Modelling humanities data with TEI-XML

Scholarly editing and manuscript cataloguing in the digital age

Dr Katarzyna Anna Kapitan (katarzyna.kapitan@chartes.psl.eu)

Mid-Term Assignment (30%)

Submission Deadline: 30 October @ 23:59 (Paris time zone)

Submission Form: Folder on GitHub consisting of:

1. A ReadMe file with a description of the materials (250–300 words)**.**
2. A plain XML file template with structured list of elements, attributes, and attribute values defined with DTD.
3. Three samples of unprocessed data in jpg, png, pdf, or txt.

Detailed Instructions:

1. Choose the type of materials you want to work on for your final project. Focus only on one type of materials, for example, letters, poems, biographies, encyclopaedia entries, manuscripts.
2. Find three suitable samples you will use in your final project. The length of your text samples should be between 300 and 1000 characters each if you work with handwritten extracts which you plan to transcribe and encode, or between 1000 and 3000 characters each if you work with typewritten extracts that you will encode. Your samples can be, for example:
    - three printed manuscript descriptions which you want to structure in XML,
    - three manuscript fragments you want to identify and describe in XML,
    - three letters/charters/poems you want to edit,
    - three extracts of prose you want to semantically annotate.
3. Prepare three samples of your unprocessed data (jpg, png, pdf, txt, html) to be shared with the teacher; these can be, for example:
    - scans of catalogue entries you plan to encode,
    - images (facsimiles) of manuscript fragments you plan to catalogue,
    - plain text of the letter/charter/poem you plan to edit,
    - plain text of prose you plan to semantically annotate.
4. Prepare a short description of the materials and justification why you want to work with them and how you think using XML for data structurisation may benefit your project (250-300 words).
5. Using the TEI Guidelines, make a list of elements, attributes and attribute values that you think you will need for your project.
6. Create a basic encoding template with DTD.
7. Upload everything to GitHub and share your folder with your teacher before the deadline.

Final Assignment (50%)

Submission Deadline: 18 December @ 23:59 (Paris time zone)

Submission Form: Folder on GitHub consisting of:

1. A ReadMe file with a description of the materials (250–300 words)**.**
2. One ODD customisation file.
3. RNG and HTML files created from ODD.
4. TEI XML file (or files) with the samples encoded in a custom-made template, following the same structure and the same elements and attributes (validated with your custom RNG schema).

Detailed Instructions:

For your ODD file:

1. Generate an initial ODD file from Roma or oddbyexample and write a structured documentation of your customisation.
2. Complete the teiHeader as precisely as possible.
3. Present your project and its possible uses in the introduction.
4. Explain the choices you made regarding elements and attributes to allow for subsequent enhancement of the encoding.
5. Define at least 3 rules, which are well documented (at least by a description) and justified by your project:
    - A rule constraining the use of an attribute and its value(s) (<attDef>);
    - A rule constraining the sequence of certain elements (<content>);
    - A rule constraining the value of an attribute or the use of an element or attribute depending on its environment (<constraint> + <sch:rule>).

For your XML file(s):

1. Complete the teiHeader as precisely as possible.
2. Encode your samples and validate the encoding with your schema.