



Science Meeting - Scientific Report

Proposal Title:

Editing texts in a digital world: Text encoding and visualisation

Application Reference N°: 5912

Summary

NeDiMAH thematic workshop was held in Cluj-Napoca, Romania, 27.04.2015-02.05.2015.

“Nowadays, when everything is becoming digital, people are trying to develop the necessary instruments in order to transform real objects into non-material/digital objects. The question in this case is: *How much of the objects real significance is kept by translating it in digital form?* In humanities, for instance, programming languages are used to enable digital processing of texts, media and data.

The thematic Working Group *Information Visualisation*, part of the NeDiMAH Network, focuses on computationally- based methods for capture, analysis and visualisation of literary works, especially of unpublished personal letters. This workgroup is a way to learn how to use ICT tools in order to create digital collections that involve data research and content re-using.

Using the markup language of the Text Encoding Initiative (TEI), which is a standard protocol for the digital encoding of texts in the humanities, participants will learn in this workshop about methods of text encoding and the specific digital formats of these texts.

The workshop will also focus on data visualization. Maps are a very important tool, as their use can transform a literary or historical description into a visual representation.

As the TEI is the most relevant standard for scholarly editing and text corpora building in an evolving digital culture, this workshop will focus on text encoding and data visualisation.

Besides the theoretical framework the workshop will provide hands-on experience with TEI technology and letter corpus format.

This workshop is addressing early-stage researchers working in the field of humanities, they will be trained to translate encoded text into data visualization.”¹

¹ <http://www.nedimah.eu/events/editing-texts-digital-world-text-encoding-and-visualisation>

Description of the scientific content of and discussions at the event

During the first half of the workshop, the facilitator, Roman Bleier introduced topics ranging from introduction to basic markup, TEI standards – providing markup syntax and vocabulary needed to produce well-formed and valid XML - to TEI Guidelines and approaches to publish TEI texts.

Day 1

- What is markup language?
- Introducing the oXygen environment
- Introduction to XML

During the first day, the training sessions focused on markup language and using oXygen XML. Because the participants were mostly at a beginner level, with few exceptions, a comprehensive introduction to markups was needed, specifically what markups are and how they are used in the modeling of humanistic data. Further on, the facilitator started by offering a short overview of oXygen environment and the technological developments that led to XML. Theoretical and practical aspects of SGML (Standard Generalized Markup Language) international standard (ISO 8879/1986) were presented as meta-language for defining encoding language (markup). XML (Extensible Markup Language), an extension of SGML, appeared as a recommendation of W3C in 1998. More exactly, a model used for the various layers of abstraction can be represented in a markup language with a structure similar to XML. This allows for a description of data and their relationships independent of the platform. The applications of this model are particularly important when the modification and visualisation of a complex model needs to be done by several persons randomly located and in case of fast calculations that need high hardware resources (for instance, the building a 3D complex model starting from fine CT data).

Practical exercises were done with oXygen, a user-friendly editor versatile editor for XML, HTML, etc.

Links to materials

- [What is the markup language and introducing the oXygen environment](#)
- [Introduction to XML](#)
- [Document Modelling with XML](#)

Days 2 - 3

- Document modeling with XML
- Introduction to TEI
- Customize your TEI schema with ROMA
- TEI encoding of correspondence

The second day, the trainer introduced TEI, TEI P5 Guidelines, the structure of a TEI file, namespace, syntactic rules and semantic recommendations, TEI schema, ODD, etc. He then introduced ROMA and TEI customization with ROMA, proceeding by working with TEI and letter encoding for correspondence and postcard encoding, practicing the theory on the Transylvania Digital Humanities Centre letter collection that had been previously scanned, labeled and transcribed by our members.

Continuing with TEI, the workshop became very hands-on as participants worked on templates designed especially for letters, (the corpus selected for DigiHUBB project).

Working on the pre-generated templates was a great way of understanding the necessary

structure of the TEI code and how it can be modified in order to better suit researchers' needs.

On the third day of our workshop participants continued during the morning session with TEI Guidelines for manuscript description and representation of primary sources and other details regarding the document encoding.

The facilitator was very efficient and clear in explaining everything and making sure every question was cleared-up before moving on to the next idea. This assured that no one fell behind, or missed a step.

Links to materials

- [Introduction to TEI](#)
- [Customise your TEI schema with ROMA](#)
- [TEI encoding of correspondence](#)

During the second part of the workshop, facilitators Vinayak Das Gupta and Shawn Day introduced theoretical and practical aspects of data visualisation. Digital Humanities typically use various visual formats such as timelines, diagrams, charts, tables, and maps. These visual representations use GIS (Geo-Spatial Information Systems) as operational systems, which can encode various theoretical approaches. Maps are very important tools for Digital Humanities, transforming artistic, literary or historic descriptions into visual representations. Mapping customisation and exploitation by adding information may include spatial, temporal and conceptual viewpoints, recording experiences and expressing hypotheses encapsulating cultural values at specific historical moments.

- Vinayak Das Gupta: A brief introduction to data visualization
- Shawn Day: Visualization and storytelling; Constructing narratives using digital objects

Days 4-5

Practical exercises were done using Leaflet js Javascript library for maps, Omeka application for creating thematic research: items, exhibits and collections of various formats (photos, sound files, videos, documents).

During the fourth day, participants continued working with LeafletJS, changing maps styles, adding markers, connecting different points on a map, adding pop-ups for the pins on the map, adding images, clustering the pins on the map and so on.

The afternoon session of the fourth day was dedicated to Visualisation and Storytelling with Shawn Day. Case studies from other universities were discussed, together with the value of data visualization, visualizing time and space, Gephi, Nightlab, TimelineJS, JuxtaposeJS, StoryMapJS and so on.

During the fifth day, Shawn Day talked about more types of metadata and some common standards (Dublin Core, GIS Metadata, VRA Core, MODS, MPEG21). A quick introduction to Dublin Core was presented, with its main components and ways to indicate such details as: title, subject, description, source, publisher, contributor, right, format, language, etc. The afternoon session of the last day of our international workshop was dedicated to some tools used to see metadata in action, such as Zotero, Omeka, Scriptor, ScholarPress, Timeline Builder, Serendip-o-matic, Web Scrapbook, etc., with a focus on Omeka. Participants talked about how to add items and their metadata, how to edit them, what options the Settings menu offers, how to add additional plug-ins, etc. Participants went on building a narrative exhibit and planning an Omeka site.

Overall, because of the diversity of themes covered during the workshop, the participants could really get the feel of working on DH project from start to finish. From the intricate work of marking-up and encoding your texts, all the way to the more imaginative showcasing of the results.

Links to materials:

Vinayak Das Gupta: [A very, very brief introduction to data visualization](#)

Shawn Day:

- [Constructing narratives using digital objects](#)
- [Visualisation and Storytelling](#)

Assessment of the results and impact of the event on the future directions of the field

The workshop provided a complex set of digital techniques and tools for archiving and content analysis that now represents a strong foundation for the development of the scientific activity of DigiHUBB, the Transylvania Center for Digital Humanities. These techniques and tools bridge the skills and competencies of the interdisciplinary team of the hub. The participants practiced markup language and text encoding on a set of letters from the private collection of Adela Fofiu, a most valuable experience for the optimal creation, storage and analysis of various collections of social-historical documents in the future projects of DigiHUBB. These techniques and tools take content analysis on cultural artifacts to a very high level of comprehension and both theoretical and empirical perspective. DigiHUBB now has the capacity to promote digital humanities to a high level in the Romanian academic community, while also serving the intersection of the business, cultural and research environments.

Another important gain from this workshop was learning how to present DH information for an audience with and without technological background. On the one hand, humanist researchers working on text encoding have gained an insight into theoretical aspects and technology-based applications and will be able to understand what to ask from IT specialists. On the other, IT specialists have a better understanding of how technology can be used to design projects in TEI and data visualisation for humanities.

Besides expert guiding from the facilitators, the workshop also provided a great opportunity for peer-to-peer learning and expert guidance. The wide variety of disciplines and interests made it even more beneficial to all those involved. Below, there are some contributions written by the participants on the impact this workshop has had on their individual research.

“The NeDiMAH workshop on TEI and Data Visualization has been an excellent opportunity to strengthen and to develop my skills in working with historical newspapers. The TEI techniques will enable me: to create more complex archives of historical media discourses to mine the text of historical newspapers for more nuanced topics and phrasings, as compared to mainstream qualitative content analysis software (such as MaxQDA, NVivo, Atlas.ti, Tropes, etc.) to improve the codification and qualitative analysis of image and text, together

The data visualization techniques and tools offered me the opportunity to express research results on historical newspapers in a more compelling fashion, not only by design, but also through the novel perspective these techniques can bring on contents that I am all too familiar with. For instance, I am currently testing Timeline JS for observing and visualizing the content of front pages in historical newspapers, with the aim of better illustrating how front page topics and design change over 80 years. My interest is to see if this form of visualization is appropriate for exploring discursive patterns that construct intercultural conflict.

To continue, the mapping exercise in which we have used Leaflet JS is extremely useful at this point in my research on historical newspapers, as it enables me to map the periodicals published on the territory of Transylvania and Banat between 1860 and 1940. I will soon present this map at the 5th Workshop of the Research Programme “Clash of Civilizations or Peaceful Co-evolution? Intercultural Contact in the Age of Globalization” - Explorations in Textual Digital Analysis for the Humanities and Social Sciences, at Babes-Bolyai University. This first map will most probably facilitate the development of a larger project in DigiHUBB, in which we would map the periodicals indexed in the catalogs of the Romanian Academy, thus creating a comprehensive interactive map that displays the information in a more complex and more user-friendly way than a mere list of entries.”(Adela Fofiu, Ph.D. Lecturer at Babes-Bolyai University, social scientist)

“The very first contact to TEI was surprisingly pleasant and more daunting than I would have imagined, this has to do with the fact that it is such a large and active community, willing to help out should any questions arise. This, along with the already existing guidelines and templates, make sure that even complete beginners can try their hand and practice encoding. Also the oXygen software was of real help, as it really quickened the entire process of encoding, while being extremely exigent and telling the user precisely where a mistake was made, which I found very comforting. Despite the steep learning curve, the trainers were wonderful in clearing out any issues the participants had with the programs, so that generated a relaxed atmosphere throughout the workshop. I personally not only discovered a particular inclination for encoding text, but also through this workshop I envisioned new and exciting ways of showcasing my older, Art History-related research. Hopefully, my future DH projects will focus on long-standing idea of mine, that of creating a timeline based resource for Art History students, and this workshop gave me many ideas of how that could be implemented.” (Voica Puscasiu, doctoral candidate, visual art researcher)

“As a linguist, I can now start using oXygen and TEI for encoding texts to be worked with from a discourse analysis point of view. The data visualization tools presented during the workshop can be used for presenting the results in a visually appealing manner, so as to engage the public with the research I would have had conducted. Moreover, during the Clinics session at the end of the training, Roman Bleier introduced me and my colleague, also a linguist, to other tools that specifically address our current research needs in our field.”(Irina Drexler, Ph.D., linguist, post-doctoral researcher)

“The TEI and Data Visualization workshop in Cluj was extremely helpful for me in several ways. I attended, very eager to learn more about the TEI standard and about new trends in data visualization, a very enticing workshop with a dense hands-on basis for learning proper ways in which one can encode letter type documents. For the first 3 days, we used Oxygen in order to write our own html code and encode a set of local letters. Roman Bleier, our trainer was very helpful in that regard, anticipating any errors we might make while coding and by giving us a plethora of best practices. The second part of the workshop, held by Vinayak das Gupta and Shawn Day, was oriented towards creating personalized maps for several types of research regarding data visualization. We used Oxygen to create basic maps and pin down and position on the map boats and ships from all over the globe as an example of data visualization technique. The practical part went hand in hand with a large number of best practices and trends regarding the field, as well as extensive explanations about the state of the art techniques employed in order to render suitable visualization types of data. I am confident that this workshop got us several steps closer to achieving our goal, which was to make proper use of these techniques in order to apply them to our local cultural heritage, and has encouraged us to gain even more experience for building our own independent projects.” (Ruxandra Elena Bularca, Ph.D., digital culture researcher)

“I believe that this workshop has helped me understand how to design courses for (1) specialists with no technological background who would like to undertake projects for the digitisation, storing and

processing of various heritages; (2) computer scientists who would like to specialise in digitisation of various fields and text encoding.” (Liana Stanca, Ph.D., computer scientist, researcher).

4) Annex 4a: Programme of the meeting

<https://docs.google.com/file/d/0B-nWbt4Kfdn-dnV5TDBxdVVwVlk/edit>

TEI and Data Visualisation workshop in Cluj

27 April - 1 May 2015

Vinayak Das Gupta, Shawn Day, and Roman Bleier

Day 1 (27 April 2015)

- Introductory session
- Introduction: Plain text vs Markup
- oXygen: introduction and setup
- XML syntax and rules, XML languages
- Practical: Write your first XML with oXygen
- describe structure and content of your XML document, XML Schema
- Introduction to Document Modelling with XML
- Practical: Modelling documents with XML

Closing discussion

Day 2 (28 April 2015)

- Introduction to TEI
- Practical: Write your first TEI document
- Document modelling using the TEI standard
- Customising TEI for your project: Introduction to ODD and Roma
- Practical: Generate a XML schema with Roma

TEI and correspondence

- Modelling of letters and other correspondence using TEI
- Examples and special cases
- Practical: Encoding of letters using TEI

Day 3 (29 April 2015)

- A brief introduction to Data Visualization
- Roundtable: Advantages and disadvantages of dealing with structured data for analysis and visualization
 - Practical: Geo-tagging with Leaflet.js
 - ❖ The first session of this exercise will familiarize participants with map tiles and GeoJSON layers, and will illustrate how to build a simple web map using HTML, CSS and Javascript.
 - Practical: Geo-tagging with Leaflet.js (contd.)
 - ❖ The second session will explore the means to incorporate user-generated data within the map. It will continue and investigate Leaflet plugins, specifically ways of clustering markers and implementing annotations.

Day 4 (30 April 2015)

Exploring Data Visualisation of Various Types of Data

- Temporal Visualisation
 - Correspondence Visualisation
- Spatial and Temporal Storytelling
- Constructing Digital Narratives from Collections I
- Introduction and Setup
 - Examples
- Constructing Digital Narratives from Collections II
- Infrastructure

- Metadata

Day 5 (1 May 2015)

Constructing Digital Narratives

- Omeka Collections and Items

- Omeka Exhibits

Digital Narratives in Practice

- Omeka Best Practice

DH Face--to--Face Clinics

Closing

4) Annex 4b: Full list of speakers and participants

Facilitators

Vinayak Das Gupta, Doctoral researcher, Trinity College Dublin, Ireland

Roman Bleier, Doctoral researcher, Trinity College Dublin, Ireland

Shawn Day, Lecturer, University College, Cork, Ireland

Participants:

DigiHUBB

1. Corina Moldovan, Ph.D., Lecturer, Babes-Bolyai University, Director of DigiHUBB, Convenor
2. Adela Fofiu, Ph.D., Lecturer, Babes-Bolyai University, social sciences researcher
3. Cristina Felea, Ph.D., Associate Professor, Babes-Bolyai University, e-learning researcher
4. Liana Stanca, Ph.D., Associate Professor, Babes-Bolyai University, data mining and statistics
5. Christian Schuster, Ph.D., Lecturer, Babes-Bolyai University
6. Voica Puscasu, M.A., doctoral candidate, Babes-Bolyai University
7. Adina Puscasu, Ph.D., GIS researcher
8. Irina Drexler, Ph.D., linguistics researcher
9. Alexandra Cotoc, Ph.D., teaching assistant Babes-Bolyai University, Internet linguistics researcher
10. Leonard Bruckner, Ph.D., GIS researcher
11. Ruxandra Elena Bularca, Ph.D., Internet culture researcher
12. Gabriela Rus, Ph.D., postdoctoral researcher in history
13. Liviu Pop, Ph.D., researcher in digital archives

External participants

1. Nicolae Constantinescu, information architect, Bucharest
2. Andrada Catavei, Ph.D., art historian, Bucharest
3. Elena Cojuhari, archivist, National Library, Bucharest
4. Ergin Opengin, Post-doctoral researcher, Bamberg University, Germany
5. Marija Segan, MISANU (Mathematical Institute of the Serbian Academy of Sciences and Arts), Serbia