Digital Editions and Diplomatic Diagrams

Christine Roughan
College of the Holy Cross
PO Box 2294
Worcester, MA 01610 USA
+1 (203) 505-5042
cmroug14@g.holycross.edu

ABSTRACT

In this paper, the Archimedes Project at the College of the Holy Cross presents an approach to digital, diplomatic renditions of Greek mathematical diagrams as preserved in ancient and medieval sources. The team creates XML-based vector images from scalable vector graphics (SVGs). The resulting tracings are composed of data on the geometric shapes that make up the diagram, the textual labels that accompany them, citable identifiers, and editorial annotations. The relationships between the diplomatic diagram and the diplomatic text and images of the physical manuscript are modeled with the CITE architecture developed by the Homer Multitext project (http://www.homermultitext.org/hmt-doc/cite/index.html).

This approach creates machine-actionable data out of raster images provided by digital photography. The team shows how they have applied this method to diagrams from Codex Bodmer 8 (16th century) and Codex C (10th century and preserved in the Archimedes Palimpsest).

Categories and Subject Descriptors

I.7.2 [**Document and Text Processing**]: Document Preparation – *format and notation, index generation, markup languages, multi/mixed media, standards.*

General Terms

Management, Documentation, Performance, Design, Experimentation, Standardization, Theory, Verification.

Keywords

Classics, Greek, Mathematics, Diagrams, Figures, Diplomatic, Transcriptions, Digital Humanities, Medieval, Manuscripts, Codices, Archimedes, Vectors, Markup, Standards, CITE, SVG

1. INTRODUCTION

Greek mathematical documents are composed of two types of information: the text and the diagram. These two pieces have stood alongside each other since they were first authored – they are scattered across print editions, handwritten manuscripts, and fragments of papyrus.

It is important to note that unlike what is found in many modern mathematical texts, in Greek mathematics the diagram does not serve a merely illustrative purpose. It is a crucial part of the argument: without their accompanying diagrams, the texts of most propositions are revealed to be incomplete and their logic unclear. This is not a fault of the text. Diagrams contain essential information. [1]

It would therefore be irresponsible for the editor of a Greek mathematical work to neglect the diagrams in his edition. As mathematical documents receive digital representation, it is necessary for the *entire* document to be represented, diagrams as well as text. Photographic facsimiles already easily preserve diagrams in the same manner that they preserve text, but the same cannot be said for other digital forms such as diplomatic transcriptions. Texts can be handled with TEI-compliant XML; a similar standard does not yet exist for diagrams.

The team at the College of the Holy Cross approaches this problem through its work on two manuscripts of Archimedes, the tenth century Codex C (as preserved in the Archimedes Palimpsest) and the sixteenth century Codex Bodmer 8.

2. PAST TREATMENTS OF DIAGRAMS

In considering how diagrams should be handled digitally, it is informative to consider how they have been handled by past editors. Recent scholarship on Greek mathematical diagrams would agree that their treatments in print editions have been lacking. Where scholars armed with the methods of textual criticism carefully pored over each individual word in a manuscript, there existed no corresponding approach for diagrams. [2]

The diagrams drawn by medieval scribes in mathematical manuscripts would be unfamiliar to a reader acquainted only with print editions. An example can be seen in Figure 1 below, which compares the diagram for *On the Sphere and Cylinder I*'s proposition 33 in the most well-known critical edition of Archimedes and in one of the most famous and important manuscripts of the mathematician.

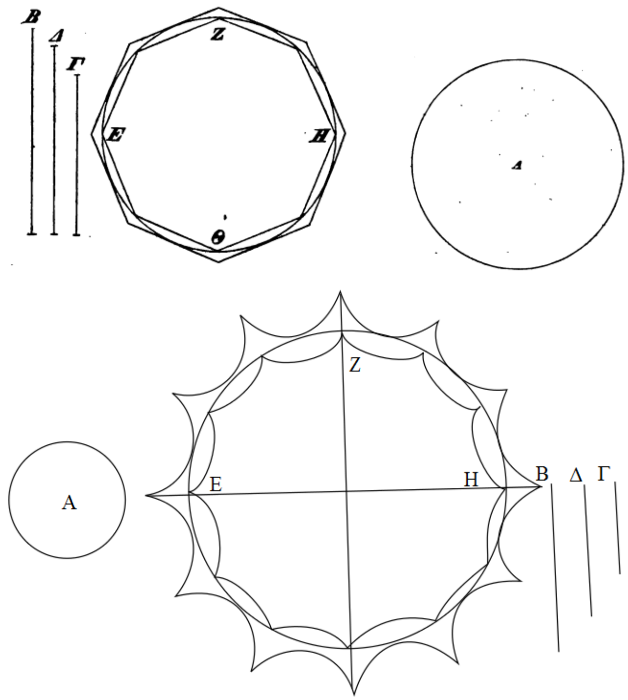

Figure 1. Above: proposition 33's diagram in Johan Ludvig Heiberg's critical edition. Below: a tracing of Codex C's corresponding diagram.

Disregarding the differing arrangement of shapes and the missing Θ (which might have been present in Codex C but lost to damage), there are still conspicuous differences in how the circumscribed and inscribed polygons are drawn. Codex C constructs them from arcs and gives them twelve sides rather than eight. Of course, this is a comparison between a critical edition and an individual manuscript, so it is not unexpected that the two do not match. Codex C, however, is not unique in drawing polygons in this way (the same curved polygons also appear in Codex Bodmer 8, which descends from Codex A). In Heiberg's edition, a medieval method of representing diagrams is abandoned in favor of a modern method.

Heiberg consulted the Archimedes Palimpsest in the creation of his critical edition, so the difference between the diagrams is not due to his unfamiliarity with Codex C. Rather, he has chosen to modernize the diagrams. The text calls for inscribed and circumscribed polygons, so Heiberg's edition draws shapes which are clearly recognizable as polygons. In the case of the text, a critical apparatus makes transparent the fact that Heiberg's critical edition departs from the manuscripts. No critical apparatus exists for diagrams, and this is true across nearly all editions of mathematical works.[1]

Older print editions do not provide historians of science and mathematics with the appropriate tools to analyze diagrams. Diagrams which are redrawn to fit modern understandings of how they 'ought' to be represented might serve a certain purpose, but they can reveal no information about the transmission of mathematical diagrams from their authors to the present day.

3. DIGITAL EDITIONS

When mathematical documents are digitized, diagrams should receive the same attention the text does: both are vehicles of the mathematical argument. In the case of diplomatic transcriptions, it would therefore be unwise for an editor to ignore the diagrams, to offer only modernized renditions, or even to supply only image selections from digital photography or other raster images.

For a comparison: digitally, diplomatically transcribing the text of a manuscript accomplishes certain goals. Such a transcription makes the text in the manuscript more accessible to those unfamiliar with the manuscript's paleography and more accessible when it is faded or obscured. It also makes the text in the manuscript machine-actionable and allows for annotations to mark certain features (personal names, unclear text, expanded abbreviations, etc.).

A digital, diplomatic diagram transcription therefore should (1) make the diagram in the manuscript more accessible, (2) make the information contained in the diagram machine-actionable, and (3) allow for editorial annotations.

4. A METHOD FOR TRANSCRIBING DIAGRAMS

In order to accomplish the above goals, the Archimedes team at Holy Cross uses the open graphics editor Inkscape to create encoded scalable vector graphic (SVG) files. The resulting diplomatic diagram tracing preserves the appearance of the diagram as it stands in the manuscript and also provides a digital framework for the encoding of information.

Over the past several months, the team at Holy Cross has focused its efforts on *Sphere and Cylinder* as preserved in Codex C and Codex Bodmer 8. Fifteen diagrams exist alongside *Sphere and Cylinder* in Codex C; thirty-one exist in Codex Bodmer 8. The remainder were not lost but rather were never drawn in the spaces left for them. The methods described below have been applied to all of the diagrams preserved for *Sphere and Cylinder I* and *II* in those two manuscripts.

4.1 Diplomatic Appearance

4.1.1 Geometric Elements

One way to understand the structure of a diagram is to break it down into its component shapes. The diplomatic diagram transcriptions therefore record the each geometric piece that makes up the figure. In the SVG file, each geometric object is defined with a machine-readable <path/> element and given a unique identifier with the label attribute.

The shape defined by the <path/> element is as exact a representation of the shape drawn in the manuscript as possible: it is a tracing produced from the digital images. This requires very precisely-taken photos, since a picture taken from an angle or a folio that is not laid out perfectly flat will result in a warped diagram tracing.

The <path/> element allows the editor to trace any shape using a combination of straight lines, circular arcs, and Bézier curves. Lines, triangles, rectangles, polygons, circles, and other more complicated shapes can all be easily represented. The tracing is

[1] Reviel Netz's *The Works of Archimedes: Translation and Commentary* (2004) is a recent exception and attempts to provide an apparatus for diagrams using written explanations and occasional thumbnails showing alternate diagrams.

well able to preserve the irregular arcs present in Figure 1, for instance, or to handle the jagged line that is present in the triangle in Figure 2.

The `label` attribute, besides providing a unique identifier, also records what kind of shape is represented in the diagram. The shape which is drawn, after all, might not perfectly match by modern standards the shape which is intended in the text. This is most clearly seen in the polygons of Figure 1. The scribes in both Codex C and Codex Bodmer 8 customarily drew polygons with curved sides rather than straight ones, but a computer, drawing the shape laid out in the `<path/>` element, would not recognize such a shape to be a polygon. The shape that is intended by the text is instead identified within the `label` attribute, leaving the geometrical primitives of the vector image to trace precisely the image which appears in the manuscript.

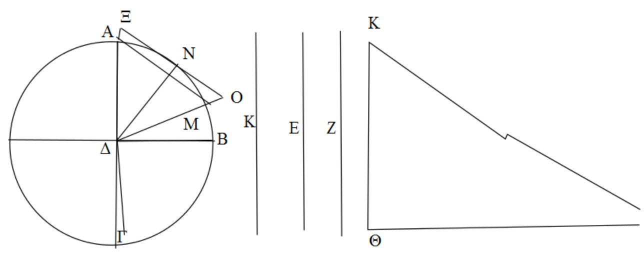

Figure 2. A tracing of *Sphere and Cylinder I* proposition 4 as it appears in Codex Bodmer 8.

Figure 2 provides a slightly different example: the text makes it clear that the jagged line Κ[Λ] is meant to be a straight one, but the scribe somehow erred in his drawing. Again, a computer drawing the shape described in the `<path/>` element would not properly identify what shape is represented in the diagram without outside help. The `label` attribute identifies the correct shape.

The Archimedes team therefore structures each unique identifier with the format `[shape][labels]`. In the case of circle ΑΒΓ, for instance, the unique identifier would be `circleΑΒΓ`. The standard used by Holy Cross is to include all the labels the text uses for that particular shape and to arrange them in alphabetic order. If a geometric object is included in the diagram but never constructed or mentioned in the text, it receives an identifier without any labels.

4.1.2 Labels

Labels serve as the markers that link text to diagram, and so the diplomatic transcription of the diagram must also include the labels alongside the geometric elements. Labels are included in `<text/>` elements and are simply the appropriate Greek letter.

The labels that are included in the tracings are exactly the labels that appear on the diagram, despite any disagreements there might be between the labels of the diagram and the labels discussed in the text.

Occasionally diagrams include extra textual information, as seen in Figure 3. Like labels, these strings of text are included in the `<text/>` element.

To parallel the approach used for geometric objects and to distinguish between labels and strings of text, each also receives a unique identifier with the `label` attribute. Label Δ has the identifier `labelΔ`, whereas a string of text is identified simply as `string1`, `string2`, etc. In cases such as Figure 3, where the same letter is used for multiple labels, they are distinguished by adding a number to the end of the identifier (`labelA1`, `labelA2`).

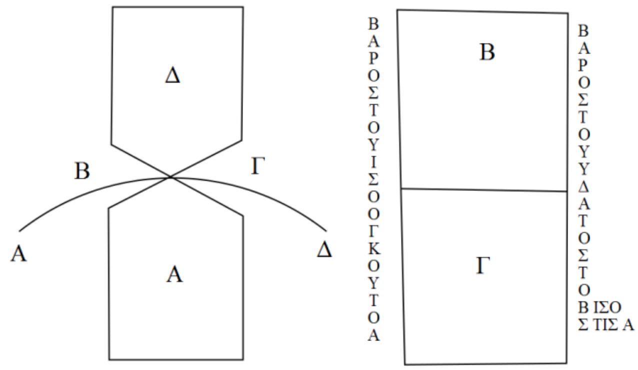

Figure 3. A tracing of *On Floating Bodies I*, proposition 6, as it appears in Codex C.

4.2 Digital Framework

4.2.1 Scalable Vector Graphics

The diagram tracings included in Holy Cross's edition are SVG files, an XML-based vector image format. The tracings incorporate data on the geometric shapes that make up the diagram and the labels that accompany them, as discussed above. The vector format allows the components of the diagram to be analyzed in ways which would be impossible for raster images. Like the XML used for texts, these SVG files also include editorial annotations.

4.2.2 Editorial Annotations

The physical diagrams which the SVG files trace are not all perfect: some are faded, obscured, or partially lost. Occasionally additions were made to diagrams or the scribe erred and corrected his mistake. The diagram tracings therefore encode all of this information to provide as complete a picture as possible. The `<desc/>` element, contained within the `<path/>` or `<text/>` element, includes editorial information.

A label or a geometric object can be `visible`, `unclear`, or `reconstructed`. When the object is absolutely certain (as it often is in Codex Bodmer 8), it is marked as `visible`. The current state of Codex C means that diagrams are often obscured or partially lost. A shape which is legible but visually faded or obscured in part receives the marker `unclear`: a line for which the two endpoints can be seen but not the middle, for example, would be marked as such. A shape which is known to have existed in the manuscript but which is too illegible for a confident tracing would be marked as `reconstructed`. A scenario where a quarter of a circumscribed polygon is faded and completely missing in parts would receive such a marker: unlike the case with the uncertain line, there is not enough information to be sure that the missing arcs of the polygon followed the specific path drawn.

There are some cases of diagrams where scribal corrections are apparent. Having drawn a line incorrectly, the scribe might scrape away the error and draw in the correct line, as in Figure 4. Such a scenario requires `sic` and `corr` markers. The same is true for labels: corrections are encoded with `sic` and `corr`. If an addition in the form of a geometric object or label is made, it is marked with `add`; similarly, deletions are marked with `del`.

Other miscellaneous markup includes using a <w/> element around complete words when parts of words cross over onto multiple lines, as in Figure 3. Proposition 6 of *Floating Bodies I* is an interesting case: the strings of text reference labels A and B, and so these must be tagged as labels. Additionally, geometric objects and labels can be grouped together for assorted purposes by using the <g/> element.

Therefore, where the state of the diagram is analogous to something which might be found in the text for which there is markup defined by the Text Encoding Initiative, the Archimedes Project at Holy Cross will use the same vocabulary.

Like XML for texts, this markup semantically encodes important information but neglects to comment on the presentation of that information. The editor can later use CSS to make unclear sections grey, for instance, or to make erasures not appear in the final diagram.

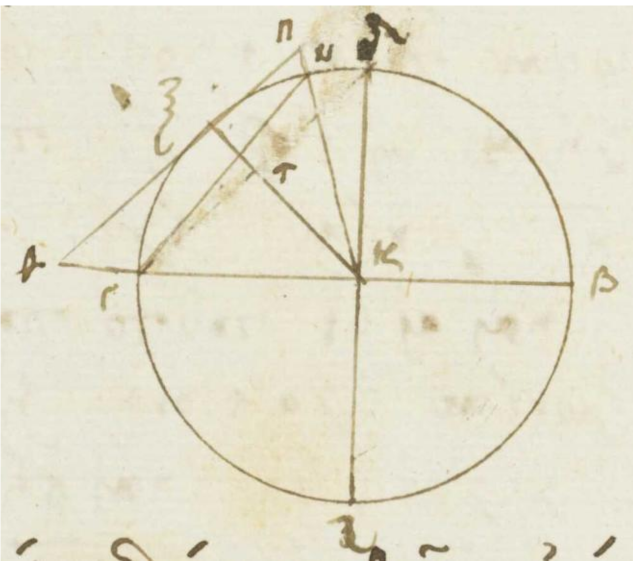

Figure 4. Detail from Codex Bodmer 8, showing an erasure. Cologny, Fondation Martin Bodmer, Cod. Bodmer 8, f. 4v (www.e-codices.unifr.ch).

5. CITE Architecture

The CITE Architecture was developed for work with the Homer Multitext project, but has since been applied to a range of other projects that deal with digital editions. CITE refers to Collections, Indexes and Texts and provides a means of citing and linking discrete texts and objects. Each text possesses a Canonical Text Service URN and each object a Collection URN which identifies the unique resource. These URNs can be associated with each other and with other data through the CITE Index, which organizes data in triples consisting of two nodes and an arc or verb in a graph. [3]

In Holy Cross's edition, both diagrams and the gaps left for undrawn diagrams possess URNs: they have the format `urn:cite:ap:apdiagrams.#` (Codex C) and `urn:cite:ap:figures.#` (Codex Bodmer 8).

The texts of *Sphere and Cylinder* are preserved in their entirety for both manuscripts. These possess the URNs `urn:cts:greekLit:tlg0552.tlg001.ap` (Codex C) and `urn:cts:greekLit:tlg0552.tlg001.cb8` (Codex Bodmer 8).

In Holy Cross's edition, these diagrams and texts are represented in two ways: digital photography of the manuscripts themselves and digital diplomatic transcriptions.

The digital diplomatic transcriptions have been discussed above: they possess the URNs `urn:cite:ap:aptrace.#` (Codex C) and `urn:cite:ap:cbtrace.#` (Codex Bodmer 8). The diagrams' representations in the digital photography are cited by referring to the coordinates for a region of interest on an image.

Particular diagrams are linked to the section of text they accompany via a <figure/> element in the text transcription XML, and this relationship is explicitly defined with triples in the CITE Index. The relationship is as follows: a diagram `illustrates` a proposition and a proposition `isIllustratedBy` a diagram.

The geometric objects discussed in the text and drawn in the diagram have two sets of labels that refer to them: the ones with which the text names them and the ones marking them in the figure. Individual geometric objects are linked to the first set through their identifiers, which come from the labels used in the text (circle ΑΒΓ is `circleΑΒΓ`, as explained above). They are linked to the second set via triples which define that a particular geometric object `hasLabel` A, for example.

Ultimately, the CITE Architecture ties together the whole digital edition, and therefore plays an important role in linking the diplomatic diagram to the evidence for its transcription as well as linking the diagram to the text and to other representations of the same diagram.

6. FURTHER QUESTIONS

6.1 The Construction of the Diagram

Where the text is sequential, the diagram is static: such is the finished product with which the reader of Greek mathematics is presented. However, the Holy Cross team is interested in determining to what extent scribes mechanically copied the diagrams, or, alternatively, to what extent they followed the construction of the diagram as laid out in the text.

To investigate this question, in the Archimedes Project at Holy Cross part of the diplomatic diagram's SVG file applies the structure of the construction from the text to the diagram. This data is in addition to the diplomatic transcription of the diagram since it looks beyond the figure to the text.

The construction that is laid out in a proposition is broken down into steps. This information is included in the diagram tracing through the group element <g/>. The geometric objects and labels that are set out in the first step of the construction are contained within a <g/> element with the attribute `label="tracesteps01"`, those set out in the second step are within a group identified as `tracesteps02`, etc. The steps of the construction provide an alternative framework through which to study the diagram.

The steps of the construction recorded in the diagram tracings also link to the text that declares those steps. Each step in the tracing has an identifier `tracesteps#`; triples link those steps with the appropriate section of text. Thus a section of text `isIllustratedBy` a step of the tracing and that step of the tracing `illustrates` that section of text.

It is possible to analyze this data to seek patterns within treatises: does Archimedes generally follow the same progression? How do constructions most commonly begin? Do constructions increase in complexity with later propositions? Additionally, it is interesting to note when the scribe included geometric objects that are not called for in the construction.

6.2 Transmission of Labels in Diagrams versus in Text

The creation of diplomatic transcriptions of text and diagrams results in a record of the labels which appear in the text and those which appear in the diagram. At the most basic level, these two records can be compared to see if the labels used in a certain proposition are the same as the labels used in its diagram. The textual labels and diagram labels can also be compared for a particular geometric object.

By looking at the labels in the text XML and nearby words for context, a computer could automatically narrow down and perhaps determine which geometric object is being discussed at each point in the text.

These data sets allow the investigation of various questions about the labels of Archimedes. Where do the labels of text and diagram disagree? How many different ways does the text refer to the same geometric object? Are there patterns in the arrangement of labels (ΑΒΓ vs ΓΒΑ) for the same geometric object? How often are geometric objects fully labeled (ΑΒΓ vs ΑΒ)? Even if all the same labels are used between the figure and the text, are they used for the same geometric objects, or do they disagree?

6.3 Diagram Overlays

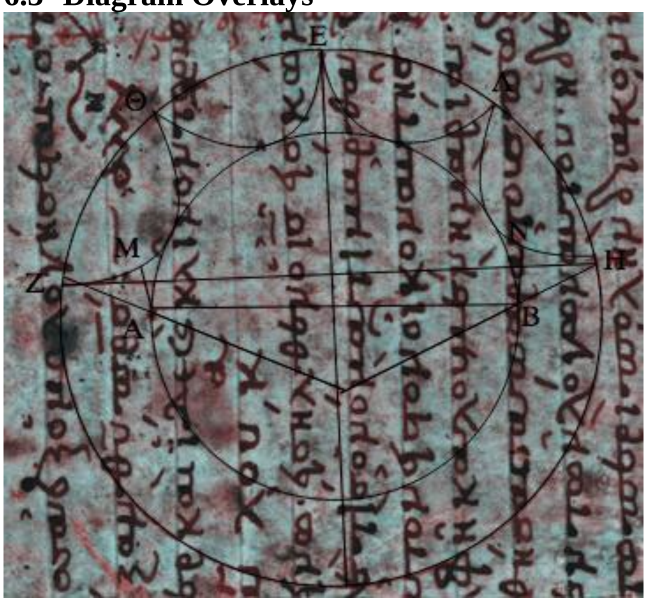

Figure 5. *Sphere and Cylinder I* proposition 39: overlay of a tracing on the diagram's image from the photography of Codex C.

The Archimedes Project at Holy Cross includes in its digital edition not only diplomatic renditions of diagrams, but also citations to the regions of interest on the digital photography those diagrams appear on. Since the diagram's diplomatic transcription comes from tracing the digital photography, the SVG file possesses the same dimensions as the region of interest in the manuscript. This and the CITE Architecture allows the editor to easily overlay the tracing directly on its manuscript region of interest; an example can be seen in Figure 5.

Such an ability is worthwhile mainly in heavily obscured manuscripts such as the Archimedes Palimpsest, where the diagram is otherwise difficult to make out. CSS can be used to change the appearance (such as the color) of the tracing, making it even more apparent.

7. CONCLUSION

Diagrams are key in mathematical texts, and the method described here which research at Holy Cross has applied to manuscripts of Archimedes could serve as an effective approach for treating Greek mathematical diagrams in general. Historians of science and mathematics are only just recently turning their attention to the historical transmission of the diagram, and there are a wide range of examples to consider spread out across a multitude of authors and hundreds of years. Digital formats, when encoded meaningfully, can be well-suited for quantitative analyses. The method applied here to two manuscripts of Archimedes could yield interesting findings if applied to the larger mathematical tradition of Archimedes and if applied to Greek mathematical works more generally.

8. ACKNOWLEDGMENTS

We thank the Digital Archimedes Palimpsest for providing digital access to Codex C as preserved in the Archimedes Palimpsest. We also thank e-codices, the Virtual Manuscript Library of Switzerland, for providing digital access to Codex Bodmer 8.

9. REFERENCES

[1] Saito, Ken. 2009. Reading ancient Greek mathematics. in Robson, Eleanor and Stedall, Jacqueline ed. *The Oxford Handbook of the History of Mathematics*. Oxford University Press, Great Clarendon Street, Oxford.

[2] Netz, Reviel. 2013. *Leaping out of the Page: The Use of Diagram in Greek Mathematics*. Lecture. The British Academy, London. http://www.britac.ac.uk/events/2013/Leaping_out_of_the_Page.cfm.

[3] Blackwell, Christopher and Smith, Neel. 2012. The CITE architecture. Web. http://www.homermultitext.org/hmt-doc/cite/index.html.