

Citations and References

Marjorie Burghart

In this chapter you will learn:

- to encode the sources used or referred to in a text, and link them to bibliographic citations;
- to encode the biblical references, and represent them as canonical references;
- to handle overlapping hierarchies if they occur.

Identifying the sources of a text is a key part of the editing process. Sources, in this context, are passages of the edited text which are quoted explicitly or implicitly from other works, or simple allusions to or reminiscences of other works. Their identification by the editor is very important to reconstruct the intellectual environment of the work at hand, the influences, and the methods used by the author, leaning on the authority of famous authors or compiling unattributed quotations. These sources can belong in various categories or styles: sometimes they are quoted silently, i.e., without any attribution, sometimes the author gives their reference (attributing them to an author and/or work), which may or may not be correct. Other categories, defined by the editor, may also be useful in order to categorise the sources: classical, patristic and contemporary sources for instance. Canonical references, like biblical quotations usually form a distinct group and need to be displayed and processed in a specific fashion.

The TEI offers a general frame for the encoding of quotations, within which you are free to develop your own strategies. The solutions suggested here are only possibilities, among many, that are offered to you with the needs of scholarly editors in mind. You are invited to consult the relevant sections of the Guidelines ¹ to further your knowledge of the various possibilities.

1. Defining your needs

The first step consists in a reflection upon your needs and what you want to achieve with your edition.

If you want to display or process the quoted bits according to their category, you will need to attach a different type to each quoted source. This will be useful, for instance, if you want to be able make a separate index or group of critical notes for the sources quoted with an attribution and those quoted silently, or if you wish to display biblical quotations in italics and other sources

1. See in particular chapter [3.3.3. Quotations](#)

between quotes. Attributing a type to the sources will also let you process the edition in a more detailed way. You could for instance count the words in order to determine what proportion of your edition each type of quoted source represents, or you could also map the use of a certain category of sources across a corpus.

If you plan to have a lot of source identifications, it may be useful to consider using a mechanism that lets you encode a link between a citation and a full bibliography (in the case of a literary source), or formally represent a canonical reference (in the case of biblical quotations). It will prevent the addition of redundant information.

The general mechanism to encode a quotation together with its reference relies upon the `<cit>` element (cited quotation), which “contains a quotation from some other document, together with a bibliographic reference to its source” ([Guidelines](#)). The element `<quote>` contains the quotation, and the element `<bibl>` the bibliographic reference.

2. A basic example

Let us start with a basic case: imagine you find this source quoted in your edition, ascribed to “Bernardus”:

```
Bernardus : Auferatur malus ne generet malos. Non potest arbor mala  
fructus nisi malos facere.2
```

2.1. Minimal encoding

If you were interested only in distinguishing quotations from the rest of the text (for statistical purpose for instance), without attempting to give a bibliographic reference or identification of the source, you could simply use the `<quote>` element:

```
Bernardus : <quote>Auferatur malus ne generet malos. Non potest arbor  
mala fructus nisi malos facere.</quote>
```

But in most cases, the point of marking up quotations in an edition is precisely to give some form of identification for each of them. To this purpose, we will use the `<bibl>` element to give the bibliographic reference, and wrap together the `<quote>` and `<bibl>` in a `<cit>` element. You could encode this quotation as follows, using the `<bibl>` element to encode the ascription given in the text (which may or may not be accurate):

```
<cit>  
  <bibl>Bernardus :</bibl>  
  <quote>Auferatur malus ne generet malos. Non potest arbor mala  
fructus nisi  
  malos facere.</quote>  
</cit>
```

2. This Latin quotation translates as follows “Bernard : He who is bad must be removed so he does not generate other bad ones. The bad tree cannot bear fruits, except bad ones.”

This is a perfectly valid encoding, but in the following examples, we will use `<bibl>` to encode the bibliographic reference that we, the critical editor, have identified as the source. In this case, the data contained by `<bibl>` is not part of the text of the edition, but of its critical apparatus, and should be displayed and processed accordingly. In any case, documenting your usage of the element in the header is useful.³

You have identified the source as a letter by Bernard of Clairvaux, numbered 102 in the reference edition of his letters. You may encode the quoted text together with the reference you have found like this:

```
Bernardus : <cit>
  <quote>Auferatur malus ne generet malos. Non potest arbor mala
fructus nisi malos facere.</quote>
  <bibl>Bernardus Claraevallensis, Epistolae, 102</bibl>
</cit>
```

If you want a more precise encoding, you can enhance the `<bibl>` element from the previous example:

```
<bibl><author>Bernardus Claraevallensis</author>,
<title>Epistolae</title>, 102</bibl>
```

This will allow you to process the different parts of the bibliographic reference according to your needs, for instance to display the author's name in small capitals and the title in italics.

2.2. Distinguishing between different categories of citations

You might want to give a type to this source. Since the author of the text has quoted it with an ascription, you might decide to add a `@type` attribute to `<cit>`, with the value `"ascribed"`, for instance. The TEI does not have specific recommendations for the possible value of `@type` here, you are free to use your own categories according to your needs.

```
Bernardus : <cit type="ascribed"> ... </cit>
```

If you wish to specify different values in `@type`, just separate them with a simple space. For instance, to express that this source is quoted with an attribution and also is a literal quotation (as opposed to a rephrasing or a mere allusion), you could use the following:

```
Bernardus : <cit type="ascribed literal"> ... </cit>
```

3. Linking sources to a bibliography

In the previous example, all the bibliographic information was contained in the `<bibl>` element. This is inconvenient when the same source is quoted multiple times in an edition,

3. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/HD.html#HD57-2>

because it forces you to repeat redundant information, at the risk of being inconsistent in your bibliographic citations of the same work.

Let's consider the example we used above. So far we have just given a short, incomplete reference, but we would need to link to the full reference edition of the letters of Bernard of Clairvaux, which is the following:

Bernardus Claraevallensis, *Epistolae*, in *Sancti Bernardi Opera*, Jean Leclercq and Henri Rochais eds., Rome, 1957-1977, vol. VII and VIII.

First, anywhere in the document, let us create a detailed description for each bibliographic citation we want to link to. You could put this bibliography in a `<front>` or `<back>` element, for instance, to clearly separate it from the `<body>` of your edition.

We offer below an example of bibliographic citation encoding. You could opt for less or more detailed options: the TEI offers many options to encode a bibliography, which go beyond the scope of this chapter. If you are interested in this topic, we recommend that you check the relevant sections of the Guidelines.

The important point here is that each bibliographic citation we want to be able to link to must have an `@xml:id` attribute, with a unique value.

```
<listBibl>
  <bibl xml:id="bernEpist"><author>Bernardus
Claraevallensis</author>, <title>Epistolae</title>, in <title>Sancti
Bernardi Opera</title>, <editor>Jean Leclercq</editor> and
<editor>Henri Rochais</editor> eds., <pubPlace>Rome</pubPlace>,
<date>1957-1977</date>, <biblScope unit="volume">vol. VII and
VIII</biblScope>.</bibl>
  <bibl xml:id="gregMor"><author>Gregorius Magnus</author>,
<title>Moralia in Iob</title> (etc.) </bibl>
</listBibl>
```

Within the body of the edition, when we want to link a source to a bibliographic citation, we can proceed as follows, using the `<ref>` element with its `@target` attribute to wrap the `<bibl>` description:

```
Bernardus : <cit>
  <quote>Auferatur malus ne generet malos. Non potest arbor mala
fructus nisi malos facere.</quote>
  <ref target="#bernEpist"><bibl><author>Bern.</author>,
<title>Epist.</title>, 102 (VIII, 257-8)</bibl></ref>
</cit>
```

This encoding is very similar to the basic example above, except that thanks to the `<ref>` element we now have encoded a link to a full description of the bibliographic citation where the edition of this letter appears, while the `<bibl>` let's us give the details for this particular quotation (which volume, and which pages).

When displaying or processing your edition, you will be able, for instance, to create an index of the works cited, to count or highlight the references to a particular bibliographic citation, or to display bibliographic notes in full or short version.

4. Canonical references: the example of biblical quotations

Canonical references are “any means of pointing into documents, specific to a community or corpus”⁴ They are identified not by bibliographic citation, like most literary sources, but by a short reference following rules defined by a scholarly community. That is the case of many religious texts (the Mishnah, the Bible, the Quran), but not only: some classical works have well-established systems of canonical references (Aristotle, for instance, is often quoted by a “Bekker number” in modern literature).⁵ A very common usage of canonical references in critical editions is the identification of quotations from or references to religious texts. It is common to treat those quotations slightly differently from other source quotations: they are generally displayed differently from other sources, have their own searchable index, and in print editions they usually have their own series of footnotes. In this section, we are going to use the example of biblical quotations to illustrate canonical references: with the adaptation of the reference scheme, these principles are directly applicable to other types of canonical references.

4.1. Particularities of biblical quotations

Biblical quotations are identified by canonical references, which take the form of:

- a book name or abbreviation thereof;
- a chapter number (optional);
- a verse number, or a range or series of verses (optional).

It might seem very simple: if I give the reference “Gn 1:1,” my readers will understand that this means Genesis, chapter 1, verse 1. But there are many potential difficulties, which may also apply to other canonical references. Some are rather trivial and pertain to the formatting of the references:

- there is no universal list of book names or abbreviations: they will differ from one language to another, and also from one editor to another, etc. For Genesis, instead of “Gn” I could have used “Gen,” “Genes,” the French name “Genèse,” the Italian name “Genesi,” etc. The names or abbreviations you are using should therefore be stated explicitly in your edition.
- the separators between the data may vary: some will use a colon to separate the chapter and verse numbers, others a comma followed by a space, etc.

Another difficulty is linked with the version of the Bible that is quoted: despite the canonical

4. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/SA.html#SACR>

5. For an introduction to canonical references, their issues and the solutions to handle them, see Kalvesmaki 2014.

character of the Bible, there are several translations, even for a single language. In a single edition, the Bible may be quoted after different translations, for instance the Vulgate and the Vetus Latina (two different Latin versions), or the Vulgate (Latin) and the Douay-Rheims version (English translation from the Vulgate).

4.2. Basic example

The encoding of biblical quotations bears many similarities with the one of literary sources. Let's consider this example:

```
Unde in Genesim: In principio creavit Deus celum et terram.
```

A basic encoding could be the same as the one used for literary sources in our first example:

```
Unde in Genesim: <cit>
  <quote>In principio creavit Deus celum et terram.</quote>
  <bibl>Gn 1:1</bibl>
</cit>
```

This is simple to implement, but also has limited possibilities: it will only let you display a note with the reference of the quotation.

Another drawback is that, as it is, it does not distinguish biblical quotations from other quotations. You can achieve that simply by adding a `@type` attribute to `<cit>`, with the value `"bible"` for instance:

```
Unde in Genesim: <cit type="bible"> ... </cit>
```

If you need more categories, you may add different values to `@type`, separated with a space, as we saw above.

But the identification of a biblical quotation is not really a bibliographic citation. More accurately, we could encode it as a reference, using the `<ref>` element as an alternative to `<bibl>`. According to the Guidelines, “`<ref>` (reference) defines a reference to another location, possibly modified by additional text or comment.” A minimal example would be the following:

```
Unde in Genesim: <cit type="bible">
  <quote>In principio creavit Deus celum et terram.</quote>
  <ref>Gn 1:1</ref>
</cit>
```

4.3. Going further with `<ref>`

The `<ref>` element may bear two optional and mutually exclusive attributes: the `@target`, which “specifies the destination of the reference by supplying one or more URI References,” and `@cRef` (canonical reference) attribute, which “specifies the destination of the pointer by supplying a canonical reference expressed using the scheme defined in a `<refsDecl>` element

in the TEI header.” For biblical quotations, `@cRef` is the best suited. To use it fully, we need to proceed in two stages.

First, within the edition, we are going to encode the biblical quotations as above, but we will now use the `@cRef` attribute to contain the canonical reference, which you must always construct following the same rules. Here I have opted for a book abbreviation followed by a space, then the chapter and verse numbers separated with a colon:

```
Unde in Genesim: <cit type="bible">
  <quote>In principio creavit Deus celum et terram.</quote>
  <ref cRef="Gn 1:1"/>
</cit>
```

Should you want to add precisions or comments about this reference, you are free to add them:

```
<ref cRef="Gn 1:1">This is a literal quotation of the first verse of
the Bible</ref>
```

Then, we must declare in the `<teiHeader>` how the canonical references in `@cRef` are constructed. To do so, we must add a `<refsDecl>` (references declaration) to the `<encodingDesc>` part of the header.

For our purpose, there are two way to explain your referencing scheme in the `<refsDecl>` element:

- either you simply explain in prose how you construct your references, giving all necessary information in plain text, as in the following example:

```
<refsDecl xml:id="biblicalCitations">
  <p>The <att>cRef</att> attribute of <gi>ref</gi> elements
citing the Bible contain a canonical reference in one to three
parts. The first part is an abbreviation for the name of the
relevant book of the Bible, and it is the only mandatory one. The
second part, preceded by a space, is a chapter number. The third,
preceded by a colon, is either a single verse number, or a series
of consecutive verses with the number of the first verse followed
by a hyphen, followed by the number of the last verse. </p>
  <p>The following list of abbreviations has been used for the
books of the Bible: Gn: Genesis; Exodus: Ex; Leviticus: Lv;
[etc.]</p>
</refsDecl>
```

This information will be very useful to future users (or even current collaborators) working on your edition, but it cannot be processed automatically.

- or you define the pattern(s) that your canonical references must match and how they must be transformed into a valid URI, using regular expressions.⁶ You can define several

6. Regular expressions are an extremely useful tool, available in many computing languages. They allow the

patterns, each in a `<cRefPattern>` element, using the two attributes `@matchPattern` (to define how canonical references are constructed), and `@replacementPattern` (to define how the different parts of the matching references must be transformed).

In the following example, we define a single rule: it matches canonical references (i.e., the values of `@cRef` attributes) composed of a string of any number of characters followed by a space, followed by another string, followed by a colon, and followed by a string. Then it transforms the parts identified into a URL pointing to an online edition of the Vulgate:

```
<refsDecl xml:id="biblicalCitations">
  <cRefPattern matchPattern="(.) (.):(.)"
replacementPattern="http://vulsearch.sourceforge.net/html/
$1.html#x$2_$3">
  <p>This pointer pattern extracts and references the
<q>book,</q> <q>chapter,</q> and <q>verse</q> parts of a biblical
reference pointing to a single verse, like "Gn 1:1", and
reconstructs a link to an online version of the biblical
text.</p>
  <p>The following list of abbreviations has been used for the
books of the Bible: Gn: Genesis; Exodus: Ex; Leviticus: Lv;
[etc.]</p>
</cRefPattern>
</refsDecl>
```

Here is how the regular expression works: in `@matchPattern`, each expression between parenthesis is a match. If the `@cRef` value "Gn 1:1" is processed, for instance, the first expression will retrieve the value "Gn", because it is a string of characters at the beginning of the reference and before the first separator, the space; the second will retrieve the value "1", as well as the third, because they are strings of characters respectively between the space and colon separators, and after the colon separator.

In `@replacementPattern`, each expression composed of a dollar sign and a number means it must be replaced with the value of the matched expression bearing this number. With the same example of "Gn 1:1", \$1 should therefore be replaced with the first match "Gn", \$2 with the second match "1" and \$3 with the third and last match "1", and the following URL would be constructed:

```
http://vulsearch.sourceforge.net/html/Gn.html#x1_1
```

You could add more `<cRefPattern>` elements defining more rules, for instance when a canonical reference points to a full chapter (Gn 1), or a full book (Gn).

definition of patterns for search and replace operations. A full introduction to regular expressions is beyond the scope of this chapter, but we highly recommend scholars interested in advanced search-and-replace operations to read further on this topic.

Note that there might be more than one `<refsDecl>` in a single file, so it may be useful to give it an `@xml:id` (unique identifier). If there are several `<refsDecl>`, you can specify in each `<ref>` which one applies with the attribute `@decls`.

```
<ref cRef="Gn 1:1" decls="#biblicalCitations">Gen 1, 1</ref>
```

Whatever solution you opt for, well-constructed canonical references will allow you to process your edition in a meaningful and useful way: extracting all quotations from a particular book or chapter of the Bible for instance, or building an index of biblical citations, etc.

5. Allusions without actual quotation

For literary as well as biblical sources, it may happen that a reference is made to a text without actually quoting it. For instance, here we have a reference to an evangelical parable, which appears in Luke, chapter 10, verses 25–37, but there is no quotation from the parable:

```
(...) as it is demonstrated by the Parable  
of the Good Samaritan (...)
```

Similarly, here we have an allusion to a text by saint Augustine, where he laments on his youthful sins, among which the theft of pears in vain. This comes from the Confessions, book II, chapter 4, but again there is only a reference and no quotation:

```
(...) Augustine says he stole pears in a  
garden when he was a young man (...)
```

Since the source is not actually quoted, we cannot use the `<quote>` element, but the `<bibl>` or `<ref>` elements are still relevant.

```
... as it is demonstrated by the Parable  
of the Good Samaritan<ref cRef="Lc 10:25-37"/>...
```

```
Augustine says he stole pears in a  
garden when he was a young man<ref target="#AugConf"><bibl  
type="source">Aug., <title>Confessiones</title>, II, 4</bibl></ref>
```

You could also choose, perhaps for considerations linked to the processing of the edition, to still wrap those `<ref>` and `<bibl>` elements in a `<cit>`, although it is not necessary.

6. Handling overlaps

It may happen that quotations, which are often numerous, overlap with other TEI elements in an edition - typically `<app>` elements dedicated to recording the textual variants. Overlapping or non-nesting information is an issue with all XML-based languages, and the TEI is not an exception. The impossibility for elements to overlap certainly poses a problem, but the TEI has various mechanisms to overcome this difficulty.

Let us consider the following example: we see there are two biblical quotations, the first from Ephesians, 5, 30 and the second from Luke, 24, 39:

And as the Apostle says in Eph. V: We are members of his body, of his flesh, and of his bones. And also in Luke XXIIII: A spirit hath not flesh and bones, as you see me to have.

But we have already encoded the variants from three witnesses A, B and C. As it happens, witness C presents a particular type of omission, caused by homeoteleuton: as two lines had similar endings, (here, “flesh”), the scribe’s eye slipped and he missed several words, right across the two citations:

```
And as the Apostle says in Eph. V: We are members of his body, of his
<app>
  <lem>flesh, and of his bones. And also in Luke XXIIII: A spirit
hath not</lem>
  <rdg wit="#C" cause="homeoteleuton"/>
</app> flesh and bones, as you see me to have.
```

The TEI Guidelines support four XML-based methods for handling overlapping information, which are exposed in details.⁷ Each method has its pros and cons, but for our purpose we recommend using in most cases the Fragmentation and Reconstitution of Virtual Elements method.

6.1. Fragmentation and Reconstitution of Virtual Elements method

This method consists in choosing a privileged hierarchy, and breaking up overlapping elements from the other hierarchy in smaller elements, connected between them but not overlapping anymore.

With our example, here is how we could solve the overlapping problem with this method, while privileging the textual variance hierarchy (<app> etc.): we would split each overlapping <cit> element into two smaller elements (or more if it was necessary) that do not overlap. To help with the virtual reconstruction of the full <cit> elements, we have given each split part of the original <cit> a unique identifier (@xml:id), and used the attributes @prev and @next to point to the previous and following part of the citation, respectively (this is called a virtual join, or a chain). We put the <ref> element, which does not overlap, only in the final part of the citation:

```
And as the Apostle says in Eph. V:
  <cit xml:id="cit01a" next="#cit01b" type="bible"><quote>We are
members of his body, of his </quote></cit><app>
  <lem><cit xml:id="cit01b" prev="#cit01a" type="bible"><quote>flesh,
and of his bones</quote><ref cRef="Eph 5:30"/></cit>. And also in Luke
XXIIII:
```

7.

```

    <cit xml:id="cit02a" next="#cit02b" type="bible"><quote>A spirit
hath not</quote></cit></lem>
    <rdg wit="#C" cause="homeoteleuton"/>
</app>
    <cit xml:id="cit02b" prev="#cit02a" type="bible"><quote>flesh and
bones, as you see me to have</quote><ref cRef="Lc 24:39"/></cit>.

```

This method is a bit verbose, but easy enough to encode and process.

One thing you must remember when processing a document created with this method is that there are more `<cit>` elements than there are actual citations. Therefore, if you want to know, for instance, how many citations there are in your edition, counting the `<cit>` elements will be misleading. With this example, you would find a result of 4, while there are only 2 biblical citations, each split in two.

This is easily solved by preparing a more precise query, for instance you could add up the number of `<cit>` elements that do not have a `@prev` nor a `@next` attribute, and the number of `<cit>` elements with a `@next` but no `@prev` attribute (i.e., the first `<cit>` elements of a chain).

Nota bene: in the example above, we have chosen to privilege the textual variants hierarchy of the sources hierarchy, but we could do the opposite if it was more suited to our edition for instance if processing the sources were our top priority, and textual variance only accessory. The exact same principle would apply: we would break up the apparatus entries into smaller elements nesting in the `<cit>` hierarchy, and virtually join the split elements.

6.2. Alternative methods

The TEI also supports three other XML methods to handle overlapping hierarchies. Depending on your data, your goal and your familiarity with computer processing of the result, they may also be a good choice for your edition.

6.2.1. Multiple Encodings of the Same Information

Multiple Encodings of the Same Information is technically the simplest method, which consists in re-encoding the same information as many times as you have overlapping hierarchies. In practical terms, it means that we would need to have one text of the edition with the variants encoded, and another text of the same edition with the sources encoded. This is therefore very redundant, and can be difficult to maintain if your text is not stable: should you spot a typo or an error in your transcription at some point, you will have to correct it consistently across all the different encodings of this text.

With our example, it would result in the following: a first text with the textual variance encoded, without information on the quotations:

```

And as the Apostle says in Eph. V: We are members of his body, of his
<app>
    <lem>flesh, and of his bones. And also in Luke XXIIII: A spirit

```

```

hath not</lem>
  <rdg wit="#C" cause="homeoteleuton"/>
</app> flesh and bones, as you see me to have.

```

And a second text with the quotations encoded, without information on the textual variance, which is not without problems in the case of a critical edition. Here we have a clearly identified lemma vs. an erratic manuscript, which makes things easier. But if we had more subtle variants, it would be much more difficult to determine what is the text:

And as the Apostle says in Eph. V:

```

<cit type="bible">
  <quote>We are members of his body, of his flesh, and of his
bones.</quote>
  <ref cRef="Eph 5:30"/>
</cit> And also in Luke XXIIII: <cit type="bible">
  <quote>A spirit hath not flesh and bones, as you see me to
have.</quote>
  <ref cRef="Lc 24:39"/>
</cit>

```

6.2.2. Boundary Marking with Empty Elements

With this method, the start and end points of the problematic (non-nesting) elements are marked with empty tags, like **<anchor>**. Although efficient, this method poses various problems. It can get confusing for the encoders and lead them to introduce errors that will be difficult to screen. The processing stage will also be difficult: no off-the-shelf application will be able to process your data satisfactorily, and you will need to write complex transformation programs capable of handling this type of encoding.

Using again our example, here is what the encoding would look like with this method. Let's say that we are going to privilege the textual variance hierarchy over the identification of sources. It means that we are going to keep classic elements for textual variance (**<app>** etc.), and represent only the start and end points of overlapping source identifications (**<cit>** and **<quote>**) with **<anchor>**. The result is cumbersome and poorly readable, on top of being difficult to encode and process:

And as the Apostle says in Eph. V:

```

<anchor type="delimiter" subtype="citBibleStart"/>We are members of
his body, of his <app>
  <lem>flesh, and of his bones <ref cRef="Eph 5:30"/> <anchor
type="delimiter" subtype="citBibleEnd"/>. And also in Luke
XXIIII: <anchor type="delimiter" subtype="citBibleStart"/>A spirit
hath not</lem>
  <rdg wit="#C" cause="homeoteleuton"/>
</app> flesh and bones, as you see me to have <ref cRef="Lc
24:39"/><anchor type="delimiter" subtype="citBibleEnd"/>.

```

6.2.3. Stand-off Markup

Stand-off Markup offers many possibilities, but it is the most complicated to encode and process. With this method, it is possible to annotate a file without writing in it: for instance, you could annotate a read-only edition made available online by a library, even if it is only a plain-text file. When you are annotating your own file, you can choose to put the stand-off markup in the same file, or in another. It is a very good method, offering many advantages for collaboration (many people can add annotations at the same time, without the risk of overlapping), and for texts with multiple, complex and overlapping layers of annotations. But it is very difficult to process, and often to encode. This is why we recommend using this method only when this is the only solution to represent satisfactorily your data.

Using this method, we could encode our example like this: first, we would need elements with a unique identifier to mark the boundaries of the bits of texts we want to annotate. We do not always need to add new elements for that, but in this case there is nothing at this place in our encoding, so we are going to use empty elements (`<anchor>`), at the beginning and end of each citation.

```
And as the Apostle says in Eph. V: <anchor xml:id="cit01"/>We are
members of his body, of his <app>
<lem>flesh, and of his bones<anchor xml:id="cit02"/>. And also in
Luke XXIIII: <anchor xml:id="cit03"/>A spirit hath not</lem>
<rdg wit="#C" cause="homeoteleuton"/>
</app> flesh and bones, as you see me to have<anchor
xml:id="cit04"/>.
```

Somewhere else in the same document (or even in a separate file), we would reconstruct the desired encoding for the annotations:

```
<cit type="bible">
<quote>
<xi:include xpointer="range(element(cit01),element(cit02))"/>
</quote>
<ref cRef="Eph 5:30"/>
</cit>
<cit type="bible">
<quote>
<xi:include xpointer="range(element(cit03),element(cit04))"/>
</quote>
<ref cRef="Lc 24:39"/>
</cit>
```

Bibliography

Joel Kalvesmaki, "Canonical References in Electronic Texts: Rationale and Best Practices,"
in *Digital Humanities Quarterly*, Volume 8, Number 2, 2014
[<http://www.digitalhumanities.org/dhq/vol/8/2/000181/000181.html>]