

Accessing emblems using XML. Digitisation Practice at the Emblem Project Utrecht

Peter Boot
Constantijn Huygens Institute
Research Institute for History and Culture (Utrecht University)

Abstract: The article describes the Emblem Project Utrecht and the XML (eXtensible Markup Language) encoding it will apply to 25 books of Dutch love emblems. The encoding procedures are based on the TEI (Text Encoding Initiative) Guidelines. It argues emblem digitisation should be based on public standards in order to ensure the longevity of its products. The article concludes by looking forward to the possible use of XML in emblem research.

Published in: López Poza, S. (ed.), *Florilegio de Estudios de Emblematica. A Florilegium of Studies on Emblematics*. Proceedings of the 6th International Conference of the Society for Emblem Studies, A Coruña, 2002. Ferrol 2004. Pp. 191-197.

Author contact: pboot@xs4all.nl

Immortality is attained by literary studies, as Alciato said, and as, indeed, his own example proves. The books which we write about have survived for three to five centuries. Those of us that work on digital editions may want to reflect about that: the printed book seems to last quite well. How will our digital editions fare in this respect? It doesn't seem excessively prudent to doubt whether they'll outlive Alciato.

In this paper I will discuss the digitisation techniques used at the Emblem Project Utrecht. While these techniques don't promise a life span of several centuries, they were designed to provide durability. And having seen a few generations of computers and computer software come and go, we should know durability is really important.

After a brief introduction about the Emblem Project Utrecht (section 1) I will show some simple examples of the encoding technique the project uses (section 2). In section 3 I will describe the public standards (XML, or eXtensible Markup Language, the *TEI*, or Text Encoding Initiative, *Guidelines* and Unicode) which these encoding techniques adhere to. Section 4 will show how application of these standards facilitates flexibility in showing the results to a reader or user. In the last section (section 5) I will conclude by showing how the same techniques may be used to encode features certainly not part of the traditional scholarly edition.

1. The Emblem Project Utrecht

The basic aim of the Emblem Project Utrecht¹ is the digitisation of about 25 books of Dutch love emblems. The Dutch love emblem was created by Daniel Heinsius (*Quaeris quid sit amor?* Leiden 1601) and lived on until the early 18th century. Originally a secular genre, Otto Vaenius created a religious counterpart when he rewrote his *Amorum emblemata* (Antwerp 1608) as *Amoris divini emblemata* (Antwerp 1615). Both the secular and religious subgenres were to have great impact on the tradition of the European emblem. In the Low

¹ The project was created by dr. Els Stronks (project co-ordination and research) and Peter Boot (technical realisation and research) at Utrecht University (the Netherlands). It received funding from the Research Institute for History and Culture (Utrecht University) and co-operates with the Digital Library of Dutch Literature (DBNL), the Constantijn Huygens Institute for text editions and intellectual history (CHI) and the Netherlands Institute for Scientific Information Services (NIWI). At the time of writing, an application for funding was under review by NWO (Netherlands Organisation for Scientific Research).

Countries, the main names associated with the genre, apart from Heinsius and Vaenius, are P.C. Hooft, Jacob Cats, Hugo, and Jan Luyken.

The project will create digital editions of these books, including full texts, facsimiles, concordances, indexes on pictorial motifs, and (possibly) translations, annotation, commentary, pointers to parallels and sources, pointers to relevant literature, etc. On the project web site (<http://www.let.uu.nl/emblems>) preliminary editions of five books are now available (besides the three books already mentioned, Jacob Cats's *Sinne- en Minnebeelden* (Rotterdam 1627) and Daniel de la Feuille's *Devises et emblemes* (Amsterdam 1691)).

The project, however, aims at more than just creating the digital equivalent of an edition on paper. The project site is to develop into the focal point for research in the Dutch love emblem. The research subjects will include the history of the genre as a web of quotations (of pictorial and textual motifs and topoi), the relevance of the genre to modern emblem theory, and its wider cultural influence.

Another main aim of the project is reflection on the consequences of the digital edition for scholarly research in the humanities. The interactivity of the digital edition will have an impact on both research methods and ways of publication. The distinction between scholarly editions and research publications may be blurred, as direct links between them can be established.² In collaboration with a.o. the CHI and NIWI institutes, experimental interfaces will be developed and made available to researchers.

2. Encoding examples

XML is an encoding technique which basically uses labels to describe relevant bits of information in a file which is otherwise best described as 'plain text'. Indeed, XML is often characterised as 'self-describing data'. I will provide some examples based on the emblem *Sit in amore reciprocatio* in Vaenius's *Amoris divini emblemata* (fig.1)³ In the illustrations you'll notice the motto, the quotations (with brief bibliographical references), the epigrams in Spanish, Dutch and French, and a pictura; you may recognise *Amor Divinus* and *Anima*, the soul, shooting their arrows at each other.

Let us look, for instance, at the Spanish epigram. It consists of three lines, abbreviated *l*, in what we call a linegroup, or *lg*. A simplified XML encoding would look like this:

```
<lg>
  <l>Ama à Dios de coraçon,</l>
  <l>Paraque se satisfaga</l>
  <l>Que amor con amor se paga.</l>
</lg>
```

What we have done here is placing labels, or, technically, *tags*, around fragments of text which we want to identify as being a fragment of a particular type or kind. Each pair of tags (<l>, </l>) defines an *element*. Elements may contain plain text (the *l*-element is an example) or other elements (the *lg*-element contains *l*-elements), or both. The structural relation between the epigram and its lines is mirrored in the relation of inclusion between the *lg*-element and the *l*-elements.

² See Section 5.

³ Otto Vaenius, *Amoris divini emblemata* (Antwerp 1615), Utrecht, Emblem Project Utrecht. At: <http://www.let.uu.nl/emblems/html/v1615front.html>.

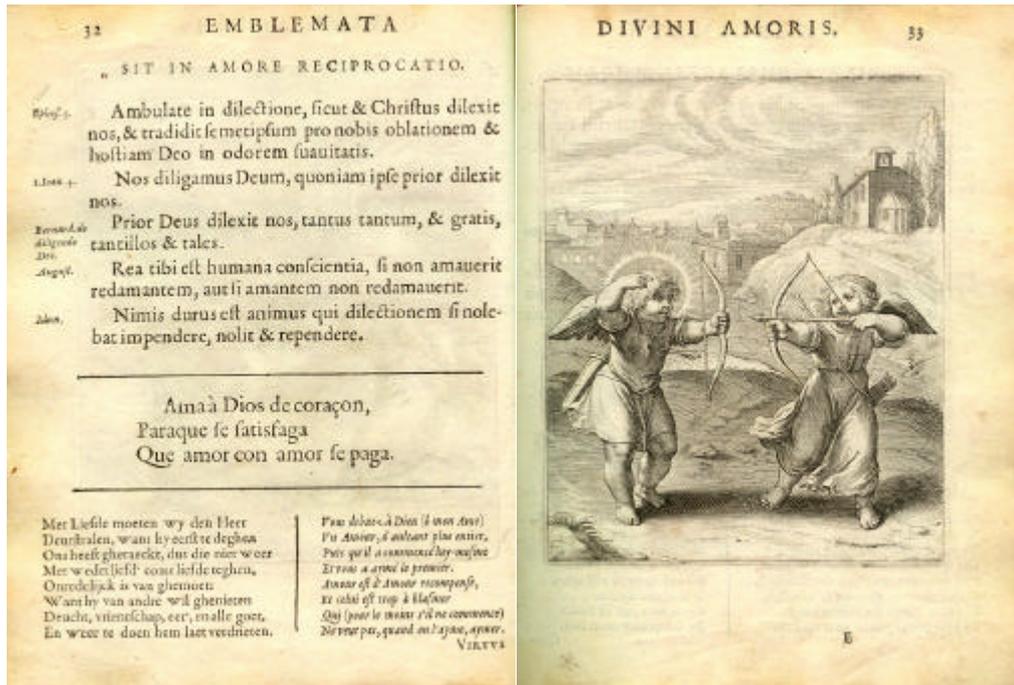


Figure 1 Otto Vaenius, *Amoris divini emblemata* (Antwerp 1615), p. 32, 33

In a similar way, we can encode one of the quotations:

```
<cit>
  <bibl>August.</bibl>
  <quote>Rea tibi est humana conscientia, si non amauerit
    redamantem, aut si amantem non redamauerit</quote>
</cit>
```

Within a *cit*-element, we have grouped the actual text of the quotation (the *quote*-element) and the relevant bibliographical reference (the *bibl*-element). Grouping the *quote*- and the *bibl*-element together makes it clear the two pieces of text are related; giving each text fragment it's own label clarifies the fragments' roles in their environment.

In the previous examples, we have seen how the structure of the emblem's smaller constituents can be made explicit using XML encoding. The same technique can be used at the emblem level and even at the level of the emblem book as a whole.

In the case of *Amoris divini emblemata* the structure of the emblems might be encoded like this:

```
<text>
  <div type="motto" lang="lat">...</div>
  <div type="subs" lang="lat">...</div>
  <div type="subs" lang="esp">...</div>
  [other epigrams...]
  <div type="pict">
    <figure entity="ref-to-pic"/>
  </div>
</text>
```

What we have here is a *text*-element containing a *div*-element (for ‘division’) for each of the emblem’s structural constituents: motto, quotations, epigrams and *pictura*. *Attributes* are used to specify the type of division and its language. The division for the *pictura* contains a *figure*-element which refers to a file (in e.g. jpg-format) containing an image of the *pictura*.

The book’s structure is again encoded similarly, using a *group*-element to group the emblem text-elements, and similar elements holding the preliminary texts (dedication, prefaces, etc). This *group*-element will thus contain all of the text from the emblem book. Adding a separate element which includes meta-information (title, author, technical information about encoding and scanning, etc) is the last step of encoding the book’s structure.

Besides structure, many other aspects of the emblem book can be described using the same techniques. For now, I just mention some of them: typography (italics, font sizes, etc.), page breaks, internal and external references, etc. Other possibilities include encoding editorial enhancements of the text: normalisation of spelling, correction, expansion of abbreviations, but also annotations, references to literature, etc.

3. Standards: XML, TEI, Unicode

Having seen a few basic examples of XML, we may want to take a step back and look at what we have been doing. The brackets (‘<’, ‘>’), the element (*l*, *lg*, *cit*, *text*) and attribute (*type*, *lang*) names have not been invented in the Emblem Project Utrecht. They are applications of two independent standards: XML and the *TEI Guidelines*. XML provides the syntax for the encodings we have seen, while the *TEI Guidelines* provide the vocabulary.

eXtensible Markup Language (XML)

XML is a standard endorsed by the W3C (the body which establishes standards for the World Wide Web) in 1998.⁴ It was originally developed for use on the WWW, as an alternative to HTML (HyperText Markup Language). HTML is mainly directed at the presentation of text, without additional semantic information. XML was meant to contain ‘self-describing data’, as said before, thereby enabling more sophisticated use than just display in a web browser.

The guiding principles in the design of XML were that XML should be (1) extensible (user communities should be able to create an XML vocabulary according to their needs), and (2) platform, vendor and application independent. XML files should work on any platform, XML should not depend on any particular software vendor’s standards, and XML should focus on content rather than presentation. The information in XML files should not reflect any specific program’s needs, but should rather mirror the structure of the underlying data.

One especially important aspect of XML is its use of Unicode.⁵ Unicode assigns a unique number to each character in almost all writing systems used anywhere in the world. Identifying characters using these unique numbers means that XML files can contain characters from any writing system (Roman, Greek, Arab, Chinese, etc.) without switching fonts. For portability among computer systems, programs and across international boundaries, this is of crucial importance.

Text Encoding Initiative (TEI)

⁴ The XML standard is available at <http://www.w3.org/XML/>.

⁵ More information at <http://www.unicode.org>.

The Text Encoding Initiative, now officially a consortium hosted by several European and North-American universities, has been working since 1987 on a set of guidelines for use in linguistic and literary applications.⁶ Its main product, the *TEI Guidelines*⁷, can be seen as providing one possible extension of XML. XML may be used in any kind of application: from a car parts inventory to financial reporting. The TEI Guidelines provide both a vocabulary for the use of XML in the humanities and guidance on how to apply this vocabulary.

The TEI Guidelines do not prescribe which features in a specific text should be encoded. They rather prescribe that if one wants to encode a certain feature, it should be done in such-and-such a manner. The application of the TEI Guidelines to the emblem corpus was designed at the Emblem Project Utrecht.⁸

4. Displaying XML

After the XML tagging has been applied to the text, the text lends itself to automatic processing. It may be searched intelligently, or it might be used as input to statistical analysis. But one of the main purposes of automatic processing is generating a suitable form of display.

Several possibilities exist. We could generate a paper-based edition of the emblem book (thus gaining our immortality after all!). Currently we have no plans to do so, but a very real possibility is to generate, for each emblem, a file in Acrobat's Portable Document Format, suitable for printing.⁹ At present in the Emblem Project Utrecht we generate HTML files, suitable for display over the web, from the XML files and the accompanying display rules defined in XSLT stylesheets.¹⁰

One of the main advantages of this procedure is that it allows for multiple forms of output. XML, as we saw before, is meant to encode data, not its presentation. Using different stylesheets, we can generate different HTML files. For example, when spelling normalisations have been added to the texts, one HTML file may contain the normalised version, another may contain the original version, and the reader may choose for himself which version he wants to read.¹¹ Another way we use this flexibility is in having a normal and a side-by-side view for each emblem.¹² Yet another application would be a search page which allows searches limited by language or by emblem constituent (e.g. search just the Spanish epigrams or Dutch mottoes).

5. Using XML in research

The foregoing paragraphs have shown that XML may serve as a basis for editions of emblem books that are both scholarly and flexible. To conclude, I would like to mention another

⁶ The TEI website is at <http://www.tei-c.org>.

⁷ Michael Sperberg-McQueen, Lou Burnard (eds.), *Guidelines for Electronic Text Encoding and Interchange*, Oxford, TEI Consortium, 2001. At: <http://www.hcu.ox.ac.uk/TEI/Guidelines2/index.html>.

⁸ The EPU Guidelines are available at <http://www.let.uu.nl/emblems/html/tech.html>. Other projects are free to use those parts of the EPU Guidelines that may be useful to them.

⁹ The XSL:FO standard describes an XML-vocabulary aimed at defining lay-out characteristics. A piece of software called a 'rendering engine' can then be used to generate (e.g.) a file in PDF-format.

¹⁰ The XSLT standard describes a language for defining transformations of an XML file into (e.g.) HTML. XSLT processors are programs which take as input an XML file and what is called an XSLT stylesheet and generate the desired output format.

¹¹ Using a 'Preferences' menu option.

¹² To view this, click 'compare' on any EPU emblem page.

application of XML which may be quite useful in the scholarly context. This is the use of XML to encode the findings of scholarly research.

This feature may be used to encode classifications of the emblems or emblem constituents. For example, the late Santiago Sebastián López, in his study of *Amoris divini emblemata*, classified the book's emblems into four groups (one group comprising general characteristics of divine love, and one for each of the three stages of the soul's ascent towards God).¹³ Using a simple keyword mechanism, classifications like these may be encoded into the XML file. The classification can then be used for searching purposes.¹⁴

A more complex example will show some of the more advanced possibilities. The example is part of a larger, as yet unfinished, investigation into the relation between the allegorical and typological modes of thought in the emblem. As students of the emblem we have all probably grown up on the views of Schöne, that allegory is somehow something unemblematical.¹⁵ Then, later on, from Peter Daly we learned allegory is a possible mode of thought of the emblem, on an equal footing with the typological variety Schöne advocated.¹⁶ Daly anticipated the modes may be combined or mixed.¹⁷ The aim of the larger investigation is to give an exact description of how this intermingling of modes works in books like *Amoris divini emblemata*.



Figure 2 Otto Vaenius, *Amoris divini emblemata* (Antwerp 1615), p. 39

¹³ Santiago Sebastián López, 'La visión emblemática del amor divino según Vaenius', *Cuadernos de Arte de la Fundación Universitaria Española* 2 (1985) pp. 3-51.

¹⁴ See <http://www.let.uu.nl/emblems/html/v1615search.html>.

¹⁵ Albrecht Schöne, *Emblematik und Drama im Zeitalter des Barock*. Zweite, überarbeitete und ergänzte Auflage, München, C.H. Beck, 1968, p. 33.

¹⁶ Peter M. Daly, *Emblem Theory. Recent German contributions to the characterization of the emblem genre*. Nendeln/Liechtenstein, KTO Press (Wolfenbütteler Forschungen, Band 9), pp. 78-99.

¹⁷ Daly, *op. cit.*, pp. 99-107.

Let us look at the *pictura* of *Superna respicit* (fig. 2). We see *Amor divinus* pointing the Soul towards the sunflower. It is an example of a more general phenomenon, where the allegorical figures point each other at an object of a typological nature - the sunflower is of course a prime example of this. So, what one would like to do is to give a formal description of the acts of pointing in the *pictura*. Now, in any act of pointing, there are three objects or persons concerned: person A is pointing person B at object C. Therefore, we need to describe three-term relationships: who points whom at what.

Here, we will not go into technical details: it may be accomplished by a list of objects, and a list of all occurrences of pointing. From the list of occurrences we refer to the list of objects to indicate who does the pointing, for whom it is intended and at what it is directed. Once the lists have been made, they can be used for searching. Questions like ‘what objects is *Amor divinus* pointing at?’, ‘to what objects is the Soul being pointed’ or ‘where is the sun being pointed at?’ now can be answered easily.

The same analysis may then be applied to related actions, such as one person holding up an object for another’s inspection; two persons together holding a single object, etc. These objects should then be analysed according to their secondary metaphorical or metonymical meaning, at which point the analysis will have to take into account the textual representation of these objects. It is an investigation which we hope to continue in the coming years.

An analysis like this is clearly beyond what one expects of a paper-based edition. I do not mean to imply a digitised emblem book should necessarily venture into this type of investigation. What I do want to stress is that a digital edition provides a natural environment for research, and may be integrated into scholarly research in ways we never dreamed of before.

Using XML, then, will not make us immortal. But in its separation of content and presentation, thereby enabling multiple views on the same content, in its independence of computer platform and proprietary software, XML represents a big step towards enduring digitisation of scholarly content. In my view, it will provide the platform of choice for digitisation of and research into the European emblem.