

Special Instructions for Euclid 1966 (Addendum to Data Entry Specs 1.1.2)

Wolfgang Schmidle, Klaus Thoden, Malcolm D. Hyman
Max Planck Institute for the History of Science, Berlin, Germany
12th December 2008

1 Data Entry

Data entry should be in UTF-8, making use of the complete character repertoire found in Unicode version 5.1.0, including characters in the following blocks when applicable:

- CJK Unified Ideographs Extension A (U+3499 - U+4DFF),
- CJK Unified Ideographs Extension B (U+20000 - U+2A6DF),
- CJK Compatibility Ideographs Supplement (U+2F800 - U+2FA1F).

If you experience difficulty with Chinese characters from the blocks in the supplementary plane (SMP), please let us know. We can provide assistance by providing fonts containing these characters and other technical assistance.

2 General Markup

The conventions for typing page breaks and numbers, running heads, headings, paragraphs, figures, and captions as specified in the standard Data Entry Specs apply, as well as the conventions for indicating illegible characters and uncertain readings (see section 2.7.1 of the Data Entry Specs).

2.1 Paragraphs

Outdented paragraphs (i.e. paragraphs where the first character is printed one space higher on the page) are marked by `<p x>`.

PLEASE NOTE: Examples for outdented paragraphs as well as the rules in section 3 can be found in the whole-page example in section 4.

3 Chinese Characters

3.1 Punctuation

Type the punctuation to the right of characters.

3.2 Spaces

Type spaces in Chinese text as the IDEOGRAPHIC SPACE character U+3000.

3.3 Small Characters

Strings of small characters are marked by `<sm>` `</sm>`. Indicate half-column breaks by `\\`.

4 Whole-Page Example

This is a transcription of Euclid p. 23 (i.e. scan 0023.jpg).

```
<pb><rh>幾何原本 卷一之首</rh>
<h>幾何原本第一卷之首<sm>界說三十六 求作四 \\ 公論十九</sm></h>
<p>泰西利瑪竇口譯</p>
<p>吳淞徐光啓筆受</p>
<h>界說三十六則</h>
<p>凡造論。先當分別解說論中所用名目。故曰界說。</p>
<p>凡歷法、地理、樂律、算章、技藝、工巧諸事。有度、有數者。皆依賴十府中。幾何府屬。凡論幾何。先從一點始。
自點引之爲線。線展爲面。面積爲體。是名三度。</p>
<h>第一界</h>
<p x>點者、無分。</p>
<p>無長短、廣狹、厚薄。 如下圖。<sm>凡圖十干爲識。干盡用十 \\ 二支。支盡用八卦八音。</sm></p>
<fig>
<desc>甲</desc>
<h>第二界</h>
<p x>線、有長無廣。</p>
<p>試如一平面。光照之。有光無光之間。不容一物。是線也。眞平眞圓相遇。其遇處止有一點。行則止有一
線。</p>
```

5 Figures

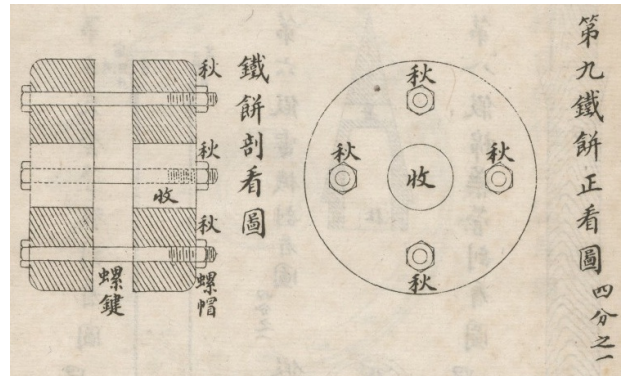
Where a figure occurs in the text, type a `<fig>` tag on a separate line. If you can identify a caption of the figure, mark it by `<cap>` `</cap>`. Additional text that describes parts of the figure is marked by `<desc>` `</desc>`.

Type the caption on a separate line after `<fig>`. A figure may have more than one description. Type each description on a separate line after `<fig>` and `<cap>`. If the same description is repeated in a figure, type it only once.

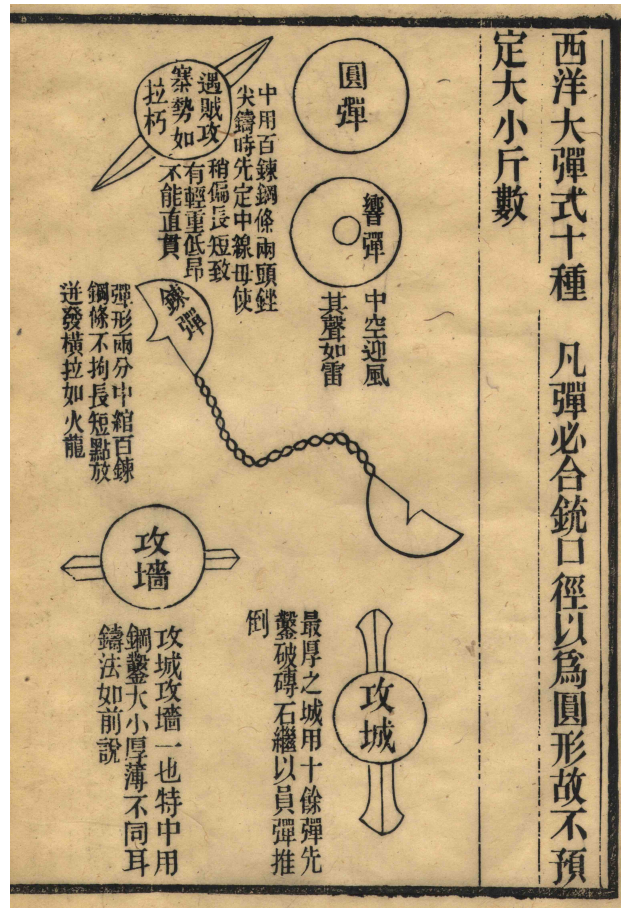
PLEASE NOTE: In the example in section 4, there is now an additional line `<desc>甲</desc>` after the `<fig>` tag.

EXAMPLES

<fig>
 <cap>第九鐵餅正看圖<sm>四分之一</sm></cap>
 <desc>秋</desc>
 <desc>收</desc>
 <fig>
 <cap>鐵餅剖看圖</cap>
 <desc>秋</desc>
 <desc>螺帽</desc>
 <desc>收</desc>
 <desc>螺鍵</desc>



<p>西洋大彈式十種 凡彈必合銃口徑以爲圓形故不預定大小斤數</p>
 <fig>
 <desc>中空迎風
 其聲如雷</desc>
 <desc>圓彈</desc>
 <desc>響彈</desc>
 <fig>
 <desc>中用百鍊鋼條兩頭銼
 尖鑄時先定中線毋使
 稍偏長短致
 有輕重低昂
 不能直貫</desc>
 <desc>遇賊攻
 寨勢如
 拉朽</desc>
 <fig>
 <desc>彈形兩分中縮百鍊
 鋼條不拘長短點放
 迸發橫拉如火龍</desc>
 <desc>鍊彈</desc>
 <fig>
 <desc>最厚之城用十餘彈先
 鑿破磚石繼以員彈推
 倒</desc>
 <desc>攻城</desc>
 <fig>
 <desc>攻城攻牆一也特中用
 鋼鑿大小厚薄不同耳
 鑄法如前說</desc>
 <desc>攻牆</desc>



<pb a><rh>泰西事物起原 <sm>第三章</sm></rh>
 (some text)
 <fig>
 <desc>第二大派</desc>
 <desc>希伯流語</desc>
 <desc>亞刺比亞語</desc>
 <pb b><rh>泰西事物起原 <sm>第三章</sm></rh>
 <desc>非尼西亞語</desc>
 (some text)

