

Workflow for Mechanics Project transcriptions

Author:	Brian Fuchs
Email:	fuchs@mpiwg-berlin.mpg.de
Date	2006.02.01
Version	0.3

1. Preparation

1.1 Tools

This workflow makes use of the following tools:

- **Text Editor:** BBedit (alt: [Jedit](#)).
-  MS Word **not** recommended!
- **Sub Table Editor:** [OpenOffice Calc](#) (= MS Excel)
- **Conversion tool:** <http://141.14.237.38:10014/cgi/convert>
- **XML Validator:** <http://archimedes.mpiwg-berlin.mpg.de/cgi-bin/archim/parse>

1.2 Registering the transcription

If the transcription is going into the Archimedes Repository, you will need to get an **id** from a project administrator.

2. Transcribing the text

2.1

Use the text editor (Jedit or BBedit) to transcribe the text. Follow the tagging specifications below.

2.2 Tagging Specifications

2.2.1 General

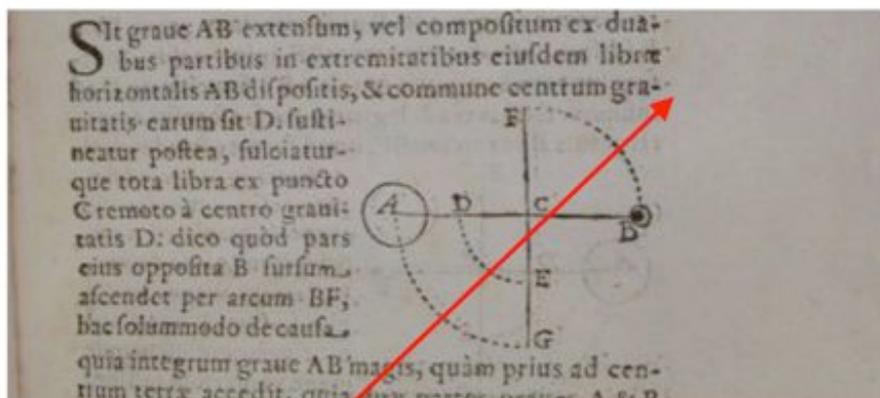
1. Encoding should be UTF-8.
2. Line-ends should be unix-style ("\n").
Note: MS Word: save as "text only".

2.2.2 What to tag

1. Place a return at the end of every line of text.
2. Use symbols to transcribe **all** accented letters. (See **Using a Substitution Table** below.)
3. At the start of each paragraph, place <p>.
4. At the end of each paragraph, place </p>.
5. At the start of every title, place <h>
6. At the end of every title, place </h>

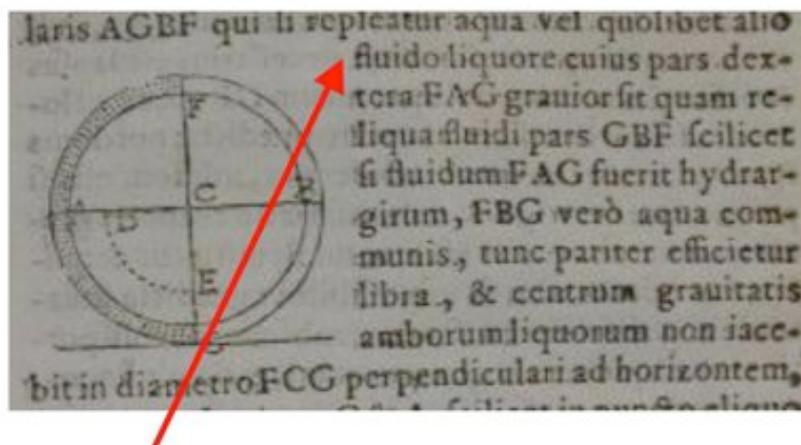
Note: sentences occur **inside** paragraphs and titles.

7. For every hyphen, type "­" or use a substitute symbol (see 2 below).
8. Place <i> at the start of italics; place </i> at the end of italics.
9. Place at the start of bold; place at the end of bold.
10. Marginalia **TODO**.
11. Figures: place <figure/> after the end of last line of text **before** the start of the figure.



Put <figure/> between *gra-* and *uitatis*.

Ex. 1 Figure with right indent.



Put <figure/> between *alio* and *fluido*.

Ex. 2 Figure with left indent.

2.2.3 What not to tag

- Sentences (normally tagged with <s>...</s>). These will be extracted interactively at a later stage.

3. Using a substitution table.

A substitution table serves two purposes. First, it can be used to avoid having to type frequently used combinations of letters—a feature also provided by most editors in the form of keyboard shortcuts / macros. Secondly, it provides a means of adding user-defined entities to the xml document. It is the second purpose that is important in this workflow.

3.1 What to put in a substitution table.

- User-defined entities.
- Other typing shortcuts (including those for standard entities).



All user-defined entities must be included in the substitution table.

3.2 How to make a substitution table.

Using a database or spreadsheet tool (Filemaker, Excel, OO [Calc](#)), make a table with 3 columns (see ex 1 below) . This table will be used by the conversion tool to replace abbreviations in the transcription (column A) with a substitute (column B). The description (column C) is used for orientation, record-keeping, and will eventually be added to the formal document definition once the file is converted to xml.

```
col1 col2 col3
abbreviation substitution description
```

In column 1, place any sign you use in transcription. In column 2, place whatever you want the sign to be replaced **by**. In column 3, place a short description of the purpose of the sign.

	A	B	C
1	<1>	æ	ae ligature
2	<2>	œ	oe ligature
3	<3>	&longs;	Long s
4	<4>	&mtilde;	m with tilde
5	<5>	á	a with acute
6	<6>	à	a with grave
7	<7>	é	e with acute
8	<8>	è	e with grave
9			

Ex 1. A substitution table in OpenOffice Calc (= MS Excel).

(The conversion tool will replace each of the signs in column A with the contents of column B; if column B contains an entity, then it will be added as an entity definition to the internal DTD of the document produced in step 4 below).

4. Save output

Save the transcription *and* the substitution table to the project repository at the end of each work session.

The **Filename** for the **transcription** should be the id of the work + ".txt". Thus if the id of a work

is **borel_demot_010_la_1670** the filename should be **borel_demot_010_la_1670.txt**.

The **Filename** for the **substitution table** should be the id of the file + ".csv". Thus if the id of a work is **borel_demot_010_la_1670** the filename of the sub table should be **borel_demot_010_la_1670.csv**.

The project repository for raw transcriptions of Mechanics Project texts and Archimedes Project texts is located at:

<http://archimedes.mpiwg-berlin.mpg.de/cvs-web/read/cvswebread.cgi/texts/archimedes/raw/>

You need a password to upload material to this repository. If you don't have one, you can get one from the repository [admin](#).



If you are following the **short version** of this workflow, you can stop here.

5. Prepare files for conversion

3a. Export the sign table from Calc / Excel as a csv (comma-separated-values) file.

6. Convert file

<http://141.14.237.38:10014/cgi/convert>

Input for text file is expected to be utf8 with unix line-ends.

The conversion service turns your file into an **xml** file.

Save this file locally, then validate it.

7. Validate

Validate using the validation link on

<http://141.14.237.38:10014/cgi/convert>

8. Adjust

Repeat steps 4-7 until you achieve a satisfactory outcome.

TODO:

instantiate this page elsewhere as a wiki.
description of conventions for step 1.