titled "Images" with a light green background. It contains three text input fields: "Image Description:", "Image file name:", and "Image file source:". A blue button labeled "Insert source URL" is located between the "Image file name:" and "Image file source:" fields. The "Image file source:" field is highlighted with a yellow background.

3. Text Section

Bibliography

Although this comes at the end of the template, (`<div type= "bibliography"></div>`), when entering inscriptions it is best to enter bibliography at this point (i.d., before proceeding to encode the text). The full bibliography (corresponding to the source you are using for metadata, transcription and translation) should be entered. We store our bibliographical references in Zotero (www.zotero.org), where you can check existing bibliography as well as add new items if needed. Follow these steps to do so:

1. In your browser, go to Zotero and log in using the following: **username:** iip (all lowercase); **password:** Maximian (capital M). Once in, **make sure to use our Group Library and disregard any other folder.**

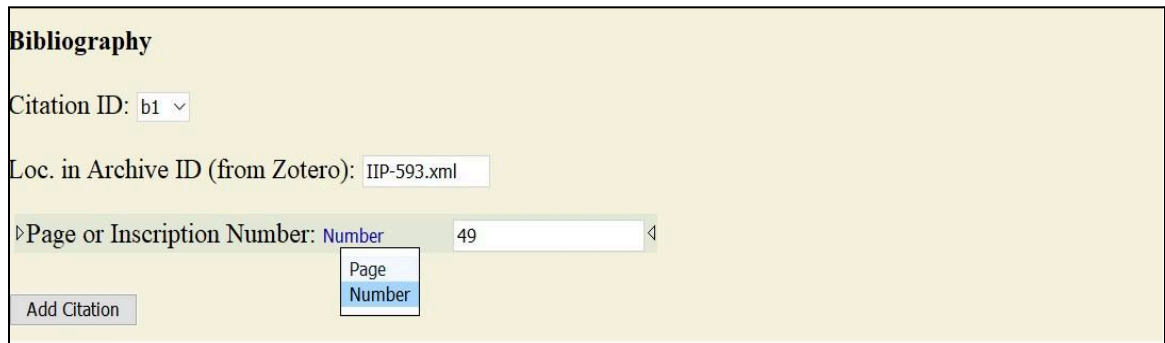


2. Search for your citation(s).

- a. If the record is present, record the value in "Loc. In Archive" (**IIP-xxx** or **IIP-xxxx**) OR the corresponding tag (they must be identical) and see below.
- b. If we do not have a record for that citation, click on the + sign to add a new record. Then follow instructions to add a new citation. Remember to record the "Loc. In Archive" and the Tag in order to create an ID specific to that item.



- Now, knowing the "Loc. In Archive"/Tag value, go back to the entry form that you are working on in Oxygen. Be sure to add ".xml" after the Location in Archive ID. Designate a unique Citation ID for each citation (b1, b2, b3); enter the Loc. In Archive ID and the appropriate page or item number (checking which value you are entering).



Bibliography

Citation ID: b1 ▾

Loc. in Archive ID (from Zotero): IIP-593.xml

Page or Inscription Number: Number 49 ▾

Add Citation

Page
Number

- For the case of titles in languages that use Latin letters, use the original title. In the case of non-Latin titles (e.g., Hebrew or Greek) use the English title that is indicated in the publication with a notation of language in the appropriate field. If there is no English translation, use the original script to copy the title.

If you have more than 3 references, you can type a number (b4, b5, etc.) into the box where the pulldown value displays.

Inscription Text

Inscriptions on stone and other objects are often in poor condition, letters are missing, or difficult to read. There are lots of abbreviations and spelling is not always as good as it should be. Epigraphers have been using special notations for a long time to indicate where lines or letters are missing, where they are conjecturing what might have been there, to fill out abbreviations and correct spelling. In a printed edition, all this takes the form of () {}, underdots and so on.

Make yourself familiar with the Leiden conventions, so to be able to understand what parentheses and other signs mean in a printed edition:

<https://academic.oup.com/edited-volume/34560/chapter/293267000>.

This is one of the most commonly used conventions, but you will find that not all editors use the same symbols to mean the same thing, especially when it comes to Hebrew and Aramaic

inscriptions. The EpiDoc encoding that we use is a machine readable translation of these conventions, with more information added, as well.

N.B. For more detailed and thorough information, please refer to the following links:

- 1) <http://www.stoa.org/epidoc/gl/latest/> (The website which contains all the information about how to use the EpiDoc markup language)
- 2) <http://www.stoa.org/epidoc/gl/latest/app-alltrans.html> (Transcription guidelines which provide detailed information for encoding the text of an inscription)

Use these links for any further doubt you may have as far as encoding is concerned: if they do not help with the issue you are dealing with or if you are unsure, ask!

The `<text>` element contains three sections, each labeled by a `<div>`. They are:

1. The *diplomatic transcription* is where the inscription is transcribed as written (i.e., there is no separation between words; all is in capital letters; no accents). These are tagged only for features relating to the appearance of the text. For example, we would mark gaps, uncertain readings, deletions (or interlinear additions) original to the text. We do not note abbreviations, supplied letters or words, alternative readings and the like.

E.g. `<div type="edition" ana="#b1" subtype="diplomatic" xml:id="sepp0014.diplomatic" corresp="#sepp0014.transcription #sepp0014.translation"><p>TEXT HERE</p></div>`

2. The *edited transcription* is the inscription as rendered by an editor. Here we mark whatever the editor has done, such as filling in abbreviations, noting original spelling mistakes or missing or extraneous words, etc. Almost all published inscriptions typographically render their texts according to something known as the Leiden conventions (see: http://papyri.info/docs/leiden_plus), in which editorial interventions are represented with various punctuation symbols. We convert these typographical symbols into their equivalent EpiDoc element in the inscription source file, and then render them in Leiden notation when we display them. For example, a printed version might say: A[B]C (meaning that the letter B was added by the editor because it was lost). We transcribe that as follows: A<supplied reason="lost">B</supplied>C. The markup allows us to do far more than just displaying the texts, however.

E.g.: <div type="edition" ana="#b1" subtype="transcription"
xml:id="sepp0014.transcription" corresp="#sepp0014.diplomatic
#sepp0014.translation"><p>TEXT HERE</p></div>

3. The *translation* is, of course, simply the English translation. These are typically provided by the editor, with the typographical symbols as written. **Translations are not generally tagged, except for line breaks where noted.**

E.g.: <div type="translation" ana="#b1"
xml:id="sepp0014.translation" corresp="#sepp0014.transcription
#sepp0014.diplomatic"><p>TEXT HERE</p></div>

Note that it is entirely possible that one or more of these items do not appear in the printed source: publications of Hebrew inscriptions, for example, often contain an edited transcription but not a diplomatic one. Just include what is in the source.

Note that IIP is careful to mark the source of the diplomatic transcription, edited transcription, translation, and commentary, which is usually the same. Occasionally, though, a translation may come from one article, and a transcription from another. In either case, as you start to encode part of an inscription, you need to note its source by selecting a value in the *Reference Citation* pull-down menu. This menu will display existing b1, b2, etc. values which refer to the bibliographic citations you entered in the Bibliography. Often, there will only be one citation, and all information will derive from it. In that case, the citation will be "b1," and all Reference Citations will be "b1." Clicking on the default b1 will bring up a checklist of all the citations you have entered for this inscription. Check one or more.

Up until now, you have been entering information in Oxygen's Author mode, using form boxes, pull down menus and checklists. You can continue to use Author mode to enter inscription text, and, indeed, you have to use this mode for Hebrew and Aramaic. To enter transcriptions, turn on Full Tags in Oxygen. You can also enter Greek and Latin texts, and do some spot checking of Hebrew and Aramaic using Text Mode, which shows you the EpiDoc elements directly. This can be very useful if you have an error or if you are trying to figure out a complex structure.

The process for entering right to left (e.g., Hebrew and Aramaic) inscriptions and Greek and Latin ones are slightly different. When you start a new file, the template will default to

presenting text as if it were left-to-right. But when entering the texts of the inscriptions, you will be presented with a choice as follows:

Texts, Translations, Notes

If text is Hebrew, Aramaic, or another right-to-left language:

To undo, click:

Diplomatic Transcription (Click to select citation reference): [b1](#)

Edited Transcription (Click to select citation reference): [b1](#)

Clicking the “Language is RTL” button will make the changes necessary for the XML file to display right-to-left text correctly. You should also see the “text here” in the Diplomatic and Edited Transcription boxes move to the right margin after clicking this button. If for some reason you clicked the “Language is RTL” button by mistake, you can reverse this by clicking “Language is LTR.”

Hebrew and Aramaic inscriptions are easier to enter working in Author Mode. When a tag is necessary, press “Enter” and select from the available tags. Some tags require attributes – for these elements select them, press alt-enter (Windows) or option-enter (Mac), and then enter the attribute value (a pulldown list). You may use the **Epidoc Cheat Sheet** for quick reference:

<https://svn.code.sf.net/p/epidoc/code/trunk/guidelines/msword/cheatsheet.pdf>

Please make sure to indicate the correct language in the metadata section on Language, i.e. Hebrew OR Aramaic (or both in case of a bilingual inscription).

For Greek: To enter Greek, change the keyboard layout to Greek. Accented characters can be accessed via the Edit Menu > Insert from Character Map. An insert window appears. Select Arial Unicode MS. You can then scroll down to the accented Greek characters, highlight them, and press Insert.

If you only have to enter a short amount of Greek and or if you have no other way of doing it, you can type the transcription in a Word document (in Unicode Greek) and then paste it into the appropriate box in Oxygen. The downside of this approach is that one must then enter the tags around the words.

A few notes about special characters in Greek (N.B. The Unicode character codes are indicated here for your own convenience, so that you can double check you are choosing the right one when you insert them from Edit > Insert from CharactersMap):

- The lunate sigma (Ϛ) is not an English C or a Cyrillic C: its Unicode character code is 03F2 (03F9 for Capital Lunate Sigma). The Lunate sigma character is used in the diplomatic transcription, while standard sigmas are used in the edited transcription. In case a diplomatic is not present, you can insert in the relevant metadata section (i.e. "Writing") a note about this
- Greek Koppa (Ϟ, ϟ) corresponds in Unicode to 03DE (capital) and 03DF (small): if needed, make sure to use this character, and NOT the Cyrillic one
- Greek numerals are Greek letters marked with a tick (´) (Unicode character code: 0374) at the upper right, which looks like an apostrophe. So, in Greek transcriptions look carefully for the difference between number signs and the very rare apostrophe. There is also an occasional number mark at the lower left of a Greek numeral, like a comma (,) (Unicode character code: 0375). You can learn about Greek numerals here: https://en.wikipedia.org/wiki/Greek_numerals.
- Greek uses two differently shaped sigmas. It has a final sigma (ς) and a sigma used in all other places throughout the word (σ). Make sure to use the final sigma where appropriate.
- Greek inscriptions sometimes make use of a large lowercase omega (which looks like a curvy "W": ω) in the place of a majuscule Omega. **Do not** encode the large lower case omega with the Cyrillic letter that resembles it! Instead, encode a standard majuscule omega (Ω) and record in the relevant box (Writing: More Details) that the inscription uses large minuscule omegas in place of their majuscule counterparts. Here is the reference to it within our template:

```
<handDesc><handNote ana="xyz"><dimensions type="letter" extent="height"
unit="cm" quantity=""/></handNote><p>TEXT HERE</p></handDesc>
```

N.B.: Our program assumes a white space anytime you press "Return." Therefore, make sure *not to hit return* in the middle of the word, especially if you are working in Text Mode, as this is crucial for searching purposes. Even when an author has used soft hyphens to mark words split over two lines, stick to our line-break conventions (see below), and do not enter soft hyphens.

The Notes box (<div type="commentary" ana="b1">) is where you include any other additional information that you feel is important for the viewer to have. These notes will appear in the Inscription display. Generally, these deal with details about the inscription and not commentary on its contents. Try to be concise.

Encoding Textual Features

A summary of the tags we use the most can be found in the [Leiden to EpiDoc Cheatsheet](#).¹ Below is a more expansive version of this, with some comments relevant to our encoding practices. Tags can be added manually by typing them out in Editor Mode, by pressing “enter” in Author Mode, or pressing CTRL+E in either.

IMPORTANT: Make sure that all attribute tags are limited to a single word with no white-space in between. Thus, rather than encoding something as

```
<supplied reason="lost">Marcus Verus</supplied>
```

you MUST encode it as

```
<supplied reason="lost">Marcus</supplied> <supplied  
reason="lost">Verus</supplied>.
```

The reason for this is that the wordlist identifies spaces as word breaks, so we do not want any attributes where there is a space. There are a few exceptions where spaces within tags are okay (i.e., `<foreign>` tags, `<choice>` tags involving multiple words within a given option).

- Abbreviations

A standard way to publish an abbreviation in an inscription is: `v(iro)`, meaning that the only letter visible on the support is `v`, while `(iro)` has been supplied by the editor to make the text understandable. In XML we represent it as follows:

```
<expan><abbr>v</abbr><ex>iro</ex></expan>
```

Abbreviations are tagged only in the edited transcription.

See: <http://www.stoa.org/epidoc/gl/latest/trans-abbrevfully.html>

- Abbreviation Marks

At times, an abbreviation is signaled through an additional letter or a punctuation mark. Any characters or symbols that appear in the abbreviated form on the support (`<abbr>`), but do not

¹ Please refer also to: [EpiDoc Guidelines 9.2](#) / [List of All Transcription Guidelines](#).

form part of the fully expended word (<expan>), should be included within the <abbr> element, and should additionally be enclosed in an <am> (abbreviation mark) element, as follows:

```
Augg(usti)
<expan><abbr>Aug<am>g</am></abbr><ex>usti duo</ex></expan>
```

In this example, the inscription text is Augg which means Augusti duo.

See: <http://www.stoa.org/epidoc/gl/latest/trans-abbrevmark.html>

- Alternative readings

When the editor provides an alternative reading *for a whole word*, encode like this: <app type="alternative"><lem>first reading</lem><rdg>second possible reading</rdg></app>

See: <http://www.stoa.org/epidoc/gl/latest/supp-app-inline.html>

- Ambiguous characters with alternatives offered

Use the following tagging. This is often indicated either in a critical apparatus, with a slash, a question mark, or the word "or". Any of the following might be encoded as follows: "α/β" "α or β" "α? or β?"

```
<choice><unclear>α</unclear><unclear>β</unclear></choice>
```

See: <http://www.stoa.org/epidoc/gl/latest/trans-ambiguousalt.html>

- Bilingual inscriptions

When two languages (e.g. Greek and Hebrew) appears on the same support and within one and the same text, the language that appears second is marked as <foreign> in the edited transcription, with the indication of the relevant language in the xml:lang attribute. For example, δοῦλος θεοῦ אבנ would be rendered:

```
δοῦλος θεοῦ <foreign xml:lang="heb">אבנ</foreign>
```

See: <http://www.stoa.org/epidoc/gl/latest/trans-foreigntext.html>

- Columns

In case you encounter an inscription where the text is in multiple columns, they should be tagged as explained here: <http://www.stoa.org/epidoc/gl/latest/trans-textpart.html>

As of now, our web page does not display columns, but a correct tagging is important for searching purposes: **be sure to describe how the columns look and are read in the relevant section regarding writing in our template** (<p></p> within <handnote>, which appears as a "More detail if necessary" box in Author Mode).

- Deletion

If an object preserves traces of an inscription that has been rubbed out, or otherwise removed, editors represent the text in double square brackets: [[Legio]]. In our corpus, **if it is still somehow legible**, erased text is represented as follows both in the diplomatic and in the edited transcription boxes:

```
<del rend="erasure">Legio</del>
```

That is, in this case, the person attempting to erase the word "legio" was not entirely successful, as it is still partially visible.

See: <http://www.stoa.org/epidoc/gl/latest/trans-erased.html>

If the text was erased and it became illegible, but is restored by the editor, a <supplied></supplied> is put around the reconstructed text *within* the tag, as follows:

```
<del rend="erasure"><supplied reason="lost">xyz</supplied></del>
```

Note that each individual word should be listed as <supplied> independently of adjacent words - DO NOT use it to encompass more than one word at a time.

See: <http://www.stoa.org/epidoc/gl/latest/trans-erasedillegible.html>

- Gap

A standard way to represent gaps when publishing inscriptions is [----] or [....]. Often the number of dashes or dots indicate the number of missing characters, so the prior examples indicate four missing characters. In XML we represent this as follows:

```
<gap reason="illegible" (OR: "lost", depending on the case) quantity="4"
unit="character"/>
```

See: <http://www.stoa.org/epidoc/gl/latest/trans-lostcharknown.html>

If the extent of the gap is unknown, we use the following mark up:

```
<gap reason="lost" extent="unknown" unit="character"/>
```

See: <http://www.stoa.org/epidoc/gl/latest/trans-lostcharunknown.html>

For the XML to validate properly and for the inscription to be displayed correctly, please make sure to always indicate: reason; quantity OR extent; unit.

If the extent is given as a circa value, for example [ca. 13], mark as follows:

```
<gap reason="xyz" atLeast="12" atMost="14" unit="character"/>
```

See: <http://www.stoa.org/epidoc/gl/latest/trans-lostcharapprox.html>

If the extent is a range, for example [-5-6-], mark as follows:

```
<gap reason="xyz" atLeast="5" atMost="6" unit="character"/>
```

See: <http://www.stoa.org/epidoc/gl/latest/trans-illegiblecharapprox.html>

If the gap is represented with units other than characters, marked as such: [-5 cm-], it becomes:

```
<gap reason="xyz" extent="5" unit="cm"/>
```

See: <http://www.stoa.org/epidoc/gl/latest/trans-lostotherunits.html>

An intentional blank space is called a *vacat*. **See below for this.**

We also tag **missing lines**. For example, a gap of three lines should read:

```
<gap reason="xyz" quantity="3" unit="line"/>
```

If there are missing lines *one after the other* there is no need to mark each individually, followed by a line-break tag. As indicated above, it is enough to indicate the quantity of the missing lines in the relevant attribute. *Used in both diplomatic and transcription.*

See: <http://www.stoa.org/epidoc/gl/latest/trans-illegiblelineknown.html>

Note that a gap should NEVER be nested inside a supplied: they should be encoded separately, as follows:

```
<gap reason="lost" extent="unknown" unit="character"><supplied  
reason="lost">abc</supplied>
```

Please make sure to understand how each kind of gap is represented in the printed editions before proceeding to encode the inscription. Editors often put a list of the conventions used in the introductory chapter: make yourself familiar with it. If anything is unclear or you are in doubt, ASK!

- Ligature

At times, two or more letters are joined together to form a single sign, as it happens, for instance, with œ (o and e are joined together), or with 8 (o and u are joined together). These characters are tagged as follows:

```
<hi rend="ligature">ou</hi>
```

Note that the letters are written normally within the mark up (i.e. as if they were not joined).

Used in both diplomatic and transcription.

See: <http://www.stoa.org/epidoc/gl/latest/trans-ligature.html>

- Line break

`<lb/>` Mark a line break on every line, except the final one. If the inscription is only one line long, do not enter a line break tag.

Note that a `<lb/>` tag can occur in the middle of the word. When this is the case, it is important **NOT** to put any spaces between the tag and the continuation of the word if you are working in Text Mode and **not to actually type a hyphen**. Instead, encode as follows: `<lb
break="no"/>`. For instance, consider something that might appear as follows in a critical edition

Legio Fre-
tensis

This would be encoded as:

legio Fre<lb break="no"/>tensis

Note that there is no hyphen and no space between the line break.

See: <http://www.stoa.org/epidoc/gl/latest/trans-linebreak.html>

- Numbers

A number or a fraction should always be placed inside a `<num>` tag, with the indication of their value, i.e.: `<num value="5">V</num>`; `<num value="5">ε</num>`.

See: <http://www.stoa.org/epidoc/gl/latest/trans-numknown.html> (numbers)

See also: <http://www.stoa.org/epidoc/gl/latest/trans-fractionknown.html> (fractions)

For numbers indicated with paleo-Hebrew letters see below.

At times, a line above or below the letters indicates that they have a numeric value. If this happens, it is tagged as follows (both in diplomatic and in edited transcription):

`<num value="5"><hi rend="supraline" (OR) "underline">ε</hi></num>`

See: <http://www.stoa.org/epidoc/gl/latest/trans-supralinearetc.html>

Greek numerals are Greek letters marked with a tick (´) (Unicode character code: 0374) at the upper right, which looks like an apostrophe. So, in Greek transcriptions look carefully for the difference between number signs and the very rare apostrophe. There is also an occasional number mark at the lower left of a Greek numeral, like a comma (,) (Unicode character code: 0375). You can learn about Greek numerals here: https://en.wikipedia.org/wiki/Greek_numerals.

- Orig/Reg

This tagging is used to indicate text normalized or regularized by the editor from a dialect or phonetic spelling, grammatical form, etc., *usually for whole words* (as opposed to sic/corr, which typically is used for single characters: see below). This is always nested within

`<choice></choice>`. For instance, in Rehovot-in-the-Negev, the word *πρεσβύτερος* takes the unusual spelling of *πρεσβίτιρος*. It would be encoded as:

`<choice><reg>πρεσβύτερος</reg><orig>πρεσβίτιρος</orig></choice>`

See: <http://www.stoa.org/epidoc/gl/latest/trans-regularization.html>

- Phoenician (paleo-Hebrew) letters and/or numerals

(cf. e.g. *ostraca* from Masada) We use a glyph tagging in these cases. Each letter needs to be tagged separately as a glyph, and identified using the Unicode name as shown in the link below. So, encode as follows:

```
<g ref="phoen_three"/>
```

```
<g ref="phoen_alef"/>
```

Please note that no spaces are allowed within "ref", so a dash needs to be added in between the words that compose each single definition.

For definitions, use this as a reference tool: <http://unicode.org/charts/PDF/U10900.pdf>.

- Raised/Lowered characters

Within the same word, some characters may at times appear above or below the rest, as in: e^ar_th. They are tagged as follows, *both in diplomatic and in edited transcription*:

```
e<hi rend="superscript">a</hi>r<hi rend="subscript">t</hi>h
```

See: <http://www.stoa.org/epidoc/gl/latest/trans-raisedlowered.html>

- Superfluous text

Erroneously included text is usually indicated as {xyz} in the printed edition. It is tagged as follows:

```
<surplus>xyz</surplus>
```

See: <http://www.stoa.org/epidoc/gl/latest/trans-superfluous.html>

- Supplied text

Letters are usually missing because the text area is damaged, or because the writer of the inscription made a mistake.

Squared brackets typographically mark letters that are supplied by an editor due to damage: Le [g]. We would typically encode this as:

```
Le<supplied reason="lost">g</supplied>
```

In cases where the restoration is less certain, we can add a certainty attribute. So, Le[g?] would be encoded:

```
Le<supplied reason="lost" cert="low">g</supplied>
```

Note that each individual word should be listed as <supplied> independently of adjacent words – DO NOT use it to encompass more than one word at a time.

See: <http://www.stoa.org/epidoc/gl/latest/trans-restored.html>

Angle brackets typographically indicate letters omitted by the writer of the inscription and supplied by the editor. L<e>g would thus render the diplomatic Lg. This is tagged:

```
L<supplied reason="omitted">e</supplied>g
```

Note that each individual word should be listed as <supplied> independently of adjacent words - DO NOT use it to encompass more than one word at a time.

See: <http://www.stoa.org/epidoc/gl/latest/trans-omitted.html>

Used only in the edited transcription.

At times, editors supply words because they are subaudible, these are often words indicated by their full inclusion in parentheses. For instance, "Cneio Pompeio Sexti filio (tribu) Quirina" would be encoded as follows:

```
Cneio Pompeio Sexti filio <supplied  
reason="subaudible">tribu</supplied> Quirina
```

Note that each individual word should be listed as <supplied> independently of adjacent words - DO NOT use it to encompass more than one word at a time.

See: <http://www.stoa.org/epidoc/gl/latest/trans-subaudible.html>

- Sic/Corr

This is used to indicate words that the writer *misspelled*. Editors often typographically mark this with (!). For example, let's imagine that the writer recorded the name "Augustus" when the correct spelling was "Augustus". The author of the article indicates this error, so it's our job to record it. We do so like this:

```
<choice><corr>Augustus</corr><sic>Augostus</sic></choice>
```

See: <http://www.stoa.org/epidoc/gl/latest/trans-erroneoussubstitution.html>.

Used only in the edited transcription.

- Unclear Letters

A standard way to indicate an unclear character when publishing inscriptions is to have the letters appear with a dot below it, for example the character ḁ. In XML, we represent this with the tag:

```
<unclear>a</unclear>
```

Please note that multiple unclear letters in a row are tagged within a single unclear, but this should not go across multiple words: `<unclear>xyz</unclear>`

See: <http://www.stoa.org/epidoc/gl/latest/trans-ambiguous.html>

Traces of **letters that are barely visible** are also signaled with an underdot, but these are encoded differently, using the `<gap>` tag. In such cases, use the following format:

```
<gap reason="illegible" quantity="x" unit="character"/>
```

Used in diplomatic and edited transcription.

- Uninterpreted characters

These characters are clearly readable, but the editor cannot make sense of them. They are tagged using `<orig>`, as follows: `<orig>xyz</orig>`. They usually appear in upper case in the printed edition.

See: <http://www.stoa.org/epidoc/gl/latest/trans-uninterpreted.html>

- Vacat

Spaces that are left intentionally blank are often indicated as (vac.). We tag this with the space element: `<space quantity="x" unit="y"/>`, where x is a number and y is usually "character". If the extent is unknown, the tag is: `<space extent="unknown" unit="character"/>`.

See: <http://www.stoa.org/epidoc/gl/latest/trans-vacatchar.html>

Symbols and Special Characters

- Cross

In order to enter a cross, simply type a plus sign: +. Crosses are also tagged as figures, as follows: `<g ref="cross">+</g>`

- Letter "ζ" as symbol for drachma(s)

Tag as follows in the edited transcription:

```
<expansion><abbr><am>ζ</am></abbr><ex>δράχμα</ex></expansion>
```

See: <http://www.stoa.org/epidoc/gl/latest/trans-abbrevsymbol.html>

In the diplomatic transcription, if present, the letter is simply enclosed within the `<am>` tag:

```
<am>ζ</am>
```

- Letter "L" as a symbol for the Greek word ἔτους ("year")

Tag as follows in the edited transcription:

```
<expansion><abbr><am>L</am></abbr><ex>ἔτους</ex></expansion>
```

See: <http://www.stoa.org/epidoc/gl/latest/trans-abbrevsymbol.html>

In the diplomatic transcription, if present, the letter is simply enclosed within the `<am>` tag:

```
<am>L</am>
```

- Greek letter stigma (ς) as a sign of abbreviation

This letter (which can be easily found through a word search in the xml file after clicking on: Edit > Insert from character map) is often used as a sign of abbreviation. It is tagged as follows in the expansion within the edited transcription:

```
<expansion><abbr>κ<am>ς</am></abbr><ex>αἰ</ex></expansion>
```

Greek capital stigma (ς) Unicode character code: 03DA

Greek small letter stigma (ς) Unicode character code: 03DB

- Middle dot and Word Separator

To indicate this Greek sign of punctuation (middle dot, corresponding to a semicolon) please make sure to use the correct character, which is 00B7 in Unicode (·). Be careful not to confuse Greek middle dot, which is punctuation, with the word separator middle dot (·), used to separate words (usually) in Latin inscriptions (Unicode character 00B7).

- Monograms

Where a monogram appears as part of the transcribed text of an inscription, there are a range of different ways to handle this in encoding, depending on our ability to read and expand any

textual content. In our inscriptions the names were often legible or understandable, and the monograms were tagged as follows: `<expan><abbr><am><g`

`type="monogram"/></am></abbr><ex>Postumius</ex></expan>`

See: <http://www.stoa.org/epidoc/gl/latest/trans-symbolmonogr.html>

- Chi-Rho

The symbol for Χριστός is usually tagged as follows: `<g ref="chi-ro"/>`

See: <http://www.stoa.org/epidoc/gl/latest/trans-symbolchirho.html>

- Glyphs

Glyphs are characters that are unique, not alphabetic or not available in unicode. Examples are Phoenician letters, decorations and symbols. They are marked with the `<g/>` tag, and briefly described in the attribute `@ref`, as follows: `<g ref="leaf"/>`; `<g ref="menorah"/>`; `<g ref="dove"/>`. Although `<g>` is almost always an empty element, it is possible to put content into this element, as we do with the cross: `<g ref="cross">+</g>`

A fuller description of this kind of glyphs and of their position within the text should be inserted as metadata in the relevant field (Decorations).

There is a further element if the glyph is a character (i.e., a letter or number) `@type`, such as `<g ref="archaic_aleph" type="char"/>`

See: <http://www.stoa.org/epidoc/gl/latest/trans-symbol.html>