Studia Universitatis Babeș-Bolyai

# Digitalia

Vol. LXII (2017), June, No. 1


Centrul Digital Humanities Transilvania


# *Studia Universitatis Babeș-Bolyai*

# *Digitalia*

**Editorial Office**:

Room 218, 31 Horea St., 400202 Cluj-Napoca, Romania
Web site: http://digihubb.centre.ubbcluj.ro/journal/index.php/digitalia
http://studia.ubbcluj.ro/serii/digitalia/
Contact: digihubb@ubbcluj.ro

**1/2017**

**June**

# *Studia UBB Digitalia Editorial Board*

*Studia UBB Digitalia*, Volume 62 (LXII) 2017, June, Issue 1, 3-4

Published Online: 2017-06-30

Issue DOI:10.24193/subbdigitalia.2017.1

**Thematic issue**

# Digitising the Humanities

*Issue Editors: Corina Moldovan and Christian Schuster*

## CONTENT / SOMMAIRE

Willard McCARTY, What are *Studia Digitalia* ..... 5

Corina MOLDOVAN, Christian SCHUSTER, *Coordinators Foreword* ..... 7

Roman BLEIER, Digital documentary editing of St Patrick's epistles. Linking the manuscript witnesses to the canonical text ..... 9

Claire CLIVAZ, Lost in translation? The odyssey of 'digital humanities' in French ..... 26

Nicolaie CONSTANTINESCU, Stretching the boundaries of publishing: the Open Web Platform and the alternatives ..... 42

Vinayak das GUPTA, Albums in the attic. An Investigation of photographic metadata ..... 57

Gábor PALKÓ, Media archaeology of Institutional archives? ..... 75

Liviu POP, A low te(a)ch approach to digital humanities ..... 83

Voica PUȘCAȘIU, The importance of data visualization: teaching art history through timelines ..... 89

Christian SCHWADERER, eXist DB or Saxon/C in PHP. A comparison between two approaches for XSLT 2.0 based websites ..... 100

Liana STANCA, Cristina FELEA, Ramona LACUREZEANU, Christian SCHUSTER,
The digitisation - measuring physical and virtual presence ..... 112

## Review Section

Voica PUȘCAȘIU, Bringing Bosch alive through multimedia ..... 133

Valentin TRIFESCU, Viorella Manolache, *Signs and designs of the virtual(izing) E@St*, introduction Ian Browne, instead of an afterword Cristi Pantelimon, LAP Lambert Academic Publishing, Saarbrücken, 2013, 124 p. . ..... 138

*Studia UBB Digitalia*, Volume 62 (LXII) 2017, June, Issue 1, 5-6
Published Online: 2017-06-30

# What are *Studia Digitalia*?

More than half a century ago an American professor of English literature, Joseph Raben, founded *Computers and the Humanities*, the first journal in the field. In the first article of its first issue Louis Milic wrote with uncommon clarity of vision that,

> The true nature of the machine is unknown to us, but it is neither a human brain nor a mechanical clerk. The computer has a logic of its own, one which the scholar must master if he is to benefit from his relations with it. Its intelligence and ours must be made complementary, not antagonistic or subservient to each other…. The computer can be made an extension of man only if it opens avenues we have not suspected the existence of. (1966: 4)

Mimesis of human behaviour is not the goal, at least not for scholarship, he suggested. Testing the machine against a human benchmark (whatever that might be) to determine whether it has reached a state of 'artificial intelligence' isn't either, for by coming into contact with digital glimmerings of intelligence we realise that it is plural and that its plurality makes it a question. "Thinking in a new way", as such questioning demands, "is not an easy accomplishment", Milic wrote. "It means reorientation of all the coordinates of our existence." (1966: 5) Milic entitled his article "The Next Step". It is a step we have only taken, if at all, falteringly.

Meanwhile, thanks to astonishing technological prowess and extravagant engineering, Alan Turing's mathematical abstraction has materialised and diversified into the multi-layered, user-friendly appliances with which we are surrounded, adorned, even interpenetrated. These have become nearly ubiquitous and mostly invisible. One manifestation of our *Studia Digitalia*, seemingly the dominant one, promotes or goes along with the notion that we are in a 'post-digital' age, i.e. that as Brian Cantwell Smith argued in the 1990s, the genius of the digital is that it renders digitality irrelevant. The subject of study which follows is, then, the effects, the *impact* on us of these ever-so-user-friendly devices. That's one answer to the question I raise in my title.

But there are other "coordinates of our existence" that, I would argue, are more important and in some danger of being completely forgotten. They are more important because devotion solely to impact cuts us off from active participation in shaping the great engine of our century (and of the latter half of the previous one). What is the 'intelligence' that legions of very bright and clever computer scientists are implementing? Who is to say? Who is to bring millennia of deep thought about 'intelligence' to bear?

Let me return to Millic on this matter via the philosophy of experimental science. In his brilliant study, *Experiment and the Making of Meaning*, David Gooding wrote that,

> The 'act' of observing something for the first time presupposes the activity that rendered it visible. Once you have learned how to see something, the activity becomes unimportant or second nature - it passes into what Polanyi called peripheral awareness. It is easy to suppose that the cases we are aware of - where we seem to 'see' first and then record what we then say we have observed - are typical of all seeing. Most of our seeing is like this because for most of what we experience, the perceptual stage is already set.... we have a repertoire of images and concepts which we bring to the task of selecting or depicting just those aspects of experience we wish to record or communicate. This repertoire is usually adequate to the task and its use is therefore usually tacit. The repertoire itself remains imperceptible until we encounter (or are asked to share) an experience which does not readily fit our available renderings. Novelty brings the repertoire into focal awareness. (1990: 74)

The question I would raise, then, is about what is novel, what is *different*, what is *strange* about how the machine artificially reasons. How does its binary, combinatorial way of doing that diverge from ours? How does it illumine our inherited "repertoire of images and concepts" with which we are only peripherally aware?

But, Milic might insist, it's not a mirror of the ordinary kind that we're dealing with. No invention is. The history of technology, the digital machine in particular, tells us that inventor and invention are co-involved in a developmental cycle of change. Some call it co-evolutionary. The imperative, I would think, is to be a conscious, active participant, and so have a chance of helping to make the developmental cycle virtuous rather than vicious, enabling rather than entrapping.

My brief argument is that we need, in much greater proportion than now, comparative studies of thought and action in human and machine, especially those that bring out the differences (which, being differences, imply the likenesses with which they contrast). In "An Essay for S. I. Witkiewicz" (1935), the Polish artist and writer Bruno Schulz wrote that "The role of art is to be a probe sunk into the nameless." Let the digital machine be considered a work of human art with exactly that role.

What are *Studia Digitalia*, then? To answer in a quite literal-minded, prepositionally centric way, they are, or could be, studies *of* the digital, in the sense that we say a person is 'of fine stock', and not merely *about* something 'out there', observed from a safe distance. Let them open up those "avenues we have not suspected the existence of".

**Willard McCarty**

King's College London
25 July 2017

*Studia UBB Digitalia*, Volume 62 (LXII) 2017, June, Issue 1, 7-8
Published Online: 2017-06-30

# Coordinators Foreword

As editors of Studia Universitatis Babeș-Bolyai Digitalia, we are proud to announce the publication of the first issue of *Studia Universitatis Babeș-Bolyai Digitalia*, the first Romanian academic journal dedicated entirely to the Digital Humanities research and practice. The main aims of *Studia UBB Digitalia* are motivating and promoting the research in the interdisciplinary field of Digital Humanities, especially in Europe, in general, and Romania, in particular.

Publishing a new journal on Digital Humanities is indeed an audacity, given that the field is still seen, after decades of evolution, as an odd hybrid between seemingly incompatible realms of scientific endeavour. Bridging this apparent dichotomy has been the focus of many researchers in both the humanities and the computer sciences for the past 50 years, but still the boundaries between the realms exist and grow strong, especially in the institutional practice of universities. Trying to break this barrier proves to be even more harder in Romania, where the classification of "traditional" fields of scientific research dictate the distribution of research funds and the evaluation process of research output. Chances are, if an historian publishes an article on his research using computational methods in a journal for computer sciences, that his or her results cannot be "capitalised" in his own field; likewise, computer scientists may see their work in Digital Humanities not accounted for, if published elsewhere than in well-known CS journals. Studia UBB Digitalia aims to give researchers from all fields the chance to get their results published, recognized and valued, at the same time proving that "interdisciplinarity" is not just a fancy term, but already the reality of cutting-edge research in all the fields.

In line with this orientation, the first issue of Studia UBB Digitalia brings together papers from different areas: theoretic papers about the meaning of Digital Humanities (Claire Clivaz), papers focused on practice with digital text analysis (Roman Bleier), photographic metadata (Vinayak Das Gupta) and media archaeology (Gábor Palkó), papers describing innovative teaching techniques (Liviu Pop, Voica Pușcașiu), as well as more technical papers ranging from different approaches to coding (Christian Schwaderer) or to digital publishing (Nicolaie Constantinescu), to testing the reliability of digital analysis instruments (Liana Stanca et al.).

We should firstly like to thank the entire team of DigiHUBB – the Transylvania Digital Humanities Centre at the Babeș-Bolyai University Cluj-Napoca for both the last three years since the creation of DigiHUBB, and for the past year of preparation for this first issue of *Studia UBB Digitalia*. The publication of this first issue is the result of the work of many hands, seen and unseen, and the outcome of a great support from both institutional partners and many enthusiastic people from all around the (mainly European) Digital Humanities community. None of our work would have been possible without the help of the members of our advisory board, of the reviewers, of the authors who joined us into this audacious venture and contributed to this first issue, and without the editorial team of Studia Universitatis Babeș-Bolyai. Thank you all for your commitment, for the opportunity to work together, and for believing in us.

Corina Moldovan and Christian Schuster

Editors

*Studia UBB* Digitalia, Volume 62 (LXII) 2017, June, Issue 1, 9-25

Published Online: 2017-06-30

DOI:10.24193/subbdigitalia.2017.1.01

# Digital documentary editing of St Patrick's epistles. Linking the manuscript witnesses to the canonical text

Roman Bleier

Karl-Franzens-University of Graz, Centre for Information Modeling - Austrian Centre for Digital Humanities, Austria,
E-mail: roman.bleier@uni-graz.at

**Abstract:** St Patrick of Ireland, the fifth-century missionary and bishop, wrote two epistles which are commonly referred to as *Confessio* and *Epistola ad milites Corotici*. These two texts survive in seven medieval manuscript witnesses which were copied centuries after Patrick's time. This article discusses digital transcriptions of these manuscript witnesses. The transcriptions were encoded using the encoding standard of the Text Encoding Initiative (TEI) following a documentary editing approach. The transcriptions will be the core of a new digital documentary edition. This edition is meant to be a research tool for the exploration of the medieval documents, but at the same time it is a mediator between the manuscript witnesses and existing text-critical editions and translations of the texts. This mediator function is achieved by including markers of lines and chapter of a canonical edition and links to an existing text-critical edition online. In the final section a few examples will briefly be look at that illustrate what kind of analysis will be possible and what visual presentations of the transcriptions will be incorporated in the new edition.

**Keywords:** Text encoding; TEI; Canonical; Manuscripts; Text; Edition; Document; Patrick of Ireland.

## Introduction

This article will discuss the development of digital transcriptions of the manuscript witnesses of St Patrick's epistles which are the core element of a Digital Documentary Edition currently developed at the University of Graz. The idea to create such transcriptions

was conceived during work on the Saint Patrick's *Confessio* Hypertext Stack website, a digital scholarly editing project of the Royal Irish Academy that makes numerous resources related to the historical St Patrick available to the general public including images of the manuscript witnesses, text critical editions and translations. Transcriptions of the manuscript witnesses are not part of the Saint Patrick's *Confessio* Hypertext Stack as the primary focus of this project was on Ludwig Bieler's text-critical edition of Patrick's epistles and, therefore, the new edition will augment the already existing online resources. However, to think of the new edition simply as a add-on to the HyperStack is not correct. As will be argued later, the new edition will be a tool that is based on documentary model and addresses different research questions related to medieval text structure. Besides this documentary focus, the edition also understands itself as a mediator between the manuscripts and text-critical editions to provide a more holistic understanding of the manuscript evidence in the context of text-critical research. The resources of the Saint Patrick's *Confessio* Hypertext Stack are central and referenced and linked to in various ways in the new edition.

The manuscript witnesses were transcribed and enriched with metadata during a PhD project at Trinity College Dublin for which the Digital Arts and Humanities (DAH) programme provided funding and support. The Marie-Curie Initial Training Network DiXiT and the University of Graz provided further funding for the revision of the encoding model and the development of the edition. The electronic transcriptions were developed following the latest recommendations of the Text Encoding Initiative (TEI) and with a documentary editing approach in mind. The purpose of this article is it to discuss the rationale and the methodology behind the transcriptions project and to point at some possible use cases in the forthcoming digital edition. A particular challenge is the integration of both the diplomatic manuscript view and the canonical text structure used in print editions. It will be discussed later how this is achieved in the encoding model and in the web presentation.

## Exploring the manuscript witnesses

St Patrick of Ireland is a popular saint who is especially well-known for the celebrations around the day of his presumed death on March 17. Patrick was a historical personage and a writer of letters, two which survive today. These are the two epistles known as *Confessio* and the *Epistola ad milites Corotici* (Bieler), and are variously referred to in what follows as Patrick's writings, Patrick's letters, or Patrick's epistles.

The *Epistola ad milites Corotici*, or short *Epistola*, is a letter pronouncing excommunication against a certain Coroticus and his men for raiding one of St Patrick's Christian communities, while the *Confessio* is a text with a strong autobiographical tendency in which Patrick tries to justify his life's work and his mission against accusations

from other church men, in that sense, an *apologia pro vita sua*. Patrick's epistles are very important historical texts. They are the only reliable sources for the life of the historical Patrick and for his Christian mission to Ireland, and they are two of the few surviving contemporary texts that inform our knowledge of fifth-century Irish and British history (O'Loughlin). Furthermore, his epistles also have much to reveal about mobility, of people and ideas, shedding light in particular on the importation of ideas from the Roman empire into Ireland.

One problem presented by Patrick's letters as a source is that his original letters have not survived and the medieval manuscript witnesses contain different versions of his texts. The earliest surviving copy of the *Confessio* dates to the first decade of the ninth century (ca. 807), and is found in the Book of Armagh, now Ms 52 in the library of Trinity College Dublin. Not only is this the oldest manuscript witness, but it is the only surviving Irish manuscript, containing Patrick's writings (albeit only one of the two, namely an abbreviated version of Patrick's *Confessio*). Besides this Irish manuscript witness, there are three surviving manuscript witnesses from northern France, and three from England. These all date from the tenth to the thirteenth century and contain a much longer version of the *Confessio* and also Patrick's other text, the *Epistola ad milites Corotici* (Bieler).

The editors White and Bieler proposed that the manuscript witnesses form two groups based on their linguistic and codicological characteristics. The first (or Irish) group consists only of one manuscript, the Confessio-text in the Book of Armagh. The six other manuscript witnesses form a second group of manuscripts of continental and English origin. The most significant differences between the two groups of manuscripts can be summarized in five points (Bleier *Re-Examining the Function of St Patrick's Writings in the Early Medieval Tradition*):

1. Irish vs. non-Irish manuscripts.
2. Codex composition: the Irish copy of the *Confessio* survived in a collection of texts related to St Patrick.
3. Age: the Irish manuscript is the earliest to survive.
4. Shorter version of *Confessio* in the Irish group.
5. No Epistola in the Irish group.

From a text-critical and content point of view the last two points in the above list are in particular relevant. The Irish manuscript witness in the Book of Armagh contains one of the most reliable readings, however, substantial parts of the text of the *Confessio* and the entire *Epistola* are missing in this manuscript. This is a problem as this witness has great value for the reconstruction of the text. In the nineteenth and early twentieth century some scholars argued even that the shorter Book of Armagh version was the full *Confessio* text of St Patrick and the longer versions found in the other

manuscripts are the product of the merging of two or three different letters of Patrick (Bleier *Re-Examining the Function of St Patrick's Writings in the Early Medieval Tradition*). While this argument has been widely refuted today, the Book of Armagh *Confessio* is central to a text-critical reconstruction of the text. Furthermore, it is an important document that gives insight into how the epistle was perceived and used in early medieval Ireland (Binchy 41-43) and is key to the understanding of the composition of the Patrician section in the Book of Armagh (Sharpe).

The diversity of the manuscript witnesses and the amount of text that is actually missing in some of them is one reason why a documentary edition is important. Some editors in the past have attempted to make the textual differences between the long and the short *Confessio* version visible by highlighted the parts of the text not available in the Irish manuscript (Dumville). The digital edition developed by the author goes even further and provides electronic transcriptions that can be compared in an automated way. Not only is it possible to see where text is and is not available, but also how the text was structured and presented in the manuscripts.

**Medieval and canonical text structure**

St Patrick's epistles have been published in print since the seventeenth century in editions, transcriptions and translations of varying quality. The variant readings of all seven medieval manuscript witnesses were first published by White in 1905 (*Libri Sancti Patricii: The Latin Writings of Saint Patrick*)[1]. White's edition has been certainly one of the most important and influential editions in Patrician studies. It has been credited to be the first truly text-critical edition (Bieler 30) as White used existing print editions with caution and applied a solid editing methodology. White also created a new chapter structure for *Confessio* and *Epistola*. White's page layout and structure, 62 short chapters for the *Confessio* and 21 for the *Epistola*, was later adopted by other editors and translators and is still used today. The mid-twentieth century edition by Bieler uses even White's page and line numbers as anchor points for apparatus and commentary. Since White's time other important print editions were presented by Bieler, Hanson, Hood and others. All these editions follow a text-critical approach and all use White's chapter organisation. Therefore, White's chapter organisation will also be called canonical chapter structure later in this article.

While the focus of interest was traditionally on the content and the reconstruction of the text of St Patrick's original epistles, in recent years more attention has been on text structure. A ground-breaking study exploring the possible text structure of the

[1] White's first edition from 1905 was based on six manuscript witnesses (the Paris manuscript was missing). However, in the same year he published an article discussing the readings of the Paris manuscript and correcting his earlier edition (*The Paris Manuscript of St. Patrick's Latin Writings*).

original epistles was presented by Howlett in 1994 in form of an edition *per cola et commata.* The edition revealed that Patrick used rhetorical devices and text structure similar to the one found in the bible. While Howlett's edition focuses on St Patrick's fifth-century texts, it triggered an ongoing discussion about text structure. The manuscript witnesses have only to a limited extent been part of this discussion. One reason being that a systematic study of layout and text structural features of the manuscripts is still missing.

Two editions that have been produced in the last decade try to make the manuscript witnesses better accessible. A key feature to make the manuscript witnesses better accessible is to somehow link transcriptions or facsimile images to the canonical chapter organisation used in most print editions. Freeman tried this by presenting his transcriptions following the canonical chapter structure rather than the layout of the manuscript witnesses. This is an interesting approach, but his transcriptions have some limitations what makes the comparison of different manuscript versions and text-critical editions cumbersome. For instance, Freeman does neither provide marker for line breaks of the manuscripts nor does he indicate the canonical line breaks and this makes granular comparison of the transcriptions very difficult. Another limitation is that Freeman focused primarily on the text of the manuscript witnesses and did not sufficiently record some features of the manuscripts such as the layout and abbreviations which are be central to a documentary edition.

Another approach to link the canonical text of an edition with the manuscript witnesses was realised in context of the already mentioned Saint Patrick's *Confessio* Hypertext Stack. The central component of the Saint Patrick's *Confessio* Hypertext Stack is a digital edition of Bieler's text-critical print edition. An additional feature is that the individual chapters of St Patrick's *Confessio* and Bieler's *apparatus criticus* are linked to facsimile images of the manuscripts. However, the possibility for comparison across different versions is again limited as the linking between the manuscript images and the text-critical edition happens on a folio and chapter level. Freeman's edition and the Saint Patrick's *Confessio* Hypertext Stack go some way towards providing better access to the manuscript witnesses. The new edition tries to overcome some of the limitations these editions have, providing a tool for the study of the manuscript witnesses on the one hand and on the other hand a better interface between canonical text structure and the manuscripts.

## A documentary encoding model

The central component of the new edition are transcriptions of the seven medieval manuscript witnesses of *Confessio* and *Epistola ad milites Corotici* transcribed from high-resolution manuscript images and enriched with meta-information using XML

following the recommendations of the TEI. The abbreviation TEI stands both for an encoding standard for electronic texts, Text Encoding for Interchange, or for the consortium that releases and continuously develops this standard, the Text Encoding Initiative. The TEI consortium was established in 1987 as an international research project to develop a standard to 'facilitate the creation, exchange, and integration of textual data in machine-readable form'. The goal was to create a standard that would support the encoding of 'all kinds of texts, in every human language, from every historical or social context'. Since its first release the TEI has become the de-facto standard for the encoding of humanities data.[2]

A documentary (or document-centred) editing approach focuses primarily on the accurate reproduction of historical documents such as the manuscript witnesses of St Patrick's epistles. This is distinct from a text-critical approach which focuses on the recreation of the original work of an author (e.g. the text of Patrick's fifth-century original epistles). Documentary editing is nothing new and has already been done in a print environment in the form of diplomatic, ultra-diplomatic and genetic editions long before the invention of electronic computers, John Gwynn's diplomatic transcription of the Book of Armagh is a good example. However, the majority of print editions in the nineteenth and twentieth centuries did not use such an editing approach. The reasons were that editors were primarily interested in the content of texts and less in the manuscript contexts texts survived in. Furthermore, advances in photography and facsimile editions questioned the value of ultra-diplomatic transcriptions. After all, high-quality photographs of a manuscript are a more accurate reproduction than a transcription can be. With advances in digital scholarly editing the view on documentary editing changed. Editors started to provide diplomatic transcriptions of manuscripts in order to interlink them with images and electronic editions put facsimile image and a reading text side-by-side. Pierazzo has pointed out that we should make a clear distinction between traditional diplomatic transcription and the semantically enriched digital transcriptions and she suggested the term Digital Documentary Edition (*A rationale of digital documentary editions*). This new generation of diplomatic transcriptions/editions goes far beyond its print predecessors and depending on one's research question a semantically enriched transcription can be a 'very powerful and versatile tool' (*Digital Documentary Editions and the Others*).

The new edition of St Patrick's epistles follows a documentary approach in the encoding and presentation of the transcriptions. The goal is a Digital Documentary Edition that is on the one hand a tool for the closer study of medieval text structure and scribe related features of the manuscript witnesses. Consequently, the TEI encoding model focuses first of all on the physical and topographic structure of the document:

[2] See: http://www.tei-c.org/index.xml.

manuscript witness, folio, page, column and line. The encoding of these features was not always easy in TEI. Until recently the representation of the logical structure of a text, such as book, chapter, paragraph, etc., or the representation of a critical-apparatus was much better supported than material properties of a manuscript and topographic layout of the text. For instance, for a long time the TEI had no specialised element for the tagging of topographic lines while elements existed for the representation of sentences and verse lines. Some scholars successfully argued for the extension of the TEI to provide more flexibility for the development of documentary and genetic editions (Pierazzo and Stokes). The most substantial change in that direction happened in 2011 with the release of TEI P5 2.0.[3] The transcriptions of St Patrick's manuscript witnesses take advantage of some of these new TEI elements such as <sourceDoc>, <surfaceGrp>, <surface>, <zone> and <line> to represent a documentary model of the manuscripts. The <sourceDoc> contains the entire transcription of a manuscript witness. The <surfaceGrp> element groups folios together and the text of individual pages is transcribed within a <surface> element. Further topographic regions such as columns were encoded within a <zone> element with an attribute @type specifying the feature that was captured. Finally, within the different major zones the <line> element is used to tag topographic lines.

**Encoding structural elements of the manuscript witnesses**

Besides the above discussed topographic features a number of features were encoded that might have been used by the medieval scribes to implement some sort of a text structure. For instance, the incipit and explicit, where present, and the decorated initial are indicators for the beginning and end of a text such as the *Confessio* or the *Epistola*. It is important that these features are included in the transcription as they give us an important clue about how the texts were perceived by the copying scribe (Bleier *Re-Examining the Function of St Patrick's Writings in the Early Medieval Tradition*). Additionally a number of features were encoded that function as indicators of internal text structure implemented by the medieval scribe. Special attention was given to the recording of highlighted characters, punctuation and differing spaces between the words.

Within the text highlighted initials (often in combination with punctuation and/or a big space) are in many cases a sign for the start of some sort of a textual division identified by the medieval scribe. The Irish manuscript witness in the Book of Armagh is an interesting case as the scribe uses a very dense script with numerous ligatures, sub- and superscript characters and abbreviations. Furthermore, sections within the text are initialised not simply be a highlighted character, but by a feature

[3] See: http://www.tei-c.org/release/doc/tei-p5-doc/readme-2.0.html.

called a ‘diminuendo’. This is a common feature of early Insular manuscripts. The characteristic of a diminuendo is that the first few characters of a word, often the first word in a sentence, are bigger than the rest of the text. The biggest character is the first character in the word and every succeeding character is written in decreasing size. The *Confessio* text in the Book of Armagh has such a diminuendo at several points in the manuscript. Highlighted characters and the diminuendo are encoded using the TEI <hi> to indicate that these characters are highlighted in some way. TEI allows to specify the form of highlighting by using @rend which may contain a description of how the text is rendered.

For the tagging of punctuation the TEI has a specialised element, the <pc> element. This element is also used in the transcriptions for the recording of various forms of text-structural punctuation. On the <pc> element an @type is provided to assign a standardised name for the punctuation in question (e.g. <pc type="period">.</pc>). In the case of a punctus or period the encoding is quite simple, because the Unicode character ‘full stop’ is used. This character is a standard Unicode character and part of the Basic Latin Unicode code range. However, not all punctuation characters used in medieval manuscripts have a standardised Unicode mapping – *punctus elevatus* or *punctus interrogativus* are such examples. For these characters it would be possible to use Unicode mappings supplied by work groups such as the Medieval Unicode Font Initiative (MUFI) for representation. In these cases the <g> element with an @ref is used within the element <pc> to link to a description of the character in the TEI header section where the element <charDecl> declared this non-standard characters and provides a set of specialised childelements that allow for the description of the character in question. Within <charDecl> an optional Unicode mapping such as the MUFI character declaration for the above mentioned *punctus elevatus* can be assigned to a character. In the assignment of xml:ids for non-standard characters the MUFI recommendations were followed.

The last text-structural feature that was encoded are spaces between the words.

The study of word separation and spaces between words is an important matter in early medieval manuscripts as spaces are sometimes used as indicators of text structure but at the same time they are used inconsistently and it is not always the case the what we would consider to be a word today is indeed delimited by spaces (Saenger 32-36). In the manuscript witnesses of St Patrick’s writings spaces between the words fall roughly into one of three categories: no space, normal space, large space. All manuscript witnesses show a high degree of word separation. The later manuscripts even have features which are signs of high awareness of word separated manuscripts, such as the *traits d’union* or the ‘accented double ii’. However, all manuscript witnesses contain occasional irregular spacing. Interestingly, the ninth-century manuscript from Dublin, the oldest manuscript, and the twelfth-century manuscript

from Arras, one of the youngest manuscripts, show the highest level of word separation. In contrast, the eleventh-century manuscript from London has longer passages in 'aerated' script with irregular spacing.

In the digital transcriptions the TEI <space> element is used to record spaces between words and the attribute @quantity is used to record the quantity of space. However, because the manuscript witnesses come from different cultural backgrounds and time contexts the question of how to measure space is quite problematic. The size of spaces, whether between words or not, depends greatly on the format of the manuscript and the script used by the scribe. For instance, the Dublin manuscript (=Book of Armagh) is in size, script and layout very different from the Paris manuscript. The Dublin text is in a codex the size of a pocket gospel, written in narrow, very compact Insular script on pages with two-column layout, while on the other hand the Paris text is written in Carolingian minuscule on one-column pages in a script with more generous spacing and character width. Consequently, the measurement of spaces has to be relative to the context in which they are found rather than being absolute. Therefore, it was decided to use the average height of minims – the vertical strokes of the characters m, n, u and i – as a unit of measurement. To apply this measurement consistently a grid overlay was used over the manuscript images. However, even with a grid overlay the measurement of spaces was not always straightforward. Sometimes it is difficult to decide if there is a space and how big it really is.

Besides text structural features also a number of 'scribal accidentals' such as additions, deletions, substitutions and abbreviations were tagged using suitable TEI/XML elements. The advantage of TEI encoded transcriptions over print editions can be demonstrated very well with the encoding of abbreviations. The abbreviated form of a word is rarely incorporated into scholarly print editions. Editors usually provide a 'reading text' rather than a replication of the abbreviation. Sometimes the abbreviated word is highlighted in italics. This certainly makes sense, because correctly resolving of abbreviations requires a high level of linguistic and palaeographical knowledge about a manuscript and the text it contains. Consequently, simply reproducing abbreviations would make an edition as difficult to use as the original manuscript. However, abbreviations are an important feature of a scribe's copying style and therefore recording it may be useful to answer specific research questions. For instance, questions related to a scribe copying habits. Electronic text encoding with TEI allows both the abbreviated and the expanded form of a word to be equally recorded. This is done using the <choice> element. For example within the <choice> element both forms the abbreviation ur̄i and the expansion *uestri* can be recorded using the <abbr> element (for the abbreviated form) and the <expan> element (for the expanded form). Additionally, may the abbreviation marker and the part that was expanded by the editor be tagged.

Such a TEI encoding allows for the presentation of different forms of an abbreviated word and additionally are the abbreviations searchable and can be statistically analysed across the manuscripts.

**Including markers of the canonical numbering**

Texts can be structured in different ways. The two text structures important for this project are the topographic text structure which follows the text-bearing material object and the semantic text structure which provides structure for the content of a text structuring using elements such as headings, sections and chapter. In the previous section the encoding of the topographic text structure was discussed. Furthermore, the encoding of text structural marker in the medieval manuscripts. These marker were included by the medieval scribes and represent the semantic structure of the text as understood by the medieval scribes. However, this is a different semantic structure than is found in recent editions and translations of St Patrick's epistles. In editions and translations the chapter structure introduced by White is used. Due to their wide adoption and the historical importance of White's edition, the chapter number introduced by him can be called canonical numbers and White's chapter structure the canonical text structure of the epistles.

The Functional Requirements for Bibliographic Records (FRBR) model can be used to describe the different levels on which these text structures operate (Kalvesmaki). FRBR is a model developed in a library cataloguing context and the main levels are work, expression, manifestation, and item.

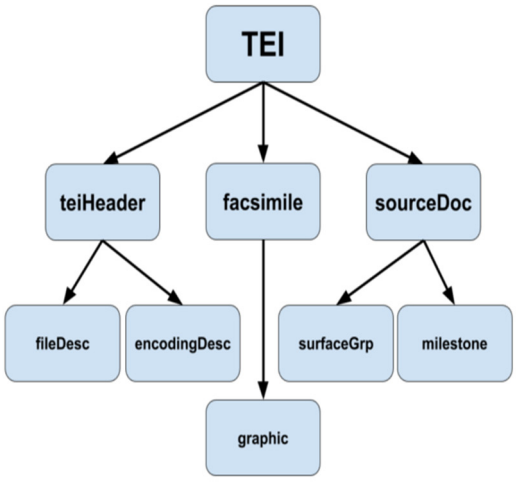
> "The entities in the first group represent the different aspects of user interests in the products of intellectual or artistic endeavour. The entities defined as work (a distinct intellectual or artistic creation) and expression (the intellectual or artistic realization of a work) reflect intellectual or artistic content. The entities defined as manifestation (the physical embodiment of an expression of a work) and item (a single exemplar of a manifestation), on the other hand, reflect physical form." (International Federation of Library Associations and Institutions 13)

While St Patrick's original works are unfortunately lost, there are several expressions of his epistles (editions and translations) that use a canonical numbering system[4] which becomes a physical reality in the manifestation and items of a particular edition or translation. In the FRBR context the manuscript witnesses are on the one hand expressions, manifestations and items of Patrick's work. Therefore, the

[4] Before the twentieth century two other systems were used, and Howlett introduced another text structure in the 1990s. However, none of these alternative text organisations are as widely used as White's chapter structure and numbering.

canonical numbering for Patrick's work could be used. However, at the same time the manuscripts may be considered works in their own right and therefore it would also make sense to use another numbering system frequently used in a manuscript studies context following the physical topographic text structure and use folios and lines to locate sections within the text (Kalvesmaki). These are two different ways to approach the text of a manuscript and depending on a researcher's interest - either in the manuscript or the text it contains - either text structure can be a useful. In order to make the new documentary edition interoperable with the manuscript witnesses and text-critical editions, it was necessary to include information not only about the topographic structure, but also the canonical in the TEI transcriptions.

One downside of XML is that it enforces a hierarchical, tree-like document structure. Each XML document should have exactly one root element, and nested in it (or, to continue the tree analogy, 'branching off from it') are all other elements and text content (Fig. 1). However, such a hierarchical, tree-like document structure does not allow for the equal representation of the topographic and canonical text structure. The first requires the encoding of folios, columns and lines of the manuscripts while the second requires the encoding of the canonical chapters of White. This would certainly cause a nesting conflict as it is very likely that a canonical chapter starts on one folio, but ends on another and therefore the chapter could not be nested fully within an element representing a folio. Using overlapping markup is a violation of one of the central rules of XML: every XML element (besides the root element) has to be fully nested within another element and therefore it needs to be escaped somehow.




```
<TEI>
   <teiHeader>
      <fileDesc> ... </fileDesc>
      <encodingDesc> ... </encodingDesc>
   </teiHeader>
   <facsimile>
      <graphic />
   </facsimile>
   <sourceDoc>
      <milestone />
      <surfaceGrp> ... </surfaceGrp>
   </sourceDoc>
</TEI>
```

*Fig. 1: The three top levels of the TEI model used for the transcriptions. Once represented as a treegraph (left) and once with the proper nesting highlighted by indentation (right).*

The issue of overlapping hierarchies is a common problem in TEI projects. TEI has its own Special Interest Group that works on solutions to overlapping hierarchy issues.[5] Furthermore, the TEI guidelines contain a section named 'Non-hierarchical Structures' which advises possible workarounds to encode overlapping hierarchies in an XML-conform manner (*20 Non-hierarchical Structures*). Essentially the TEI guidelines suggest four ways to deal with data that leads to overlapping hierarchies in XML:

1. Multiple encoding of the same data, e.g., the different hierarchies should be modelled and transcribed separately;
2. The choice of one hierarchy as the main hierarchy, with others encoded using empty elements;
3. The fragmentation or segmentation of one or more hierarchies. In this case too one hierarchy is usually favoured as the main one;
4. Stand-off markup.

To encode the canonical chapters in the digital transcriptions of St Patrick's writings, the second method, empty elements, was chosen. The beginning and end of the canonical chapters was marked with the <milestone> element, an empty element that can be used to indicate a 'boundary point'. In an empty element no further child elements or text is allowed. To provide further information about what kind of structural feature is represented the attributes @type, @unit and @n are used on the <milestone> element and the attributes @xml:id and @spanTo are used to indicate how far a chapter goes. This kind of markup allows for the identification of what transcribed text belongs to what canonical chapter without violating the documentary encoding model.

In addition to provide more granular anchor points that can be used to compare the transcriptions on a line level with White's and Bieler's editions, the line and page numbers of White's edition were encoded with empty elements for line break <lb/> and page break <pb/>. It is hoped that the inclusion of these features enables a maximum level of interoperability between the transcriptions, text-critical editions and translations. Furthermore, as each manuscript witness has a different text layout and structure, the canonical chapters and White's lines can be used as the basis for cross-version comparison and reference.

## Different presentations of the transcriptions

A main benefit of TEI/XML encoded texts is that they can be used as source for a number of outputs. The new edition of St Patrick's epistles takes advantage of this by showing the transcriptions in different presentational views. As the TEI model of this project follows a documentary editing model the primary view presents the

[5] See: <https://wiki.tei-c.org/index.php/SIG:Overlap>.

transcriptions following the layout of the manuscripts, a diplomatic view. It could be argued that such a diplomatic presentation is not useful as it simply tries to mimic the original manuscript. Facsimile images of the manuscripts are freely available on the Saint Patrick's *Confessio* Hypertext Stack website and interested people should simply look up the manuscript images. However, the digital diplomatic transcription have the benefit that they come with a lot of data that was encoded in the TEI. They are a research tool to explore the text structural features from the medieval manuscripts by utilizing the markup used for the encoding of capitals, highlighted letters, punctuation and spaces. For instance, this can be done in form of visual highlighting and overlay over the diplomatic transcriptions to show where in the text highlighted initials are preceded by punctuation (Fig. 2).

Another benefit of the digital documentary transcriptions is that they facilitate the study of 'scribal accidentals' that were encoded in the TEI such as additions, deletions and abbreviations. The TEI files can be used as the basis for visualization and automated analysis. In that way existing patterns can be spotted more easily which provides unique insight into the copying process of the individual scribes. For instance, a simple query shows how many deletions were made in a manuscript or what kind of abbreviations a scribe preferred.

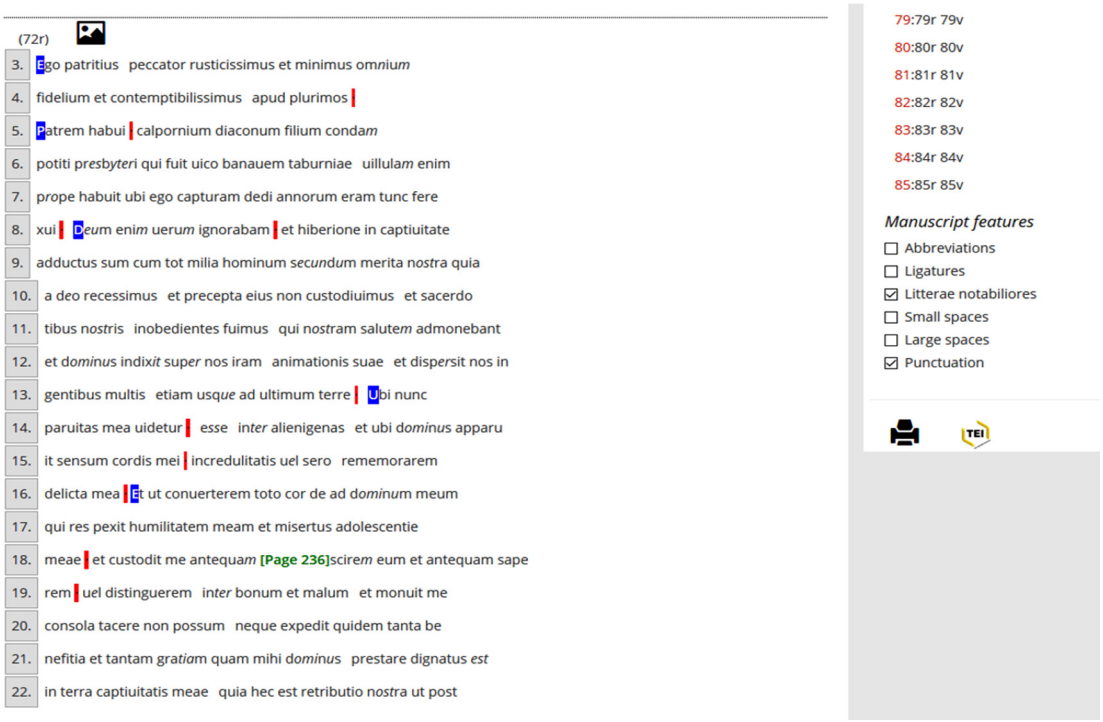
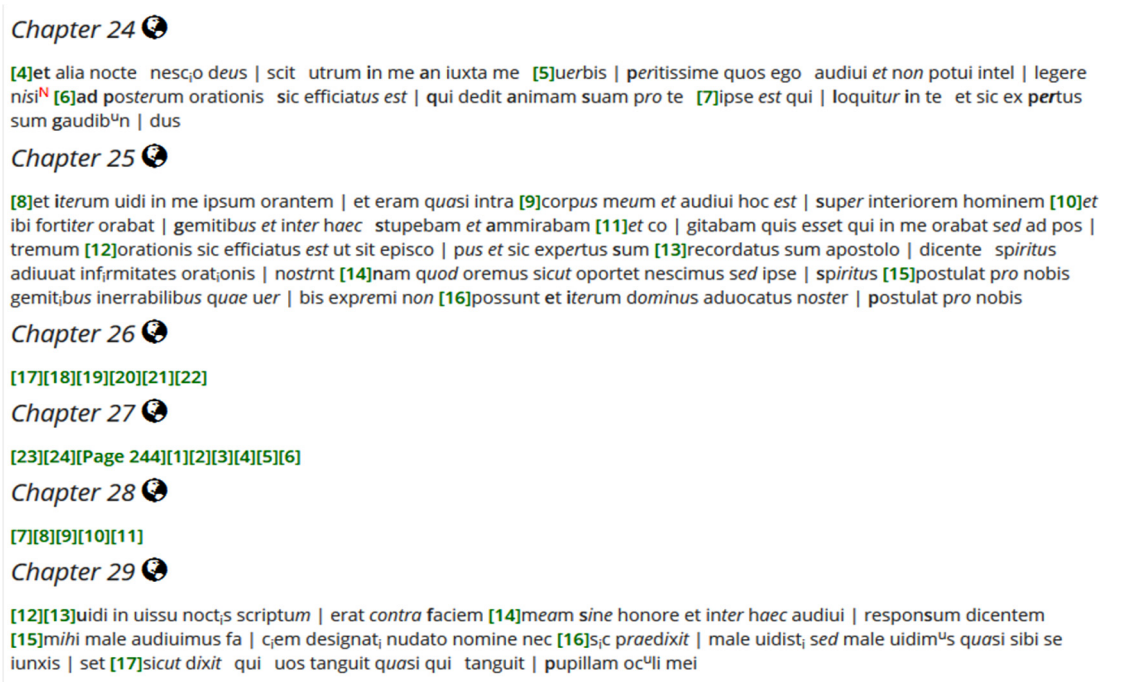



*Fig. 2: Screenshot of the prototype of the digital documentary edition showing the transcription of folio 72r of the Paris manuscript witness. Capital letters and punctuation are highlighted.*

Finally, the inclusion of White's canonical text structure allows for interaction with editions and translations that follow White's system. One way that shall facilitate this interaction is the presentation of the text of each transcription following the canonical chapters. In that way the reader can spot immediately where in a manuscript text is missing. Even if the text is missing hyperlinks corresponding to White's pages and line numbers are present that allow to explore what is written at this point in the other manuscripts. In this presentation the canonical structure can easily be compared across all manuscript witnesses and a user can explore missing passages and textual variation. For instance, a user can immediately see the big sections of text missing in the Dublin *Confessio* in the Book of Armagh (Fig. 3), the missing folios of the Arras manuscript or what parts of text are damaged in the Rouen manuscript which was the victim of a fire.


Chapter 24

[4]et alia nocte nesc$_i$o deus | scit utrum in me an iuxta me [5]uerbis | peritissime quos ego audiui *et non* potui intel | legere nisi$^N$ [6]ad posterum orationis sic efficiatus *est* | qui dedit animam suam pro te [7]ipse *est* qui | loquitur in te et sic ex pertus sum gaudib$^u$n | dus

Chapter 25

[8]et iterum uidi in me ipsum orantem | et eram quasi intra [9]corpus meum *et* audiui hoc *est* | super interiorem hominem [10]*et* ibi fortiter orabat | gemitibus *et* inter haec stupebam *et* ammirabam [11]*et* co | gitabam quis esset qui in me orabat sed ad pos | tremum [12]orationis sic efficiatus *est* ut sit episco | pus *et* sic expertus sum [13]recordatus sum apostolo | dicente spiritus adiuuat inf$_i$rmitates orat$_i$onis | nostrnt [14]nam quod oremus sicut oportet nescimus sed ipse | spiritus [15]postulat pro nobis gemit$_i$bus inerrabilibus quae uer | bis expremi non [16]possunt et iterum dominus aduocatus noster | postulat pro nobis

Chapter 26

[17][18][19][20][21][22]

Chapter 27

[23][24][Page 244][1][2][3][4][5][6]

Chapter 28

[7][8][9][10][11]

Chapter 29

[12][13]uidi in uissu noct$_i$s scriptum | erat *contra* faciem [14]meam sine honore et inter haec audiui | responsum dicentem [15]mihi male audiuimus fa | c$_i$em designat$_i$ nudato nomine nec [16]s$_i$c praedixit | male uidist$_i$ sed male uidim$^u$s quasi sibi se iunxis | set [17]sicut dixit qui uos tanguit quasi qui tanguit | pupillam oc$^u$li mei

*Fig. 3: Screenshot of the prototype of the digital documentary edition showing the transcription of the Dublin manuscript following the canonical text structure. This view clearly shows that in this manuscript the text of three chapters (26, 27, 28) of St. Patrick's Confessio are missing.*

The new edition also tries to use existing online resources by linking to them. For instance, it does not contain any images of the manuscript witnesses, because high-resolution images are already available online on the Saint Patrick's *Confessio* Hypertext Stack. Instead of reproducing these images in the diplomatic view each

folio provides a link the digital facsimile on the Saint Patrick's *Confessio* Hypertext Stack. Similarly are all markers of the canonical chapters in the canonical view linked to the same chapter in the electronic version of Bieler's text-critical edition (with three apparatus) on the Saint Patrick's *Confessio* Hypertext Stack. In that way a user of the new edition can also take advantage of the already existing digital resources about Patrick and the manuscript witnesses provided by this Royal Irish Academy project.

## Conclusion

This article discussed the rationale behind the documentary editing of the manuscript witnesses of St Patrick's epistles and some features of the new digital documentary edition currently being developed. While in the last two hundred years several text-critical editions of *Confessio* and *Epistola* were produced, a digital documentary edition is something new and was still missing. Such an edition is necessary as important text-structural questions have been raised about *Confessio* and *Epistola* in the last decades and in order to bring these discussion to the next level a close study of the actual manuscript evidence together with existing editorial suggestions is crucial.

The digital documentary edition was designed for two main purposes: Firstly, to allow a close study of individual manuscript witnesses, their layout and text-structural features and 'scribal accidentals'. Secondly, to function as a mediator between the manuscripts and text-critical print editions and translations. For this second goal features from White's canonical edition were included into the TEI encoded transcriptions such as canonical chapters. As was shown, trying to model the manuscript layout and canonical chapter in the same TEI/XML document cannot be done in the same way as this would violate the XML nesting rule. The paper also discussed how this issue was escaped using empty elements.

The new digital edition comes with a number of benefits provided by the underlying TEI data model. First and foremost, providing a tool for automated study of the text structure of the individual manuscript witnesses and various scribal activities in the manuscripts which will lead to a better understanding of these documents and how they were produced by medieval scribes. An additional benefit of TEI/XML is that it can be transformed into any number of visualisations and views. A diplomatic and a canonical view will be possible access points for a user to explore the manuscripts and make cross-comparisons between the manuscript witnesses and other editions of the texts. In that way the new edition understands itself as a research tool that between the manuscripts, editions and translations based on the canonical text structure. Consequently, through the new documentary edition, St Patrick's *Confessio* and *Epistola* can be researched in ways that have never previously been possible.

**Works Cited**

Bleier, Roman. Encoding St Patrick's epistles: history and electronic editing of the manuscript witnesses. Dissertation, University of Dublin, Trinity College, 2016.

Bleier, Roman. “Re-Examining the Function of St Patrick's Writings in the Early Medieval Tradition.” *Proceedings of the Royal Irish Academy. Section C: Archaeology, Celtic Studies, History, Linguistics, Literature*, vol. 116C, 2016, pp. 95–117, www.jstor.org/stable/10.3318/priac.2016.116.07. Accessed 11 April 2017.

Binchy, D. A. “Patrick and His Biographers: Ancient and Modern.” *Studia Hibernica*, no. 2, 1962, pp. 7–173, www.jstor.org/stable/20495717. Accessed 11 April 2017.

Dumville, David N. “The form of St Patrick's Confessio in the ‘Book of Armagh’.” *Saint Patrick, AD 493–1993*, edited by D.N. Dumville et al., Boydell, 1993, pp. 191–202.

Gabler, Hans Walter. “The primacy of the document in editing.” Ecdotica, vol. 4, 2007, pp. 197–207.

Gwynn, John. *Liber Ardmachanus = The Book of Armagh*. Hodges, Figgis & Co. Limited and Williams & Norgate, 1913.

Hood, A.B.E. *St. Patrick: his writings and Muirchu’s life.* Phillimore, 1978.

International Federation of Library Associations and Institutions. Functional Requirements for Bibliographic Records. 2009, https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf. Accessed 11 April 2017.

Kalvesmaki, Joel. “Canonical References in Electronic Texts: Rationale and Best Practices.” *Digital Humanities Quarterly*, vol. 8, no. 2, 2014, http://www.digitalhumanities.org/dhq/vol/8/2/000181/000181.html. Accessed 11 April 2017.

O'Loughlin, Thomas. *Discovering Saint Patrick*. Paulist Press, 2005.

Patrick. “Confession et lettre à Coroticus.” Edited by R.P.C. Hanson, Sources chrétiennes, vol. 249, 1978.

Patrick. *Liber Epistolarum Sancti Patricii Episcopi = The book of letters of Saint Patrick the bishop.* Edited by D.R. Howlett, Four Courts Press, 1994.

Patrick. *Libri epistolarum sancti Patricii episcopi*. Edited by Ludwig Bieler, vol. 1, Irish Manuscripts Commission, 1952.

Pierazzo, Elena. “A rationale of digital documentary editions.” *Literary and Linguistic Computing*, vol. 26, no. 4, 2011, pp. 463–477.

Pierazzo, Elena. “Digital Documentary Editions and the Others.” *The Annual of the Association for Documentary Editing*, vol. 35, 2014, http://scholarlyediting.org/2014/essays/essay.pierazzo.html. Accessed 11 April 2017.

Pierazzo, Elena and Peter Stokes, “Putting the text back into context: a codicological approach to manuscript transcription.” *Kodikologie und Paläographie im digitalen Zeitalter*, vol. 2, edited by F. Fischer, C. Fritze and G. Vogeler, Books on Demand, pp. 397–429.

Saenger, Paul. Space between Words: The Origins of Silent. Stanford University Press, 1997.

Sharpe, Richard, "Palaeographical considerations in the study of the Patrician documents in the Book of Armagh." *Scriptorium*, vol. 36, 1982, pp. 3–28.

White, Newport J. D. "Libri Sancti Patricii: The Latin Writings of Saint Patrick." *Proceedings of the Royal Irish Academy. Section C: Archaeology, Celtic Studies, History, Linguistics, Literature*, vol. 25, 1904, pp. 201-i., www.jstor.org/stable/25502726.

White, Newport J. D. "The Paris Manuscript of St. Patrick's Latin Writings." *Proceedings of the Royal Irish Academy. Section C: Archaeology, Celtic Studies, History, Linguistics, Literature*, vol. 25, 1904, pp. 542–552, www.jstor.org/stable/25502730. Accessed 11 April 2017.

**Electronic Sources**

"20 Non-hierarchical Structures." *P5: Guidelines for Electronic Text Encoding and Interchange, Version 3.1.0, Last updated on 15th December 2016,* http://www.tei-c.org/release/doc/tei-p5-doc/en/html/NH.html. Accessed 5 May 2017.

*P5: Guidelines for Electronic Text Encoding and Interchange, Version 3.1.0, Last updated on 15th December 2016*, http://www.tei-c.org/release/doc/tei-p5-doc/en/html/. Accessed 5 May 2017.

*Medieval Unicode Font Initiative (MUFI)*. http://folk.uib.no/hnooh/mufi/. Accessed 5 May 2017.

*Saint Patrick's Confessio Hypertext Stack*. The Royal Irish Academy, 2011, http://www.confessio.ie/. Accessed 5 May 2017.

*Studia UBB* Digitalia, Volume 62 (LXII) 2017, June, Issue 1, 26-41

Published Online: 2017-06-30

DOI:10.24193/subbdigitalia.2017.1.02

# Lost in translation? The odyssey of 'digital humanities' in French

Claire Clivaz

Swiss Institute of Bioinformatics, Switzerland,
E-mail: claire.clivaz@sib.swiss

**Abstract:** By examining the case of the French translation of the expression "digital humanities" (DH), this article argues that cultural diversity and multilingualism could be fostered in digital culture. If other languages have been invited and forced to welcome this English phrase, its translations have to be studied since they could potentially have strong epistemological backwash-effects on it. Through an historical etymological inquiry, it can be demonstrated that the use of the outmoded French word *humanités* is the most significant element in the two French expressions *humanités numériques* or *humanités digitales*. This single word opens up a specific space for humanist approaches within the open-ended digital approaches. On this base, the encounter between Humanities and hard sciences can be reconsidered, as it happens already in two examples of new DH masters in French-speaking countries.

**Keywords:** Digital Humanities; French language; Multilinguisme; epistemology; Humanities; Master in DH; Etymology.

*To my late mother,*
*who read so many books aloud to me,*
*building my cultural memory*
*of the forgotten meanings of words*

By examining the case of the French translation of the expression "digital humanities" (DH), this article argues that cultural diversity and multilingualism could be fostered in digital culture. At first glance, the international success of this expression seems to contradict this statement: isn't it a clear example of English language domination over other Western and non-Western languages? Used in written form for

the first time in 2004 (Kirschenbaum 56), tirelessly discussed in DH conferences and works, “DH” has quickly been used in professorship titles, in undergraduate and postgraduate degrees, or to qualify centers, laboratories, and research projects (Author “Title1” 41). If other languages have been invited and forced to welcome this expression, its translations have to be studied since they could potentially have strong epistemological backwash-effects on it. French is an example worth examining: it can be demonstrated that the use of the outmoded French word *humanités* is the most significant element in the two French expressions *humanités numériques* or *humanités digitales*. This single word opens up a specific space for humanist approaches within the open-ended digital approaches.

The introduction below aims to present the specific impact of a study of the phrase “digital humanities” and its translations within the general problematic of the phrase's definition. The second part of this article summarizes the main progressions and arguments in the discussions surrounding *humanités numériques* (humanities computing) and *humanités digitales* (digital humanities) in the French-speaking sphere. The third section examines the historical epistemology of *humanités* while the final section considers the resulting confrontation between the humanities and the ‘hard’ sciences: this underlines their potential synergy and the proper role of the humanities.

## 1. Introduction

The multitude of debates around the definition and delimitation of “digital humanities” are well-known and prolific. Nevertheless, a clearly prominent progression is an enlargement of the notion during the last few years. In 2009, Patrick Sevensson proposed that there was a “discursive shift from humanities computing to digital humanities”. He suggested that tensions were born from the relative lack of humanities computing engagement with the “digital as a study of object” (“Humanities Computing”). Seven years later, in the second volume of *Digital Debates*, Steven Jones chose to use the word “everted” to qualify the extension, and even the overall presence, of the digital:

> New practices and areas of interest for computing in the humanities correspond to changes associated with the eversion of cyberspace in the culture at large. In one sense, the new digital humanities is humanities computing, everted. […] The term also reflected a larger change: from implying a separation between the stuff of the humanities – manuscripts, books, documents, maps, works of art of all kinds, other cultural artifacts – and computing, to more of a mixed reality, characterized by two-way interactions between the two realms, physical artifacts and digital media. (“The Emergence”)

This “everting digital” is also clearly represented in Bernard Stiegler’s definition of the expression he prefers to DH: “digital studies”, which has a sociological flavor. Indeed, he considers the digital studies as a total global phenomenon: “absolutely all the fields are concerned: all practices, all forms of life, all enterprises, personal or collective” (221).[1] The digital studies should have the purpose to produce “new institutional programs as well as industrial” (218). In questioning or even resisting the overall presence of an “everting digital”, one should examine the words themselves to find limits and complexity, while keeping in mind the work of Jacques Derrida, *Sauf le nom*. His work explains depth of the names, the words: “The name: What does one call thus? What does one understand under the name of name? And what occurs when one gives a name? What does one give then? One does not offer a thing, one delivers nothing, and still something comes to be, which comes down to giving that which one does not have”. (Derrida XIV)

Consider thus whether something *happens* to the English expression “digital humanities” when other languages try to find their own words to translate it: something “comes down to giving that which one does not have” (Derrida XIV). Consider this in the context of French hesitating between “humanités numériques” and “humanités digitales”, or when Hebrew chooses to speak about *rouach digitalit*, the “digital spirit” to translate DH. The shock that occurs between languages when it comes to DH is sometimes so significant that “digital humanities” can remain untranslated, like a type un-embeddable UFO: the *Manifeste des digital humanities*, written in Paris in May 2010, kept the English expression (Dacos “Manifeste”). The French scholar Olivier Le Deuff relates how he began by using in French writings the English expression “digital humanities” before daring to use *humanités digitales* (“Humanités digitaltes” 263). Institutions too, sometimes prefer to keep the English phrase, like the national Austrian DH center, the “Austrian Center for Digital Humanities”: all the other institutions on the webpage devoted to the departments in the Arts, Humanities and Social Sciences have German titles (*Geistes-, Sozial- und Kulturwissenschaften*). One gets the impression that DH bears something that is not embeddable in the traditional Humanities.

Digital culture definitively has effects on different languages. The linguistic experiment done at the DH2014 conference enlightened this aspect: the English call for papers was translated into 23 languages, and several of these translations had to choose words to translate the English call (*DH2014 Multilingual Calls*). For example David Wrisley and his team in Lebanon had to find Arabic terms and expressions to give life to DH concepts and landscapes. Such an experiment clearly demonstrated the English constraint on other languages but it also represented a real opportunity for

[1] English versions of French bibliographical references are my own translations. I thank Harley Edwards for the English proof-reading of this article. This article is based on a paper presented in Warsaw on the 10th of May 2017 at the CLE2017 conference.

those languages. The French example shows the potential backwash-effect of the translations of “digital humanities”. Similar inquiries should be done in as many languages as possible.

## 2. A French language odyssey: “humanités digitales” versus “humanités numériques”

The French expression “humanités digitales” is a direct translation of “digital humanities”, first used in 2008 by Valérie Carayol and her colleagues in Bordeaux (Le Deuff “Humanités digitales” 263). Simultaneously, another team on the Lausanne campus began to speak about *humanités digitales* (Author “title 2”). In 2015, Carayol and Morandi published a collection of essays that was supposed to be entitled *Humanités digitales* (“Humanités numériques” 1), but they finally published it as *Le tournant numérique des sciences humaines et sociales*. The collection’s brief introduction thus mentions only “humanités digitales”, without commenting on the discrepancy between the introduction and the title of the book (9–12). Such hesitations, changes, or adaptations between *humanités digitales* vs *numériques*, are visible in diverse circumstances, such as in the titles of the academic degrees for example. In 2015, the University of Lausanne (CH) first started a master's degree specialization entitled “Humanités digitales”, in 2016 it was replaced by a complete and inter-faculty master's degree entitled “Humanités numériques” (*Humanités digitales*). In 2014, Milad Doueihi noticed, not without a certain Parisian disdain, that *humanités digitales* was used in “certain regions of the French-speaking sphere” (8).

In 2016, Dominique Vinck estimated that “the expression *humanités digitales* is probably going to be abandoned at the expanse of the intellectual project associated to its use” (9–10). Inversely, also in 2016, Olivier Le Deuff wrote a blog post entitled “Ten reasons to prefer digital to numeric”, whereas the new journal *Etudes digitales* devoted several pages of its first issue to the debate, with scholars globally supporting *humanités digitales* (13-15, 215-227, 251-268). Previously, ThatCamp 2013 in St-Malo (FR), a French language event, welcomed the pre-general assembly of the future French-speaking DH association. Interestingly, there was a vigorous discussion regarding the name of the association and the distinction between *numérique* and *digital*, the record thereof has been collectively written in online proceedings. A range of arguments were put forth: from Charlotte Touati underlining that German is never afraid to welcome English words, to Aurélien Berra lucidly recalling that the attribution of budgets is often dependent on labels and names (“Humanités numériques” 17). General agreement was reached for nominating a name for the new association, in the end neither *humanités numériques* nor *humanités digitales* was chosen, instead a third option came to the fore. *Humanistica*,

proposed by Marin Dacos (“Humanités numériques” 19), was later confirmed as the association’s name (*Humanistica*).

Is such a recent, brief and intense history a signal of “French coquetry” or does it point to something more important? Instead of dismissing this as mere ‘coquetry’, one should consider that deep epistemological dimensions are at stake in this complex history of “digital humanities” reception in French. Names and language are such a finely cultural phenomenon that it is here only possible to point to some milestones; more can be achieved only within the framework of a collective inquiry. However, there is no doubt that such an inquiry is needed, and hopefully it will occur in diverse languages beyond the examination of only the current French example.

Among the scholars inclined to speak about *humanités digitales* or *études digitales*, a huge consensus appears about the fleshly aspect present in the term *digital/e* that means in English the digit and the finger, the first tool used to count[2]. For Bernard Stiegler, *digital/e* joined to the computing vocabulary underlines in French “*the hand's role, that of its fingers and of what they create*, in the generation of thought via its grasp on life, creation and conception” (217). Whilst the collective *laudatio* of the corporeal aspect of *humanités digitales* is well founded, it is nevertheless surprising that only a few scholars have noticed the return of the outmoded French word *humanités*.[3] In 2012, Aurélien Berra, recalling the Renaissance roots of the expression “faire ses humanités” (to study humanities), argued that by translating humanities as *humanités*, the French would enlarge the English academic definition by including the social sciences (3).

However, this is not an easy or light affirmation, indeed, the DH international milieu traditionally does not include the social sciences in its structures such as ADHO or its constituent organizations. At the level of ERIC – the *European Research Infrastructure Consortium* – the Humanities fall under DARIAH and CLARIN, whereas the social sciences are present mainly in CESSDA (*ADHO, CESSDA, CLARIN, DARIAH, ERIC*). While Berra comes to this conclusion by scrutinizing the word ‘*humanités*’, Cormerais and Gilbert similarly expand the definition by focusing on the fleshly dimension of the French word *digital/e*:

> The choice of ‘digital’ is an attempt to reintroduce the sensation of touch, this in turn creates a ‘double digitization’: referring, not only to numbers themselves, but to the contact between digits and the calculated matter. […] We claim this ambivalence between physical carnality [charnalité] and the calculation as the

[2] See for example *Etudes digitales* 13-15, 215-227, 251-268 ; Le Deuff “10 raisons” ; “Humanités numériques” ; Author “Title 1” 52–53.

[3] See notably Berra 2–3, Author and XX 9, Author “Title 4” 255.

> most adequate rendering of the tension inherent to the entirety of digital studies. This is not reducible to either the code or the disciplines that define Humanities. ("Introduction" 14-15)

Such an expansion is also present in English when it comes to the most recent debates about the definition of the DH, as evidenced by section 1. Section 3 allows one to test and verify a specific example of a definition by immersing oneself in the history of the French word *humanités*: Is the corporeal dimension evoked by the word '*digital/e*' a recent addition to what would otherwise be a strictly cerebral concept, *humanités*? Is there no limit to the digital studies implied by the *humanités*, which themselves include the Arts and Humanities and the Social Sciences? Can everything be included in the new modes of digital knowledge? One should focus on words and their philological history to orientate oneself more clearly with respect to these questions, if not their answers.

### 3. The word 'humanités' and its forgotten meanings

The outdated French word '*humanités*' was almost forgotten until its recent association with computing. At least one part of the word's meaning has been attested to since the 17th century and defined as such in 1831 by the *Société des grammairiens*: "one names *Humanités* what is usually taught at the colleges until philosophy exclusively" (89). Looking backwards at the long history of this word is impressive with regards its diverse meanings. The ancient French dictionary (9th to 15th century) indicates the following meanings: "profane goods", "active feeling of benevolence, or alms" (Godefroy IX 774). However the basic senses at the time referred to "life", the "body", and even "sexual parts" or "shameful parts" (Godefroy IV 526). To know somebody *humainement* means "carnally" (*charnellement*) (Godefroy IX 774; IV 526). In the early 16th century, the womb of a woman was also referred to as *le moule d'humanité*, "the mold of humanity" (Huguet "Humanité").

As astonishing as it is, while many scholars emphasize the corporeal elements related to the use of the word '*digital/e*' in its relation to the word '*humanités*', they all ignore that such potential traces belong even more so to the history of the word '*humanité*' itself. Whereas Cormerais and Gilbert relate *digital* to "carnality" (*charnalité* "Introduction" 15), the association between *humanité* and carnality is at least as profound. How can cultural memory – in this case Western –forget meanings, or stop using words to describe certain things, according to the Foucauldian relationship between words and things? In fact, the corporeal and carnal traces in '*humanités*' can rarely still be found in our present cultural expressions. From the pen of a psychiatrist,

Marie-Rose Moro, who in 2008 wrote an article about female bodies referencing the corporeal aspects of '*humanités*' in the title.

When was the corporeal element of '*humanité*' forgotten by mainstream Western cultural memory, to remain only as a tiny and seldom remembered *souvenir*? A Foucauldian analysis of the double epistemological Western turn proves to be valid in answering this question (Foucault 42-47)[4]: in the Classical Age, in the preliminary edition of the *Dictionnaire de l'Académie* in 1687, one finds the simultaneous emergence of the scholarly meaning of *humanités* (Classical training), and the disappearance of its carnal meaning, a situation that remains in the editions that came thereafter. In the first middle of the 19th century, however, language took its revenge and the corporeal or carnal meaning of humanité found refuge in the Complement to the Academy *Dictionary* published by Louis Barré in 1842: "Humanité: Expr. prov., *Reposer son* humanité, Se mettre à son aise. // Avoir humanité, se disait autrefois pour, Être en vie. // *Humanité*, s'est dit Des parties sexuelles" (586).

Barré, in his definition, explicitly explains that the corporeal and sexual meanings in '*humanité'* were employed in previous eras (*autrefois*, *s'est dit*); he nevertheless indicates them in his "complementary" definition. His complementary dictionary is a delicate exercise in scholarly communication: the 6th edition of the *Dictionnaire de l'Académie* in 1835 provoked polemics and Barré had to justify why he created a complementary dictionary without hurting the French Academy authorities too greatly (Marosavri 95). It is particularly interesting to notice that the corporeal and carnal traces of *'humanité*' subsist in the 19th scholarly memory in this complementary dictionary written by a professor of philosophy (Barré), whereas the philological domain takes power over the word '*humanité*' in the official editions of the *Dictionnaire de l'Académie*. Deep shifts in the linguistic tectonic plates are at stake here.

When one considers the near disappearance of the corporeal and carnal elements of '*humanité*', combined with the focus on the corporeal elements of '*digital/e*' in French, and the stubborn presence of *humanités digitales* behind all linguistic critics, one may perceive new possibilities of understanding the recurrent difficulty we have in linking *materiality* to the digital culture (Author "Title 3"). In his provocative 2009 interview about the future of books, Umberto Eco considers the Internet as radically different from writing because "the writing is the continuation of the hand and is in this sense is almost biologic. […] Internet is not biologic" (22), whereas in that same year Robert Darnton pointed to the illuminating German word '*Fingerspitzengefühl*' (Author Title 1 53). Darnton uses it to explain the new relationship emerging between our bodies and digital matter:

[4] For a presentation of the Foucauldian Western *episteme* pattern, see Author *Title 5* 99–101.

> our way through the world by means of a sensory disposition that the Germans call *Fingerspitzengefühl*. If you were trained to guide a pen with your finger index, look at the way young people use their thumbs on phones, and you will see how technology penetrates a new generation, body and soul. (XII)

Darnton's linguistic intuition was more than true: at the end of 2014, biological research has shed light on "use-dependent cortical processing from fingertips in touchscreen phone users" (Gindrat et al.). The "corporeal gap" between these great thinkers in their evaluation of the digital world can now be better understood in the context of the different traces present in '*humanité*'. In French, as long as we continue to focus on '*digital/e*' and its relation to the human body and to forget the carnal dimensions of '*humanité(s)*', we will continue to be unable to join the digital domain to our sense of materiality. Bridging this gap and considering '*humanité*' as relevant to bodies and life becomes more urgent since this bridge could lead to a better grasp of the specific role of Humanities in global knowledge, section 4 focuses on this notion. Furthermore, this bridge is of necessity since digital humanities has a certain chance of losing its 'digital' qualification in the next few years or generations.

DH is not a field as such but a necessary label for the present generations: we need it to be more conscious of the immense epistemological turn we are facing. However, if we consider what has happened to the expression 'digital computer', we can argue that at a certain point in time within the next few years or decades the word 'digital' might no longer be joined to the word 'Humanities'. The first written trace we have of the junction between the English words 'digital' and 'computer' as opposed to an analogue computer, goes back to a 1942 scientific report by George Robert Stibitz (William 310, Dennhardt). Similarly, scholars use the phrase 'digital humanities' to speak about what are not considered the usual humanities. In 1950, Turing was still using the complete phrase 'digital computer' in his seminal article "Computing Machinery and Intelligence": "The reader must accept it as a fact that digital computers can be constructed, and indeed have been constructed, according to the principles we have described, and that they can in fact mimic the actions of a human computer very closely" (5).

When did we stop speaking about 'digital computers' and start simply saying 'computers'? Such questions of linguistics generally remain without clear answers. In future, surely similar questions will be asked with regards to the digital humanities. If in many years or decades to come we speak again about 'humanities', will humanities be considered evidently digital. If this is to be, it will demonstrate that the important word in the phrase '*humanités digitales*' is the former, and that humanities was not condemned to be absorbed into an open-ended digital studies domain.

## 4. The face-off between the hard sciences and the (digitized) humanities

### *4.1 Observing the knowledge discourse*

On the 5th of July 1919, the *British Journal of Medicine* published a lecture given by William Osler, often called the father of modern medicine, in front of Oxford Classical scholars: "The old humanities and the new science". In a way typical of the modern period, Osler recognized an "unhappy divorce" between humanities and sciences:

> We make boys and young men spend ten or more years on the study of Greek and Latin, at the end of which time the beauties of the languages are still hidden because of the pernicious method in which they are taught. […] The so-called Humanists have not enough science, and Science sadly lacks the Humanities. This unhappy divorce, which should never have taken place, has been officially recognized. (4)

Full of good will, and fascinated by the Humanities, Osler presented an persuasive metaphor: scientists are like brave working ants and as faithful nurses, they take care of the ant larvae that are the Humanists. Ant nurses are awarded with the honey produced by the larvae: similarly, Humanists provide nourishment to intellectual life (3). Secreting a nice, sweet substance, the Humanities are the "hormones" of intellectual life (3). Trying to overcome the "unhappy divorce", he reminds us that "in biology Aristotle speaks for the first time the language of modern science, and indeed he seems to have been first and foremost a biologist, and his natural history studies influenced profoundly his sociology, his psychology, and his philosophy in general" (4).

There is no doubt that Osler was genuinely attached to the Humanities and attempting to honor them, but the point of view that he presents confirmed the divorce between fields: "the new science" – perceived as brave working ants, and the Humanities – whatever old or new – perceived as producing something nourishing but as larvae. We know the evolution that occurred in the following decades of the 20th century: the pre-digital age saw several fields of the Humanities almost starved to institutional death. Bernard Stiegler rightly reminds us that the digital turn arrived within the context of a deep crisis for all the academic educational systems (215).

During this "unhappy divorce", the Humanities have progressed along certain paths away from the hard sciences. Humanist scholars have convinced themselves of this divorce, so much so that we can sometimes read resulting simplifications: Lorna Hughes, Panos Constantinopoulos and Costis Dallas in 2016 quote a 2000 paper by John Unsworth. They transmit Unsworth's information as if he was presenting "scholarly primitives" as the "common elements of humanistic inquiry" (5725): "Definitions of digital

humanities are as prolific as the field itself […], but frequently cited as an initial conceptual framework are the 'scholarly primitives'. This was used in the contest of the digital humanities by Unsworth (2000) to denote basic functions that have been common to scholarship across the disciplines: discovering, annotating, comparing, referring, sampling, illustrating and representing" (Hughes et al. 5706). By reading this list, it would seem that hard scientists are definitely using the same term "scholarly primitives". Unsworth is indeed referring to axioms used in mathematics as well as in philosophy, to elements coming from Aristotle. As he wrote in 2000:

> "According to Aristotle, scientific knowledge (*episteme*) must be expressed in statements that follow deductively from a finite list of self-evident statements (axioms) and only employ terms defined from a finite list of self-understood terms (primitives). [Stanford Encyclopedia of Philosophy] […] These 'self-understood' functions form the basis for higher-level scholarly projects, arguments, statements, interpretations–in terms of our original, mathematical/philosophical analogy, axioms. My list of scholarly primitives is not meant to be exhaustive, I won't give each of them equal attention today, and I would welcome suggested additions and debate over alterations or deletions, but here's a starting point". ("Scholarly Primitives")

The unhappy divorce between the Humanities and the hard sciences has so invaded our cultural minds that we can easily forget what there is in common, at least from an epistemological history point of view. We might have forgotten that Aristotle's axioms concerned all the episteme, and that he was a biologist as much as a philosopher. The growing definition of the fields that we observe in digital culture is a real opportunity to recombine pieces of knowledge that modern times have divided. Thanks to the French return of *humanités*, it forces us to remember that language analysis and philology rely on a corporeal heritage, as put forth by Lakoff and Johnson (*Philosophy in the Flesh*). *Corpus* and *corpora* are related, that's the gift of the body, or carnal, *humanité*.

How then could one distinguish between differences and complementarities, between what is or not Humanities or its enlarged, French version '*les humanités*'*?* In these digital times this become an urgent question because new curricula are already being developed, as the section 4.2 underlines: *things* (the signified) are present mainly before *words* (the signifiers) in digital culture, according to the Foucauldian binome word/thing. To discern limits and specificities of *les humanités*, one should return to the period anterior to Osler's diagnostic of unhappy divorce. One should return to Louis Barré's iconoclastic *Complément au Dictionnaire de l'Académie* (1842).

Barré explains the "Humanists" – those who consider that Classical languages should build the basis of teaching – as a label existing in Germany and opposed to the *Realists* (586). In France, he recognizes a similar opposition between *les classiques et les industriels*, the "Classics" and the "industrials" (586). It is absolutely necessary to realize that the tension between the Humanists and others in this context is based on their point of view on the *realia*. We can verify that such a tension/opposition is still at stake by keeping in mind Stiegler's aspiration to see digital studies producing "new institutional programs as well as industrial" (218, quoted in part 1). If almost all knowledge approaches seek to understand the human being, its body and its environment, *les humanités* analyses the *realia* with a particular point of view.

*Les humanités* – beyond the English Humanities – are definitively attached to asymmetry, details, complexity and plurality or even to divergent plurality. Digital humanities is the act of studying and practicing "ontologies", always within the plural and with a strong consciousness of the potential "ontology" trap. Even taking into account this plural aspect, the Humanist attachment to asymmetry requires attention and resistance (Author Title 1 48-52). However, the singular form of "ontology" – a successful word even in the hard sciences – seems to be the rule in biomedical investigations (*Ontology for Biomedical Investigations*). From a "realist" and/or "industrial" perspective, the challenge is to reduce the perturbing diversity of the *realia* in order to take care of disturbed or disrupted human bodies and organisms. Whereas the binary system (0/1) is the basic language that stands at the roots of the computing sciences – an electronic system that is either open or closed – *les humanités* will always be looking for a third term and for asymmetry. *Les humanités* are interested in disturbing *realia*, such as closed but non-functioning electronic systems. It is a "culture of interpretation" according to the French thinker Yves Citton (21).

While attempting to outline the Humanities posture, the *real* world progresses to another step: new master curricula are being developed, claiming to train students far beyond the usual delimitation of the academic fields. Even so, it remains seemingly possible to discern an attempted outline to *les humanités*. Consider two different examples of interdisciplinary master's degree, one born from a Humanities background at the University of Rouen (FR) and the other stemming from a Computing background at the EPFL, the *Federal Polytechnic School of Lausanne* (CH).

### *4.2 Observing the emergence of new curricula at the frontiers of fields*

The University of Rouen has just announced a new master's degree program in digital humanities or humanities and the digital world, *master en humanités numériques ou humanités et monde numérique* (Massin). It is preceded by a *Licence Humanités*, a bachelor's degree in Humanities, in which the French word *humanités* signals an

humanist interdisciplinary perspective: it includes History, Letters and Political Sciences (*Licence Humanités*). It corresponds to the wishes for a general bachelor degree formulated by Yves Citton in his book *L'avenir des Humanités* (2044). The master's degree in DH is thought of as either the next step after the bachelor's in *Humanités*, or for computing students wishing to be trained in the interface with the humanities (Massin, "Un nouveau master"). This new master's degree integrates computing scientist and humanist paths. All the following different undergraduate degrees allow one to be enrolled in this postgraduate course: bachelor's in Humanities, Letters, History, Philosophy, Law, Computing, Computing-communication (*Master Humanités numériques*).

It is interesting to highlight that the master title do not necessarily link the humanities to digital, since the second way of phrasing is "Humanities and digital world": the Humanities seem to resist their own perspectives, a point in concordance with the observations in sections 3 and 4.1. It should also be underlined that a computing undergraduate degree does not seem to be missing the basic humanist elements to enter in such a postgraduate. It is not an insignificant choice. Indeed, in the EPFL *Master of science in digital humanities*, bachelors in Humanities are not accepted as pre-requisite:

Candidates must have Bachelor Degree in Computer Science or in Communication Systems from a recognized university, with excellent academic records. A Bachelor Degree in Mathematics, Physics, Electrical Engineering, Micro Engineering or related field may also be accepted. You must have solid basis of programming, algebra, statistics, signal processing. A strong interest in the digital humanities is expected. (*Master of Science in DH*)

Such a discrepancy should be quickly noticed by Humanist scholars: are we really considering that everybody can jump over our bachelor in Humanities, whereas a master of Science labeled "DH" is not accessible without an undergraduate in the hard sciences? It is clearly worth considering closely and urgently since this influences the definitions and perceptions we have of the Humanities. If a whole three years of Humanities undergraduate study are considered equivalent to less than a semester of complementary of studies, it signals that these undergraduate degrees should probably be revised. An intense, collective discussion on this point is needed.

This intriguing EPFL master's degree in science and digital humanities, opening in autumn 2017, requires further examination. It has an innovative profile that opens new gates to future academic training, if one infers according to the current website. It has an innovative profile that opens new gates to future academic trainings, as far as it is possible to read it on the website. L*es humanités* are clearly not disappearing at all; they have not dissolved in a digital studies 'soup'. To the contrary, many of the greatest names in DH are proposing courses gathered in a "Humanities block" (David Boullier,

Sarah Kenderdine, Franco Moretti and Martin Rohrmeier), whereas the "Computing block" presents courses given by computing colleagues (Daniel Gatica-Perez, Frédéric Kaplan, Sabine Süsstrunk, Robert West; *Master of science in DH*).

What kind of knowledge will come out of this new postgraduate program? It is really hard to say at this step, but such an innovation should be encouraged, monitored and analyzed especially if one considers it to be the "fabric of the future knowledge". *Les humanités* clearly manages to maintain an autonomous role, as in the *Master en humanités numériques* at the Rouen University. Such master's degrees, born on French-speaking grounds, reveal that the French-speaking sphere is reworking and transforming the English label 'digital humanities'. *Les humanités sont de retour*, *les humanités* are back, but reshaped by digital culture. Rediscovering their corporeal vocabulary heritage, they will soon be ready to fully plan their role in digitized studies on this strange global phenomenon, the human.

## Works cited

ADHO, https://www.adho.org. Accessed 14 May 2017.

Barré, Louis. "Humanité", "Humaniste." Complément du Dictionnaire de l'Académie, Firmin Didot frères, Paris, 1842, p. 586. https://archive.org/details/fre_b1886109. Accessed 14 May 2017.

Berra, Aurélien. "Faire des Humanités Numériques." Read/Write Book, vol. 2, edited by Pierre Mounier, OpenEdition Press, 2012. http://books.openedition.org/oep/238. Accessed 14 May 2017.

Carayol, Valérie, and Morandi, Franc (dir.). Le tournant numérique des sciences humaines et sociales, Maison des sciences de l'homme d'Aquitaine, 2015.

Carrière, Jean-Claude, and Eco, Umberto. N'espérez pas vous débarrasser des livres, Seuil, 2009.

CESSDA, https://cessda.net/. Accessed 14 May 2017.

Citton, Yves. L'Avenir des Humanités. Économie de la connaissance ou culture de l'interprétation ?, La Découverte, 2010. Kindle edition.

CLARIN, https://www.clarin.eu/. Accessed 14 May 2017.

Author. Title 5, 2010.

Author. "Title 2" 2011.

Author. "Title 1" 2012.

Author. "Title 4", 2016.

Author. "Title 3" 2016.

Author and XX, 2014.

Cormerais, Franck, and Gilbert, Jacques Athanase (dir.). Etudes Digitales, vol. 1, 2016.

Cormerais, Franck, and Gilbert, Jacques Athanase. “Introduction.” Etudes Digitales, vol. 1, 2016, pp. 11–16.

Dacos, Marin. “Manifeste des Digital humanities. ThatCamp Paris 2010.” http://tcp.hypotheses.org/318. Accessed 14 May 2017.

DARIAH, https://www.dariah.eu/. Accessed 14 May 2017.

Darnton, Robert. The Case for Books: Past, Present, Future, PublicAffairs, 2009.

Dennhardt, Robert. The Term Digital Computer (Stibitz 1942) and the Flip-Flop (Turner 1920), Grin Verlag, 2016.

Derrida, Jacques. On the name, translated by Thomas Dutoit, Standford University Press, 1995.

DH2014, http://www.dh2014.org. Accessed 14 May 2017.

DH2014 Multilingual Calls, https://dh2014.org/2013/08/31/call-for-proposals-online/. Accessed 14 May 2017.

Doueihi, Milad. “Préface. Quête et enquête.” Le temps des humanités digitales, directed by Olivier Le Deuff, FyP éditions, 2014, pp. 7–10.

ERIC, https://ec.europa.eu/research/infrastructures/index_en.cfm?pg=home. Accessed 14 May 2017.

Foucault, Michel. The Order of Things. An archaeology of human sciences, Tavistock, 1970.

Geistes-, Sozial- und Kulturwissenschaften, https://www.oeaw.ac.at/forschung-institute/forschungsinstitute-der-oeaw/geistes-sozial-und-kulturwissenschaften/. Accessed 14 May 2017.

Gindrat, Anne-Dominique et al. “Use-Dependent Cortical Processing from Fingertips in Touchscreen Phone Users.” Current Biology, vol. 25, n° 1, 2015, p. 109-116. doi: http://dx.doi.org/10.1016/j.cub.2014.11.026. Accessed 14 May 2017.

Godefroy, Frédéric. “Humanité.” Dictionnaire de l'ancienne langue française et de tous ses dialectes du IXe au XVe siècle, vol. IV, F. Vieweg, Paris, 1885, p. 526 ; vol. IX, Emile Bouillon, Paris, 1898, p. 774 ; http://gallica.bnf.fr/ark:/12148/bpt6k50634z. Accessed 14 May 2017.

Huguet, Edmond. “Humanité.” Le dictionnaire de référence pour la langue de la Renaissance, du 15e au 17e siècle, 1925-1967, edited by Claude Blum, Classiques Garnier Numérique, Editions Champion Électronique, no year. https://tinyurl.com/l6dsxfc. Accessed 14 May 2017.

Hughes, Lauren et al. “Digital Methods in the Humanities.” A New Companion to Digital Humanities, edited by Susan Schreibman et al., John Wiley & Sons, 2016. Kindle edition. doi: 10.1002/9781118680605.ch11. Accessed 14 May 2017.

Humanités digitales, Université de Lausanne (CH), https://www.unil.ch/lettres/fr/home/menuinst/master-et-specialisation/ma-avec-specialisation/humanites-digitales.html. Accessed 14 May 2017.

"Humanités numériques: une question de lexique." ThatCamp St-Malo 2013. Non actes de la non conférence, La Non-Collection, Editions de la maison de sciences de l'homme, 2014. http://books.openedition.org/editionsmsh/2189. doi : 10.4000/books.editionsmsh.2181. Accessed 14 May 2017.

Humanistica, www.humanisti.ca. Accessed 14 May 2017.

Kirschenbaum, Matthias G. "What Is Digital Humanities and What's It Doing in English Departments?" ADE Bulletin, vol. 150, 2010, pp. 55–61, http://mkirschenbaum.files.wordpress.com/2011/03/ade-final.pdf. Accessed 14 May 2017.

Lakoff, George – Johnson, Mark. Philosophy in the Flesh. The Embodied Mind and Its Challenge to Western Thought, Basic Books, 1999.

Le Deuff, Olivier, "10 raisons de préférer digital à numérique," Le guide des égarés, 2016, http://www.guidedesegares.info/2016/01/31/10-raisons-de-preferer-digital-a-numerique/. Accessed 14 May 2017.

Le Deuff, Olivier. "Humanités digitales versus humanités numériques, les raisons d'un choix." Etudes Digitales, vol. 1, 2016, pp. 263–264.

Le Dictionnaire de l'Académie françoise (de A à Neuf), Le Petit, Paris, 1687. Pre-edition.

Licence Humanités, Université de Rouen (FR), http://lsh.univ-rouen.fr/licence-humanites-543140.kjsp. Accessed 14 May 2017.

Marosvari, Mária. "Barré et la sixième édition du 'Dictionnaire de l'Académie' (1835-1842)." *Romantisme, vol.* 86, 1994, pp. 89–100. doi: 10.3406/roman.1994.5989, www.persee.fr/doc/roman_0048-8593_1994_num_24_86_5989. Accessed 14 May 2017.

Massin, Fabien. "Un nouveau master 'Humanités numériques' à l'Université de Rouen." 76actu, 10 Mai 2017, https://actu.fr/normandie/rouen_76540/un-nouveau-master-humanites-numeriques-luniversite-rouen-normandie_891923.html. Accessed 14 May 2017.

Master Humanités numériques, Humanités et monde numérique, Université de Rouen (FR), http://lsh.univ-rouen.fr/master-humanites-numeriques-humanites-et-monde-numerique-543132.kjsp. Accessed 14 May 2017.

Master of science in digital humanities, EPFL (CH), http://cdh.epfl.ch/dh-master; How to apply, http://cdh.epfl.ch/dh-apply; Study Plan, http://cdh.epfl.ch/dh-studyplan. Accessed 14 May 2017.

Moro, Marie-Rose. "L'infinie matière des humanités. Corps, espace, temps, au féminin pluriel." Cahiers de l'infantile, vol. 6, 2008, pp. 119–129.

Ontology for Biomedical Investigation, http://www.ontobee.org/ontology/OBI?iri=http://purl.obolibrary.org/obo/OBI_0001285. Accessed 14 May 2017.

Osler, William. “The Old Humanities and the New Science.” British Medical Journal, 5th July 1919, pp. 1–7.

Société des grammairiens. “Humanité.” Dictionnaire de l’Académie française A-K, Guillaume, Paris, 1831, p. 89.

Steven, E. Jones. “The Emergence of the Digital Humanities (the Network Is Everting).” Debates in the Digital Humanities, edited by Matthew K. Gold and Lauren Klein, University Minesota Press, 2016. http://dhdebates.gc.cuny.edu/debates/text/52. Accessed 14 May 2017.

Stiegler, Bernard. “Dans la disruption. La main, ses doigts, ce qu’ils fabriquent et au-delà.” Etudes Digitales, vol. 1, 2016, pp. 215–227.

Svensson, Patrick. “Humanities Computing as Digital Humanities.” DHQ, vol. 3, n°3, 2009. http://www.digitalhumanities.org/dhq/vol/3/3/000065/000065.html. Accessed 14 May 2017.

Turing, Alan M. “Computing Machinery and Intelligence.” Mind, vol. 49, 1950, p. 433–460.

Unsworth, John. “Scholarly Primitives: what methods do humanities researchers have in common, and how might our tools reflect this?” Paper presented at a symposium on ‘Humanities Computing: formal methods, experimental practice’, King's College, London, 13 May 2000. http://www.people.virginia.edu/~jmu2m/Kings.5-00/primitives.html. Accessed 14 May 2017.

Vinck, Dominique. Humanités numériques. La culture face aux nouvelles technologies, Le cavalier bleu, 2016.

Williams, Bernard. Computing with Electricity, 1935-1945, PhD Dissertation, University of Kansas, 1984. University Microfilms International, 1987.

*Studia UBB Digitalia*, Volume 62 (LXII) 2017, June, Issue 1, 42-56

Published Online: 2017-06-30

DOI:10.24193/subbdigitalia.2017.1.03

# Stretching the boundaries of publishing: The Open Web Platform and the alternatives

Nicolaie Constantinescu

Information architect, kosson.ro, Romania,
E-mail: kosson@gmail.com

**Abstract:** Today's scientific practice makes data the first territory for analysis and publishing becomes the rich context for cross referencing and interpretation. Scholarly publishing practices, editing and publishing in general have been shaped by the printing press and later on fashioned by the launch of the scientific journals starting in 1665 with the Philosophical Transactions by the Royal Society. Trends are mounting towards a self-publishing model, let that be under a known editorial label or in a self-sustained environment under university's press umbrella or by the help of online free existing tools and services. All of the practices lend themselves to one purpose: being indexed, being parsed, mined for data, and emerging as valuable contributions to the field and become relevant. This paper explores some the new tool chains the Open Web Platform offers and some alternatives to be considered in the daily editing workflows. The envelope of the existing technologies is pushed further to step into the future world of deep learning and artificial intelligence. An exploratory tab will be set on the possibilities self-publishing is presenting the researchers in terms of tools and publishing platforms other than the *regular* channels.

**Keywords:** publishing; editorial trends; editorial tools; editorial services; editorial business models; open web platform; markdown; publishing platforms; html5; css3; epub.

## Limits and conditions of today editing and publishing framework

There are confinements with regards to what you may achieve with the traditional file formats and software used to edit and disseminate research findings. There is a continuous search for new, more flexible tools and formats able to achieve more

audience, let that be persons or machines. Having flexible tools and better articulated technologies that allow cross-sectorial information to be expressed is a scope of activities extending way beyond computer and information science. The main goal is to achieve a better expressibility through non-sophisticated toolchains which are flexible enough to accommodate all fields of research geared up towards the final goal: reproducibility.

"[…] the global scientific publication output is growing at a rate of approximately 3% annually. The volume of publications doubles approximately every 24 years"[1] reported a study done in 2014 that has analysed figures coming from natural, medical and health sciences. But there is also one aspect important for understanding of the new publishing trends: "nine out of 10 academic papers—which both often take years to research, compile, submit, and get published, and are a major component by which a scholar's output is measured—contribute little to the academic conversation".[2]

Concurrently, members of scholarly community are "being hoodwinked into writing books nobody can buy"[3], and sometimes acceptance comes as a part of satisfying "publish or perish" cycle.

These findings mark a continuous race to get one's contribution to the most visible spot concerning a certain field of research or in a particular learned society. The best visibility is gained through content exposure (intrinsic to document exposure), and this sends the researchers and the managers to devising and adopting the best strategies and tooling suites concerning exposure, visualisation and mining. It was wildly acknowledged that only Open Access, an accomplished policy throughout European Union[4], enabled through Open Licenses is able to give the best chance for the content to be exposed through distribution channel exponential multiplication offered by the automated and semi-automated processes.

Part of the strategies employed with content exposure are closely linked with copyright involving open licensing enabled through wavering certain rights retaining only those needed for acknowledgement. This leads to a readership empowerment which in turn resolves to extending precious findings and most of all enables reproducibility.

---

[1] Bornmann, Lutz, and Ruediger Mutz. "Growth Rates of Modern Science: A Bibliometric Analysis Based on the Number of Publications and Cited References." arXiv.org, Digital libraries, May 8, 2014. doi:arxiv.org/abs/1402.4578v3.

[2] Gordon, Aaron. "Killing Pigs and Weed Maps: The Mostly Unread World of Academic Papers." Pacific Standard, Books & Culture, March 18, 2014.

[3] "Academics Are Being Hoodwinked into Writing Books Nobody Can Buy." The Guardian, September 4, 2015, sec. Academics Anonymous. http://www.theguardian.com/higher-education-network/2015/aug/21/academics-you-need-to-be-managed-its-time-to-accept-that.

[4] "Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020." Directorate-General for Research & Innovation, March 21, 2017. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.

Reproducibility might be an issue if we are to look at the humanities, but even in these fields data would establish, at least a context for further understanding the reasons for a certain result.

The late comer in this concerto is data as a base component, and how to design *Data Management Plans*[5] in such a way that the research efforts would establish extended context needed for higher reproducibility. Some data policies require even to have the data sets accompanying the articles right at the moment of publication[6].

Reign of LaTeX, XML schema based formats, HTML and PDF seems to go well and undisturbed being the majority of digital resources aggregated in the worldwide repositories where the scholarly scientific output usually lands. This is set to be changed profoundly with the continuous rise of the Open Web Platform being "the collection of open (royalty-free) technologies[7] which enables the Web". Using the Open Web Platform, everyone has the right to implement a software component of the Web without requiring any approvals or waiving license fees"[8]. OWP offers a better integrated platform for information and data, let that be metadata, data microformats or canonical datasets based on Comma Separated Values (CSV) or JavaScript Open Notation (JSON).

"One of the biggest bottlenecks in Open Access publishing is typesetting. It shouldn't be" says Rob Welsh, Product Designer and Co-Founder of Scholastica[9]. Besides bottlenecks there is a lot of documented wasted time in peer-review processes working against "disseminating information useful for researchers and society."[10] There is a whole lot of time spent by the researchers in putting together the research results and formatting and often times adapting their own writing behaviours to the editor's publishing conditions. Picking the right tools to write and make sense of the data collected could be seen as an overhead, but in the end these personalised work-flows need another shape shifting in order to respond to publishing rigours.

Let us not forget the fact that most of citations need to be resolved because the vast majority are hyperlinks. No matter the mechanism employed, the citations should be regarded as context understanding enablers for their essential role of setting the piece of information among the sources which led to it.

---

[5] "Guidelines on FAIR Data Management in Horizon 2020." EUROPEAN COMMISSION, Directorate-General for Research & Innovation, n.d. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.

[6] Bloom, Theo. "Data Access for the Open Access Literature: PLOS's Data Policy." Public Library of Science. Accessed September 17, 2015. https://www.plos.org/data-access-for-the-open-access-literature-ploss-data-policy/.

[7] Browser Technologies. http://www.w3.org/wiki/BrowserTechnologies.

[8] Open Web Platform. http://www.w3.org/wiki/Open_Web_Platform.

[9] Scholastica is a journal and publishing platform. http://blog.scholasticahq.com/post/85756650973/one-of-the-biggest-bottlenecks-in-open-access.

[10] Shashok, Karen. "Authors' Editors in the 21st Century: Promoters of Publication Quality and Efficiency." European Science Editing 40, no. 3 (August 2014): 60–62.

Although there are many possibilities to realise this wider context, even to this day simple enabling technologies like Cool URIs needed to establish a true semantic web, are efforts with little impact. “I fully agree with him and others that publishers need to think and act a lot more like web developers”[11].

The technologies used and the final medium should acknowledge that “most of the time, a person sits down at her personal computer not to create, but to *read, observe, study, explore, make cognitive connections, and ultimately come to an understanding*. This person is not seeking to make her mark upon the world, but to rearrange her own neurons. The computer becomes a medium for asking questions, making comparisons, and drawing conclusions—that is, for *learning.”*[12]

**Editing, presentation, interactivity, disseminating, reuse**

The fast pace of research dissemination is guarded by the continuous debate of a needed capturing context for the whole research output starting from data up to publication and presentational phase.

In the same time, there is a need to use simpler file formats both understandable by the machine and human. The classical steps of putting together a finding has translated the final form of having a paper, and slowly migrates to being able of capturing a workflow with all the pieces this open science paradigm entails: datasets, lab notes, blog posts, debates on social media, presentations, software and all the versions of it, the preprint and if accepted a version for print. But all these apparently chaotic steps are the expression of the modern ways in which science is conducted which leads to some interesting controversies linked to the need for fixity[13] in a dynamic web pages world let alone publishing under a widely known label or enter indie row. It is easy to acknowledge the fact all the scientific content ends up on the web, in a repository or on a webpage, be that a dynamically generated by a content management system or a simple version of HTML.

The new publishing open technologies, are currently amassing to realising the long-envisioned goal of having a modular expressive set of instruments able to combine and complement to realize the full potential of research and innovation expressions. All the scientific output should be accounted through the eyes of the machines as pure data if we are to ripe all the benefits. Text and Data Mining considers machines to be first class citizens of the digital world capable of offering a rich context for the humans

[11] McCoy, Bill. “Portable Documents for the Open Web (Part 1).” TOC Tools of Change for Publishing, August 16, 2012. http://toc.oreilly.com/2012/08/portable-documents-for-the-open-web-part-1.html.

[12] Victor, Bret. “Magic Ink: Information Software and the Graphical Interface.” Blog. Bret Victor, September 16, 2015. http://worrydream.com/#!/MagicInk.

[13] Anderson, Kent. “The Mirage of Fixity — Selling an Idea before Understanding the Concept.” *The Scholarly Kitchen*, n.d. https://scholarlykitchen.sspnet.org/2012/01/09/the-mirage-of-fixity/.

to evolve a broader insight many times soon with the help of a neural network upon which deep learning algorithms are processing research inputs, and possibly having it all fed into an AI.[14]

**Overview of the existing building blocks**

There are several technologies that lead to a new upheaval. All of them are linked to the web regarded as a platform rather than a distribution network. Some of them are combined for the sum of parts offers better flexibility like in the case of Web Publications: *A Web Publication (WP) is a collection of one or more constituent resources, organized together in a uniquely identifiable grouping that may be presented using standard Open Web Platform technologies*[15]. This idea is not a new one as Object Reuse and Exchange standards issued by the Open Archives Initiative proposed a similar approach in 2008: "Open Archives Initiative Object Reuse and Exchange (OAI-ORE) defines standards for the description and exchange of aggregations of Web resources"[16]. Although both are similar they share the same traits of the *Architecture of the World Wide Web.*

In order to explore future possibilities, there is a need to understand better the parts because all of them in their own right have a transformative effect upon the others.

**Open Web Platform: HTML5, CSS3, JavaScript and SVG**

HTML is a recommendation[17] of W3C meant to evolve the entire web technologies ecosystem starting with the markup and ending with the different APIs[18]. Together with Cascading Style Sheets 3, JavaScript (an implementation of ECMAscript standard), and Scalable Vector Graphics (SVG) form the Open Web Platform.

One might argue it is only the beginning of the implementation of the new technologies, but all the browser producers are actively supporting it. Also, there are a lot of good editors that would dramatically reduce authoring time in order to arrive to a fine articulated solution. HTML5 technologies act as the enabling context and JavaScript as a glue between the other components of the Open Web Platform. World Wide Web Consortium – www.w3.org – began to show interest to the publishing issues since

[14] McGuire, Hugh. "What Books Can Learn from the Web / What the Web Can Learn from Books." Medium, April 21, 2016. https://medium.com/@hughmcguire/what-books-can-learn-from-the-web-what-the-web-can-learn-from-books-64670171908f.

[15] "Web Publications for the Open Web Platform: Vision and Technical Challenges." W3C - Interest Group, May 2, 2017. https://w3c.github.io/dpub-pwp/.

[16] "ORE User Guide - Primer." Open Archives, October 17, 2008. https://www.openarchives.org/ore/1.0/primer.

[17] HTML5: a vocabulary and associated APIs for HTML and XHTML. http://www.w3.org/TR/2014/REC-html5-20141028/.

[18] "Application Programming Interface." Wikipedia, n.d. https://en.wikipedia.org/wiki/Application_programming_interface.

June 2013, when Digital Publishing Interest Group was initiated (https://www.w3.org/dpub/IG/wiki/Main_Page). The activity of the group is essential for understanding the shape of the digital publishing using OWP in the future. The use of these technologies let to many attempts in finding a formula able to fit the bill of the rich expression palette of the traditional publication world and the innovative and experimental one. Some of them will be reviewed for a better understanding of the different avatars, but first the building blocks.

*HTML5*

The latest iteration of the standard says that *HTML is the World Wide Web's core markup language. Originally, HTML was primarily designed as a language for semantically describing scientific documents*[19]. HTML5 is the context in which many other web technologies are bind like CSS and JavaScript. To the markup structures used to represent the information, there is a programmatic node equivalent for every piece of it called Document Object Model. DOM is the key to realize dynamic and interactive documents because every node is manipulable and mutable.

*CSS3 modules*

When it comes to the first possible useful aggregation using web technologies, the best combination yet is pure HTML5 and CSS3[20]. For the publishing sector, the magic comes from the CSS Paged Media Module Level 3 which allows arrangement of a webpage for print. One possible experimentation for these two main components is PubCSS[21].

*JavaScript – ECMAScript standard*

JavaScript is the programming language used in the browser to manipulate the elements on a HTML page through the Document Object Model. In publishing sector, it has its uses when it comes to interactivity, animation of the SVG nodes and as a general enabler for other technologies coupled with the HTML APIs.

*SVG*

Scalable Vector Graphics (http://www.w3.org/Graphics/SVG/) is another markup language for rendering 2D vector images. The advantage of using SVG is the capability of generating animations. The technology is used as the basic layer for a whole range of libraries aimed at producing illustrations and diagrams.

[19] https://www.w3.org/TR/html51/introduction.html#introduction.
[20] https://www.w3.org/Style/CSS/specs.en.html.
[21] Park, Thomas. "PubCSS: Formatting Academic Publications in HTML & CSS." Personal blog. Thomas Park, n.d. http://thomaspark.co/2015/01/pubcss-formatting-academic-publications-in-html-css.

The technologies presented are only the components for realizing an articulated platform for publishing scientific information using the Open Web Platform.

**Useful additions for the STEM – Science, technology, engineering and mathematics tool chain**

*MathML*

MathML is a markup XML language released in 1999. MathML is part of the HTML5 namespace meaning that you may drop MathML code into HTML code and the user agent will render correctly. If there are some older technologies involved there is a great JavaScript polyfill library called MathJax that will do the proper rendering. In June 2015, MathML 3.0 became ISO/IEC International Standard (ISO/IEC 40314:2015). “The standardization of MathML is key to "friction-free" interoperability among software components for e-learning and technical collaboration in the academic world”. The best place to start learning about it is http://www.w3.org/Math/. For LaTeX users, there are converters hence diminishing the burden of migration.

*MathJax*

MathJax[22] is the product of MathJax Consortium and it is an open source JavaScript platform for high-quality display of mathematical notation across all browsers. Its structure is highly modular to accommodate the use of LaTeX and MathML directly into the HTML source. MathJax is the primary option for the mathematical representation for all major science publishers and some more.

*Chemical Markup Language – CML and ChemDoodle*

CML is an XML schema designed to support the chemical representations and is accepted by most of the publishers. ChemDoodle Web Components is a JavaScript library released by iChemLabs under a GPL license. It is making use of the WebGL and canvas element.

*LaTeX*

“LaTeX is a high-quality typesetting system; it includes features designed for the production of technical and scientific documentation. LaTeX is the *de facto* standard for the communication and publication of scientific documents”[23]. You may take into account LaTeX when you prepare documents. LaTeX is here to stay, being more and more included in the online documents through different plug-in mechanisms. It is not unusual to see fragments of LaTeX in web documents rendered through special parsing scripts.

[22] https://www.mathjax.org/.
[23] LaTeX – A document preparation system. http://www.latex-project.org/.

**The boundary stretched – online and offline boundary blurred**

From the researcher point of view much of the format and afterwards the publication paths are dictated in large amounts by the field or by the sphere of the multidisciplinary road taken.

What binds the practices together is the need to arrive at a version of the document that is acceptable by the publisher. Usually the editorial end of line is a Portable Document Format despite the start and intermediate forms and file formats distilled in the end in the PDF version. Through the use of conversion tools like Pandoc, markdown, R and LaTex files melt into a final version, let that be Microsoft Word (.docx), Open Document Format (.odf), PDF or HTML for the matter.

The Open Web Platform is offering everything in order to present enticing opportunities for text structuring as well as data embedding and on-the-fly visualizations. There is also the capacity of further aggregation and a large cover in terms of standards and fields for descriptive, administrative and technical metadata. Here we will explore some of the technical solutions developed using OWP. First, some of the developments are there to form the main traits of the current trend and the approach of every individual initiative, editorial or experimental leaves a picture rendering the toils of the beginnings.

*HTMLBook*[24]

Is a XHTM5-based standard for authoring print and digital books. HTMLBook is a subset of XHTML5. Many of today O'Reilly books are published this way. The tool employed is called Atlas[25] and it is also available to the general public. The best case to date for this new editing platform is Lea Verou's[26] book *CSS Secrets: Better Solutions to Everyday Web Design Problems.* Even O'Reilly's new editing tools platform makes use of the Git for version control purposes.

*EPUB3 and Web Publications*

EPUB is a standard proposed by the International Digital Publishing Forum, and is intended to be a general purpose document format. It is a combination of XHTML5, SVG 1.1, CSS3, JavaScript and XML to produce a special zipped file.

Out of the many formats available, EPUB is the one striding to become the best for distribution and interchange. The key feature is the use of Web Standards. In January, 2017, the efforts of the IDPF were transferred to the World Wide Web Consortium (W3C). Immediately after, in February, 2017, the efforts of both organisations were

[24] http://oreillymedia.github.io/HTMLBook/.
[25] https://atlas.oreilly.com/.
[26] Invited Expert in the W3C CSS Working Group.

combined into a *New Roadmap for Future of Publishing is Underway as W3C and IDPF Officially Combine*[27]*.* Publishing Working Group established soon after published a Charter with the goal of having "all publications—with all their specificities and traditions—to become first-class entities on the Web"[28].

*EPUB+WEB*

Is a “vision”[29] fostered by the W3C to advance the establishment of a document representation to be used in the context of Open Web Platform. This public working draft has a section dedicated to scholarly publishers and STM publishers. This vision achieved a new embodiment in the form of the Web Publication.

*Web Publications*

Given the web technologies, scholars engaged to find a unifying answer and one notable example is the concept of “book in browser”[30] which was followed independently by an intensified activity giving birth to a new community around the new concept of “web publication” enabled by the Open Web Platform.

*A Web Publication (WP) is a collection of one or more constituent resources, organized together in a uniquely identifiable grouping, and presented using standard Open Web Platform technologies*[31].

The example of Organisation for Economic Co-operation and Development is showing already an integration platform for the Web Publications in the electronic library of OECD accessible at the following link http://www.oecd-ilibrary.org/. Besides the classic PDF, the document is represented as a Web Publication in the form of an HTML version and also as an aggregated form as EPUB.

## The case of lab notes

In case the researchers need to document their laboratory work, there are many tools, but only a handful are also capable to offer dynamic interactions. I have chosen the case of Jupyter Notebooks as it is a popular choice thee days. Jupyter Notebooks (https://jupyter.org/) is introducing yet another open document format based on JSON.

[27] “New Roadmap for Future of Publishing Is Underway as W3C and IDPF Officially Combine.” W3C, February 1, 2017. https://www.w3.org/2017/01/pressrelease-idpf-w3c-combination.html.en.

[28] “Publishing Working Group Charter.” W3C - Publishing Working Group, n.d. https://www.w3.org/2017/04/publ-wg-charter/.

[29] Advancing Portable Documents for the Open Web Platform: EPUB+WEB. White Paper. Unofficial draft. 30 June 2015.

[30] Montana State University Library developed a project aiming to explore the combination of web technologies in search for a unified answer: http://arc.lib.montana.edu/book/.

[31] “Web Publications Use Cases and Requirements.” W3C, May 2, 2017. https://www.w3.org/TR/pwp-ucr/.

It includes a record of the user's sessions code sequences needed, text, mathematical equations and rich output. To render the content, a Jupyter notebook needs a specialized piece of software called kernels. The kernels are in charge with running the embedded fragments of code and the text, all being transmitted to the kernel using a specialized protocol based on JSON. More example could be consulted at the following link https://github.com/jupyter/jupyter/wiki/A-gallery-of-interesting-Jupyter-Notebooks.

**Embedding data**

There is a leap of faith every researcher has to take: regard your text as data, as structured data capable for much more than conveying information. The research output most than any human information and knowledge productions must become intelligent, context-reach and extendible. Structured and unstructured data are the opportunities to realise a true semantic web[32]. HTML5 already allows data embedding through the use of special baked-in attribute: "data-something". But there is a much more to be done and the text could be further endowed with "intelligent" marking through attributes or tagging as it is the case of Text Encoding Initiative. There is more.

Using the existing technologies already presented, there is a path to add structured data to the existing markup and this is possible due to the use of RDFa - Resource Description Framework in Attributes. "RDFa, like Microformats and Microdata, enables us to talk about *things* on the Web such that a machine can understand what we are saying[33]". If only the subset proposed in RDFa Lite finds a way into the documents this minimal step would represent a leap forward. Use of microdata[34] would add further valuable data into the document. For a preview of what could be accomplished http://rdfa.info/play/ should be accessed, and further down this path http://rdfa.info/ should be acknowledged as a valuable portal.

Schema.org is a consistent effort of the web commercial services to create a vocabulary compatible with RDFa, microdata and JSON-LD. This vocabulary is extended in every walk of life, but it is capable of expressing bibliographic[35] relationships and periodical structure as of 2014[36] having a group[37] dedicated to bringing this type of data expressiveness into HTML documents. Going a bit deeper we find RDF – Resource Description Framework which is the "standard model for data interchange on the

[32] http://www.w3.org/standards/semanticweb/.
[33] http://www.w3.org/TR/rdfa-lite/.
[34] http://www.w3.org/TR/microdata/.
[35] http://www.w3.org/community/schemabibex/wiki/Bib.schema.org-1.0.
[36] Wallis, Richard, and Dan Scot. "Schema.org Support for Bibliographic Relationships and Periodicals." Schema Blog, February 9, 2014. http://blog.schema.org/2014/09/schemaorg-support-for-bibliographic_2.html.
[37] https://www.w3.org/community/schemabibex/.

Web[38]". It allows structured or semi-structured data to be mixed. RDF 1.1, the latest version was published in 2014 as a series of W3C (World Wide Web Consortium) Recommendations[39] and Working Group Notes. RDF Schema which is an XML schema provides a "data modelling vocabulary for RDF data.

**One important alternative to the OWP: Markdown**

If there is a growing movement of establishing the right web framework for expressing scientific literature, there is also anther towards simplifying the whole writing and editing experience. The most popular format growing in importance is Markdown. This is a "a plain text formatting syntax"[40] aiming to realize readability in the best possible way without markup excess. At this point there are many flavours of markdown dictated by the needs of their adopters (GitHub[41], reddit, Diaspora, Stack Exchange, OpenStreetMap and Source Forge[42]). Markdown is one of the fastest growing syntax for document "intended for one purpose: to be used as a format for *writing* for the web"[43].

Markdown is eroding the traditional approaches. It is more flexible and taxes very little. There is also a fantastic opportunity and this translates the text written in Markdown to be treated as data having a minimal markup to indicate structure.

An entire ecosystem grew gradually and also editorial businesses start to appear. For getting across to the known file formats, there was a need to develop software behaving like a digital Rosetta Stone. One having a large footprint in the open community is Pandoc (command line). Pandoc is a universal document converter answering the need of transforming one markup format into another. There are even collections of templates[44] designed for different types of documents.

For the scholarly publishing a special Pandoc adaptation is available called Pandoc Scholar[45] which gives the instruments to easily introduce citations because it has integrated CiTO, the Citation Typing Ontology (http://www.sparontologies.net/ontologies/cito/source.html).

[38] http://www.w3.org/RDF/.
[39] http://www.w3.org/standards/techs/rdf#w3c_all.
[40] Markdown. Grubber, John. Daring Fireball. http://daringfireball.net/projects/markdown/.
[41] https://github.com/markdown.
[42] Markdown. https://en.wikipedia.org/wiki/Markdown.
[43] Markdown: Syntax. Grubber, John. Daring Fireball. https://daringfireball.net/projects/markdown/syntax#philosophy.
[44] https://github.com/jgm/pandoc/wiki/User-contributed-templates.
[45] Krewinkel A., Winkler R. (2017) Formatting Open Science: agilely creating multiple document formats for academic manuscripts with Pandoc Scholar. PeerJ Computer Science 3:e112 https://doi.org/10.7717/peerj-cs.112.

### Self-publishing and dynamic dissemination

Signals of new publishing patterns are seen everywhere. Researchers are using their own blogs hosted by themselves or by the research institutions they are affiliated. Besides blogs, there are universities portals having dedicated pages where many researchers choose at least to keep a record of their published results. Some choose to use online publishing services and some keep online a record of their entire work in the form of complex web sites.

### Nanopublications

"A nanopublication is the smallest unit of publishable information: an assertion about anything that can be uniquely identified and attributed to its author[46]". Nanopublications are the works of Concept Web Alliance, a collaborative community *that is actively addressing the challenges associated with the production of unprecedented volumes of academic and professional data[47].*In short as the community behind it states, nanopublications are to be regarded as "core scientific statements with associated context".

### Crowdfunding pattern

Kick-starter offers a resounding example in the field of computer science. On the 1st of May 2013, Kyle Simpson, an Open Web Evangelist from Austin, Texas, asked the community to back his project of writing a series of books named "You Don't Know JS". Not only he got successful with the campaign, but he successfully published the books with O'Reilly Media[48], and the interesting part is that he was allowed to publish the books on Github[49] as well under a Creative Commons license. This move and new approach to publishing led to a continuous improvement of the series, being considered now as one of the best authority sources in JavaScript programming.

### Just put it out there – self-publishing for the commons

One good example comes is *Eloquent JavaScript* by Marijn Haverbeke. This book is licensed under Attribution-Non Commercial 3.0 Unported. The book is published in print with No Starch Press[50]. Besides packing a version of the book as PDF and Epub there is also the online version as HTML5[51].

---

[46] "What Is a Nanopublication." People, Tools and Knowhow for Datapublishing. Accessed September 20, 2015. http://nanopub.org/wordpress/?page_id=65.
[47] http://www.nbic.nl/about-nbic/affiliated-organisations/cwa/introduction/.
[48] http://search.oreilly.com/?q=Kyle+Sympson&x=0&y=0.
[49] https://github.com/getify/You-Dont-Know-JS.
[50] https://www.nostarch.com/ejs2.
[51] http://eloquentjavascript.net/index.html.

Other example comes from Cody Lindley having his books online as HTML5 documents http://www.domenlightenment.com/. The online version are the “pre-edited / draft form” of the O’Reilly edited version (printed, epub, mobi, pdf). The HTML5 files are kept as online versioned variants.

Other example is Python Data Science Handbook by Jake VanderPlas https://github.com/jakevdp/PythonDataScienceHandbook.

With the growing numbers of alternatives for writing and editing scientific contributions, the number of editorial platforms empowering a new bread of writers are growing to sustain these practices. One interesting example is the Leanpub, a platform allowing “to write, publish and sell in-progress and completed books”. The generated output from your contributed plain text is a PDF, an EPUB or a MOBI version of your content. Markdown content is allowed as well as Word files.

O’Reilly Media Inc. has opened its publication workflow to the individual authors and set out a model for self-publishing that is easy using the new formats. The platform is called Atlas and offers all the modern publishing instruments needed for the scholarly publishing. There is some ingredient that is common for any modern publishing system: the tools for versioning the content. O’Reilly is making use of GitHub integration allowing the author to feed into the system the content available in the GitHub account. Atlas is not the only platform that has understood the need for versioning of the content. Atlas content is managed with Git. As many others follow Atlas developers understood that the movement of ideas and changes occurs at a ever growing beet rate to which a publishing, a self-publishing ecosystem must be set.

GitHub is a platform preferred more and more for content creation, not only for writing software. More often is the place where code, data and ideas sum-up is put as a perpetual versioned resource. GitHub is not a platform only to create, but also to curate literature and a vivid example would be Project GITemberg declaring to be “a Free and Open, Collaborative, Trackable and Scriptable digital library”[52].

All those publishing platforms supporting integration with GitHub have understood one essential truth of the current editorial flow: text is data. Data is mutable and information is evolved continuously.

## Discussion

A better adoption of new technologies, formats and models are to be impacted by influencers of the scientific fields. The new scientific rock stars will be those able to cross cut several disciplines in a growing data analysis capacity built by software and hardware advanced skills. How to accelerate the process?! By taking into account the power laws. Let’s identify the hubs for open science and convert them into shift processes with regards on how science knowledge is recorded.

[52] https://www.gitenberg.org/.

Today scholarly communication is no more an effort of aggregation and quantification indices. It is how to expose content, if possible linked content in better ways both for humans and for machines. *As the amount of scholarly communication increases, it is increasingly difficult for specific core scientific statements to be found, connected and curated. Additionally, the redundancy of these statements in multiple fora makes it difficult to determine attribution, quality, and provenance[53].*

A possible trend for publishing is reserved to the mutable document and data if the data is the document. In fact, mutable to a form that will enable others to "see further" through possible re-use, forking and versioning. The is a nuance in the sense that a difference has to be made between the peer-reviewed publication of one author as part of the scientific record which is immutable and the research database or data set that is mutable: "seamless integration between the two is not desirable[54]".

> "The future of building interactive user interfaces should lie not in the hands of programmers, but in the hands of the expert of a given field. [...] Using technology to create increased levels of abstraction for the documentation of science can push education and society forward[55]".

There is also the need to mine the scientific facts already existing in the scientific literature. "The right to read is the right to mine!" is the case of the ContentMine[56], a project designing software[57] instruments and building a community dedicated to the goal of extracting 100.000.000 facts from the scientific literature that will be put in Wikidata.

It is wise as a creator, as a researcher to retain all documents produced alone or in collaboration. The reasons are multiple, but the biggest threat is web services evolution sometimes leading to site-deaths[58] taking off and breaking all the permalinks[59] (links retrieving a single post) in the process. Another reason for self-publishing is linked to ownership of the resources created, which in turn leads to a direct chain of citations. And one of the most valuable gain is having the URLs of the resources in your domain. In time, this resolves to better searchability and better metrics through local analytic.

---

[53] http://www.nbic.nl/about-nbic/affiliated-organisations/cwa/introduction/.
[54] Shotton, David, Katie Portwin, Graham Klyne, and Alistair Miles. "Adventures in Semantic Publishing: Exemplar Semantic Enhancements of a Research Article." Plos Computational Biology, Open Access Collection, April 17, 2009. doi:10.1371/ journal.pcbi.1000361.
[55] Lynch, Dan. The Art of Digital Publishing A Foundation of Combined Standards to Support the Future of Publishing. Accessed September 15, 2015. http://www.mathapedia.com/books/31.
[56] http://contentmine.org/.
[57] https://bitbucket.org/petermr/ami/wiki/Home.
[58] http://indiewebcamp.com/site-deaths.
[59] http://indiewebcamp.com/permalinks.

A perfect scenario would be for a research output to be seamlessly integrated with the others via semantic technologies in such a way that the contribution would be instantly recognizable. That would be the very moment when everyone, no matter the field will be interested to contribute.

Future of publishing is interwoven with the capacity of transforming a “publication” – research output – into a hub, a proxy or a “small world” that is highly connected with others on the semantic levels. There is already a common denominator: HTML5 with all its components and browser APIs. Due to the fragmentation in the market of electronic publishing dictated by the commercial interests, a wide adoption of HTML5 as a universal expression format is lagging behind. All the components for rich expression of scientific ideas regardless of the field are there to be used in any means necessary. In turn, there are efforts aiming towards adapting to new “industry standards” like EPUB3.

The trend is set by the use and embedding of the data related to the content – the case of STM, and in the case of humanities the new semantics of HTML5 markup allows a better expression and a ubiquitous presence, not to mention a better preservation. Most of the drawback are directly linked to the print form, most of the hesitations are pointing the culprit leading to supplementary resources being allocated to fit the paper.

*Studia UBB Digitalia*, Volume 62 (LXII) 2017, June, Issue 1, 57-74

Published Online: 2017-06-30

DOI:10.24193/subbdigitalia.2017.1.04

# Albums in the attic. An investigation of photographic metadata

Vinayak Das Gupta

Maynooth University, Ireland,
E-mail: vinayak.dasgupta@gmail.com

**Abstract:** This article investigates the challenges of creating online research collections, specifically for photographic objects. This examination considers the material aspects of the photograph and considers the manner in which they may be made accessible within a web resource. The article explores the challenges of attributing metadata for visual resources in an attempt to address some of the key issues in the production of resources in the Digital Humanities.

**Keywords:** photographs; images; art history; metadata; material culture; ontology.

In the late nineteen-eighties, my father owned a camera which would imprint the time of the photographic moment on the film. Every print would bear a sequence of faded, red numbers on its edge. Perhaps the camera had a fault, or perhaps my father never bothered to change the batteries, but the imprinted time would invariably be incorrect. Those numbers had their own logic, one that, to this day, I haven't quite managed to understand. But, as a child I would marvel at the manner in which even that incorrect inscription would guide the way in which the photograph was perceived. Those faded red numbers on the bottom-right corner of the photographic paper would enter the image into a series that would potentially enable me to understand the sequence in which the photographs were taken. This was my first encounter with metadata, even if it was one without realisation. It is now common to expect the digital camera to record a host of information from shutter speed to the geo-location of the photograph. These machine-recorded data guide the way in which the image is perceived. Metadata not only shapes the way the object is understood but also aids in

its search and retrieval within digital repositories. The metadata of each individual object affects the experience of the entire collection. Metadata, then, provides a significant way in which descriptions can be provided for objects that are not self-describing.[1] This article contends with, specifically, the possible ways of articulating the material aspects of the photographic image. It is argued that beyond the content image (the indexical quality of the photograph), contextual clues derived from inspections of its physical properties are important for a coherent understanding of the object. Following an examination of the physical nature of the photograph, its presentational forms are inspected; the content image, the material existence, and the presentational forms of the photograph, together, construct a consistent body of knowledge that affords a deeper understanding of the object. An accurate digitisation of the physical object into a digital image (specifically for research purposes), then, must present both the content image and the physicality of the object with sincerity. While current digitisation techniques have found accurate methods of copying the content image, the description of its materiality remains a challenge. If the physicality of the photograph is central to its understanding, this article inspects the possibilities of providing the digital referent with the material information. It presents an examination of both the materiality of the photographic image (and its transformation into a digital object), and the means through which the presentational forms of the original may be inscribed in the digital referent.

## 1. Photographic materiality

While the image imprinted on photographic paper is important, it would be an error to assume that the photograph may only be understood as an image; scholarly study of the photographic image, specifically with the 'material turn' in anthropology and cultural studies,[2] has attempted to inspect the photograph as a physical object. This 'material turn' expresses the complexity of the social existence of objects and allows the investigation of the photograph as a physical object whose understanding is augmented by its form. Photographs have specific modes of circulation, production, and consumption, and their inspection has potential beyond the critiques of representation

[1] These are objects that cannot articulate any information about themselves. Unlike the printed book which reveals the author, the date, the publication house etc. (inherent in the structure of the book), visual objects or pieces of music for instance, cannot reveal any information about their nature. So, they are not 'self-describing.'

[2] Material cultural analysis, from an anthropological position, questions the assumed superiority of language over other forms of expression (Miller 1987: 9). In the context of photography, Elizabeth Edwards (2002: 67) suggests the need to 'break, conceptually, the dominance of the image content and look at the physical attributes of the photograph.'

alone. The inherent bias of the indexical qualities of the image over its material properties may overlook the social and cultural contexts within which the photograph is born and used. '[A]s material objects whose "currency" and "value" arise in certain distinct and historically specific social practices' (Tagg 1988), the stability of photographic meaning remains a descriptive challenge. Arjun Appadurai (1986) explains how objects and the practices within which they are embedded are interwoven, cannot be read independently of each other. He writes (1986: 5), 'things have no meaning apart from those that human transactions, attributions, and motivations endow them with.' Geoffrey Batchen (1997: 6) contends that the meanings of photographic objects are dependent on the context in which they find themselves at any particular instance. He argues that, '[a] photograph can mean one thing in one context and something else entirely in another' and that the identity of the object is contingent on its 'use' in the physical world. Gillian Rose (2010: 18) echoes similar sentiments when she says, 'what people do with photographs is not an optional analytical extra; it is fundamental to exploring photographs' effects in the world.' The examination of photographs, as material objects, appears to present two inter-related concerns: the first discusses the 'plasticity of the image itself, its chemistry, the paper it is printed on, the toning, [and] the resulting surface variations' (Edwards and Hart 2004). The rationale here is that the choices made in the making of photographs are rarely random. As Schwartz (1996: 58) expounds, 'the choice of ambrotype over paper print implies a desire for uniqueness, the use of platinum over silver gelatine intimates an awareness of status, the use of gold toning a desire for permanence'. The second concern is that of the presentational forms — *cartes de visite*, cabinet cards, albums, mounts and frames[3] are intrinsically linked with photographs and have constituted a major consumer market since the twentieth century, especially after the Kodak revolution[4] of the early twentieth century (Edwards and Hart, 2004).

Digital objects (both born-digital and digitised artefacts) present a conceptual challenge: Joanna Sassoon contends that '[b]y the direct conversion of light into a digital format to create a stable image "photographs" that only exist in the digital form can be seen in one context as a truer version of photography (writing with light) than

[3] Photographs were, and still are, used in the production of many curiosities. For instance, see Susan Legêne's (2004) account of playing cards illustrated with photographs (though the production of these were in the first half of the twentieth century).

[4] The technological changes, specifically with the introduction of the Kodak camera, resulted in the device being smaller, portable, and cheaper. This allowed enthusiasts to practice the form and the portability of the camera allowed them to take landscapes and views with relative ease. This, to some extent, accounts for the larger range of photographic subjects in the early twentieth century. See Mia Fineman's 'Kodak and the Rise of Amateur Photography.' *Heilbrunn Timeline of Art History*. New York: The Metropolitan Museum of Art, 2000. Web. 21 November 2013.

those that require the creation of a physical intermediary to view the image in a material form' (2004). The central rationale of Sassoon's argument takes premise on the assumption that digital objects have no material existence. This reveals a fallacy in our inspection of digital objects: the digital artefact, without question, is a material object. To observe this materiality, a separation between content and carrier of the object is required. The photographic image is printed on paper, and this paper is the carrier of the content image. Similarly, the content of the book is carried in the physical form of the book — the paper, glue, and ink that hold it together. In the physical object (as opposed to its digital counterpart) the content and the carrier are closely inter-linked to the point where their separation is difficult. However, the carrier may change at different moments of time, which may provide the object with different contexts: consider an image that was first printed on photographic paper, then printed in a newspaper, and then, perhaps, in a coffee shop book, displayed on a bookshelf, never quite read.[5] The different material existences of the image provide contexts that locate it within different points in history. The digital object is, similarly, carried by electronic circuits — a complex creation of technology that is able to house a large amount of data within a tiny space. That the digital object has materiality, then, is undeniable. The problem, perhaps, in identifying this materiality lies in the degrees of separation between the circuits and the perception of the object. To view an electronic image, a screen — an enabler — is required. The experience of the object then is governed by an intermediary system. The dislocation between the carrier of the object and the experience of the object is perhaps the source of the material fallacy. If it is acceptable to think of this separation between the content and the carrier, it is possible to argue that the digital object is merely one iteration of a different carrier. The digitised image, then, is the original image, in a new material form. However, it must also be recognised, that the materiality of the digitised object does not mirror that of the original photograph. Within our current technological limitations, the only possible method of attributing the material specifics of the original photograph to the digital referent is through textual means. The metadata of the digital object relates the materiality of the original. The process of digitisation focuses on the content image, converting the original image into a series of pixels,[6] while the carrier of the photograph is described through

[5] The circulation of the photograph moving between carriers and put into new uses, derive new contexts. Brent Harris (1998), for instance, in his investigation of photographs from Africa (1850-1950) provides a strong argument for examining the circulation of the photograph. His examination reveals the contested meaning at various stages of the photograph's production, circulation, and consumption.

[6] The process of digitisation, from a computational perspective, is quite fascinating. A researched description, specifically investigating the process, may be found in Jähne, 1995.

the metadata of the object. Sassoon's investigation of the process of digitising photographs addresses some major concerns. She writes (2004):

> [I]t is no longer an accepted canon that a photograph is merely a print on paper, nor is it a simple and uncomplicated translation between reality and its mechanical representation. […] the digitising process can no longer be seen as merely changing the physical state of a photograph from the material to the pixel. If a photograph can be seen as a more complex object than simply an image, digitising can be seen as more than simply a transformation of state, or a transliteration of tones. The process of digitising involves a more complex cultural process of translation — or a change between forms of representation.

What Sassoon identifies in this transformation is the complexity that lies in articulating details about both the content image and the carrier, even if this distinction is not explicitly made. The metaphor of translation, or even transliteration, that she uses, is perhaps not entirely convincing; the transformation from physical to digital is a re-configuration of state, one that is consistent with any other similar transformation — from the shellac disc to the audio tape or from celluloid to the video cassette. Translation, even if the word is used loosely in Sassoon's work, indicates an inherent change in content *and* carrier. As discussed, digitisation (or effective digitisation) is focused on preserving an accurate content image, even if an accurate representation of materiality is beyond its purview.[7] What the word 'translation' does point towards is, in fact, the tension between original and the copy, in the accurate representation of the source. What she stresses on repeatedly in her essay is the importance of cultural contexts within which these material objects reside. Any transformation of state, whether physical or digital, creates a new context for the object; the digital state has its own context, while our attempt is to record the context of the source within the digitised.

All digitisation activities are carried out for some specific purpose. The purpose might be merely to record the content image for the purposes of the collection. As we have already discussed, the desire of the collector is central to the affect of the collection. For research collections, however, the need to record more than the content seems to be necessary. The focus of digitisation, then, cannot be only on a singular aspect of the photograph — the indexical quality of the image. The relationship

[7] Beyond the content image of the photograph, its physical peculiarities are also recorded. The back of the photograph, any tears or marks, and any inscriptions are also captured with the indexical image (this depends, though, on institutional policy or individual motivations). The tactile response of the photograph or its smell are more difficult to record. They may be described, but an inevitable bias is very likely.

between the negative and the paper forms of the photograph may reveal visible clues about its technical origins. The physical dimensions of the photograph may reveal details about the kind of camera used, the negative size, and a possible date of production. The tonal ranges, and the photographic texture may indicate the type of process used to develop the image. All these point to the social existence of the object. Viewed on computer screens of different sizes and different calibrations, it is easy to lose sense of the proportions and tonal hues of the original image.[8] Digitisation effectively alters the way the photograph is viewed; using a loupe to magnify the details of the content image, we are far readily aware of the material aspects of the photograph than that of the digital referent. The use of keyboards and mice might allow new interactions, and the ability to magnify to extreme points, but they fail to offer the tactile response of the physical object. The fundamental difference in viewing the digital object, brought about through mediation (the computer screen) might lead to questions about the perception of the object. Beyond the material peculiarities, the imprint of time on the object — the dirt and the tear, the slow fading of chemicals on paper — is evidence of the lives of the photograph, about the provenance of the image.[9] Paratextual material such as captions, scribbles on the back of the photograph, hand-colouring of black and white images, are invaluable to its contextual understanding. While research on advanced techniques of image representations point to the future of digitisation, it is rarely available for average consumer use.[10] How are we to record these elements — the material, the contextual, and the paratextual aspects of the image — with the range of technology afforded to us?

[8] Enablers, for instance computer monitors, present their own problems. The calibration of the screen (brightness, colour fidelity) along with its physical dimensions change the perception of the digitised object. In most digital collections, the digital referent does not possess a reference point against which it can be compared to derive an understanding of the original. Cultural institutions follow strict rules when digitising objects; unfortunately, they have no control over the enabler and the perception of object, for the viewer, may be fundamentally different from what its natural form is.

[9] Roland Barthes, in his seminal work titled *Camera Lucida: Reflections on Photography,* begins by describing the the material aspects of the image. He writes, 'The photograph was very old. The corners were blunted from having been pasted into an album, the sepia print had faded, and the picture just managed to show two children standing together at the end of a wooden bridge in a glassed-in conservatory, what was called a Winter Garden back then' (1981: 67). The imprint of time on the photograph is clear from his description; the material contextualises the index.

[10] There is a lot of research on object recognition in images. Vinyal et al.'s (2015) paper shows how complex images are provided a dynamically-generated caption. The work uses computer vision and natural language processing to form a complete image description. Their real-world applications haven't reached the average consumer. Also, these are only able to articulate what is in the photographic frame.

## 2. The Photograph and its Presentational Forms

The photograph has multiple lives; it exists within socio-cultural contexts, and to understand it, the content image must be seen in conjunction with its material form. Since the inception of photography, photographic images have been used in a variety of contexts, and have been presented in a multitude of ways; the carrier often determines the use the content image is put to. Whether preserved in photo-albums (arranged thematically or sequentially), sold as postcards for the curious, or published as documentary evidence the presentational form of the photograph weighs heavy on the readings of the image. Presentational forms, in particular, guide the way in which photographs were used after their inception, and also the way they were understood. It is, here, important to distinguish between the carrier and the presentational form: the carrier is always material, while the presentational form is ideological. The materiality guides the technical production of the image, and bears the imprint of time on it. The presentational form reveals the social, political, cultural, and religious contexts within which the photograph is *used*. The ideological and the material may exist in the same form.

To elaborate on the affect of presentational forms, and to see more clearly the separation between the carrier and the presentational form, it may be prudent to inspect the photographic album. The photographic paper is attached to the album; the album is the presentational form that is imbued with ideology. Photo albums are unique cultural artefacts in the sense that they look beyond the content and the creator of the photographs to the creator of the object in which the photographs reside. Glenn Willumson (2004), in an essay on the displaced materiality of photographic objects writes:

> The performance of thumbing through the photographs, selecting and sequencing, and gluing them into an album breaks the bond of the materiality of the photograph from its links in commerce and mass production. In choosing, sequencing, organising and captioning the photographs for the album, the person responsible transforms the meaning of selected images into an intensely individualistic expression. At the moment of creation, the photo album is a personal artefact, a record of people and events that are rich with biography and personal memory.

Photographic albums are embedded with the traces of their owners and their practices. The family album, for example, is immersed in ‘mnemonic frameworks’ (Langford 2001). It is argued that while textual communication (captions, titles etc.) may aid the reading of photographs, oral communication plays an important role in the experience of family albums. The oral stories surrounding the family albums are ‘performatively intertwined’ with the photographs (Edwards, 2009: 39) and thus are

construed within the etiquettes of social interactions. From this, it can be observed that presentational forms bring photographs into use, immersing them within social practices and cultural contexts. The travel album — a form so ubiquitous in the nineteenth century that they are sometimes invisible' (Nordstrom, 2004) — exist as a presentational form that provides layered subjects for examination. They bear testimony to the use, and consequently the context, of the photographic images. Though the examples in this thesis are mostly derived from South Asian contexts, the nature of photography and photographic presentation is remarkably similar is other spatial contexts. Nordstrom's study of the Tupper scrapbooks, specifically the material aspects of it, highlights the distinction between the photographs and the album as an object. Made by hand, the albums are visually unremarkable in comparison to other surviving examples from the time. What is remarkable, however, is the precision and detail with which extended captions have been afforded to each photograph. The presentational form has its own history, its own purpose, its own intent: the way a series of photographs are collected and presented in an album can relate a tale on its own accord. At this point, an important question arises: how do we record the material information within the existing structures that digital technology offer us? To coherently articulate the context and the history of the photograph, we must, first, be able to create the distinction between the indexical and the material at the level of its description. The challenge here is not merely to describe the photograph, but to delineate the separation in a form that is comprehensible to the computer; this separation must happen at the level of the metadata of the photograph.

### 3. Metadata of the Digitised Photograph

Recent writings on metadata have focused on the more functional aspects of its creation and usage. While examining learning objects and e-print communities of practice, Barton, Currier, and Hey (2004: 5-20) point out the lack of formal investigation of the metadata creation process. While some collection professionals, specifically those working in libraries have written about descriptive practices (Caplan 2003, Haynes 2004, Cole and Foulonneau 2007, Taylor and Joudrey 2008), metadata content standards (Roe 2005, Hillman 2015, Baca 2006) and the use of metadata in digital projects (Kenney and Rieger 2000, Stielow 2003, Hughes 2004) the literature has not yet developed to the point where it affords a complete picture (Park and Tosaka, 2013: 110). A number of existing surveys attempt to provide an overview of metadata practices (Ma 2007, Smith-Yoshimura 2007, Park and Tosaka 2013). Surveys and analyses of the work done by cataloguing and metadata professionals, while very important work in terms of understanding the usage of metadata across institutions, betrays the problems of such approaches. As Ma (2007) points out, the limited number of responses (in comparison to the vast amount of digitised resources), and sometimes,

the veracity of the responses prove problematic in gaining a firm perspective on the issue. The attribution of quality metadata is a process that requires significant time and effort. The results of these surveys are, usually, not very surprising: In Ma's study of the use of metadata schemata, the overwhelming use of Machine-Readable Cataloguing (MARC), a favourite in libraries, followed by Encoded Archival Description (EAD) is what can be expected considering digitised verbal resources far out-number non-verbal ones. This is confirmed by Smith-Yoshimura (2007: 27-29) and Park and Tosaka (2013: 111) who present remarkably similar results regarding the use of metadata schemata. For digitised resources on the web, Unqualified Dublin Core (DC) is the most popular. If the previous studies had only considered libraries in the United States of America, Palmer, Zavalina, and Mustafoff's study (2007) attempts to widen the research lens to include both research and non-research institutions. They note an overwhelming presence of Dublin Core, with almost half or more of the digital collections using it alone or in combination with other schemata. The criteria in selecting specific metadata schemata are derived from collection-specific considerations of the type of resources, the nature of the collections, and the needs of primary users and communities. Existing technological infrastructure, encompassing digital collection or asset management software, archival management software, institutional repository software, integrated library systems, and union catalogues also greatly affect the selection process. The skills and the knowledge of metadata professionals and the expertise of staff also are significant factors in understanding current practices in the use of metadata schemata and controlled vocabularies for subject access across digital repositories and collections (Ma 2007: 21-28; Park and Tosaka 2013: 113).

Surveys carried out mostly within archival institutions and digital libraries do not contend with the very large number of digital collections produced by individuals who are untrained in preservation methods. Melissa Terras (2010) refers to the growing trend in the creation of amateur online museums, archives, and collections, as an example of how individual endeavour may influence traditional memory institutions in creating useful, interesting repositories. These collections, perhaps not always built along the best practices guidelines, still provide a valuable resource when inspecting our cultural heritage.[11] The digitisation of private collections are increasingly becoming more popular. The experience of metadata lies not with the creator, but with the user of the digital resource. Considerations on metadata usage, then, needs also to be made from the perspective of the user who may not be well versed with the protocols

[11] With the advancement of technology, it has become easy to commit analog material to the digital space — specifically private collections. These, when committed to the digital space, might not necessarily adhere to the guidelines set by institutions, but provide resources that can be explored for both scholarly and private pursuits.

and vocabularies practiced within the institutions. Kathleen Fear's study, an excellent investigation of metadata from the user's perspective, discusses the value of Dublin Core in digitised image collections, in an attempt to understand the use of metadata. The study, conducted through surveys, focus groups, and search and usability testing, tries to identify the nature of information that non-expert users rely on when searching for images and to locate the vocabularies that best express that information. The study of metadata then has to contend not only with the semantics of expression (with regards to interoperability), but also with a certain lucidity and familiarity that may engage the user (Fear, 2010).

Visual resources are inherently different from textual ones. Their needs, the modes of their description, and the way they need to be conceptualised require a different standard to the long established (and very popular) MARC format. To describe the complexity of visual resources Categories for the Description of Works of Art (CDWA) affords a comprehensive guideline and framework. It has 532 categories and subcategories[12] to include every possible aspect of a visual object. The exhaustive nature of CDWA requires a careful identification of what needs to be described. In other words, collections built using CDWA display the intent of the curator at the very outset. The inherent difference between visual objects and printed verbal resources is that the data related to visual objects, like photographs, is notoriously difficult to articulate and often contradictory. A work of art, like a painting, might have been attributed to one painter at a certain point in time, and another at a later date; it may have multiple titles or multiple dates assigned to it. CDWA, built specifically for art objects, was designed to accommodate these issues (Baca 2001: 3).

The curation of historical photographs proves a difficult challenge in this respect. In a pragmatic or even manageable sense, one might think of the metadata sheet as the space for setting out the physical and verifiable properties of a digital file. This is certainly so; but the lack of accompanying information about photographs acquired from private collections continually presents difficulties. One, particularly, is with the date of recordings: even if some collectors were assiduous providing details of dates and places, others were content with the slightest of notings on the physical carriers. Of course, in their original location, it was almost invariably the collector's memory that would supply a whole range of details, but in the passage from collection to digital repository, the lack of recorded information results in irretrievable gaps. The problem is compounded when we consider that some of these markings on physical carriers might have been misremembered or misattributed: questions on the veracity

[12] For the enthusiast (as opposed to the expert) there exists a CDWA Lite schema with 19 categories (there are further sub-categories for these) to describe a work of art ('CDWA Lite: Specification for an XML Schema for Contributing Records via the OAI Harvesting Protocol').

of information stored within metadata is substantial. Knowledge of the object, then, becomes paramount in our attempts to create this envelope of tags and markers that we call metadata. The process of curation also has to grapple with the circulation of the historical photograph. The same photograph may appear as an image in the newspaper, or it may also be used as the symbol of a political movement, or, perhaps, it is displayed on the walls of an art gallery. The social life of the historical photograph, then, is determined by the circulation of the image. The use of the photograph — its ideological presentational form — determines its context, and in turn renders new layers of meaning to the same image. How are we, as curators of digital collections, to gain an understanding of the object if its nature is continuously changing? The little that we can articulate about the photographic image is derived from an understanding of the history of the object. How we proceed to classify the image, place it within numerous other photographs is dependent on how we recognise the photographer, the period, the photographic plate, photographic process and the photographic context. The history of the object is vital in our attempts to place it within the structures of a digital archive. A problem that is persistent in issues of digital curation of photographs is related to the provenance of the caption. For the archival image, the caption may be attributed by the curator at the time it is committed to the digital archive or it may have been inserted at some intermediary point in its history. This intermediary caption is of archival value as it may locate the image within the larger discourse of its history; the provenance of the caption determines its archival significance. If the caption, then, is of archival value does it merit a separate level of curation?

The manner in which we write the metadata — the vocabulary — conveys our knowledge of the object as series of facts: the name of the photographer, the date when the photograph was taken etc. The tradition of equating knowledge with facts has exists from a philosophical and scientific perspective that can be dated as far back as Aristotle. This view was augmented through the renaissance and enlightenment in order to systemise knowledge. The expression of knowledge becomes fundamental in our attempts to understand and locate the digital object. Traditionally, the efforts to represent knowledge were largely seen as an attempt to manage collections of facts relating to the physical world. The contemporary interest in ontologies can be seen to originate within this tradition and can be taken as an extension of this monolithic view of knowledge. This view on knowledge has been argued over the centuries: Bacon and Locke can be seen to consider knowledge as a single system of beliefs to which new concepts are added. This view would be challenged by Quine who would consider knowledge to be like a ‘field of force’, which impinged on experience only along the edges (Quine 1963: 42). However, what can be agreed on, is that a body of formally represented knowledge is based on a conceptualisation: the objects, concepts, and other entities that are assumed to exist in some area of interest and the relationships that hold among them (Genesereth & Nilsson 1987: 9). A conceptualisation is an

abstract, simplified view of the world that we wish to represent for some purpose. Every knowledge base, knowledge-based system, or knowledge-level agent is committed to some conceptualisation, explicitly or implicitly. We use common ontologies to describe ontological commitments for a set of agents so that they can communicate about a domain of discourse without necessarily operating on a globally-shared theory. Pragmatically, a common ontology defines the vocabulary with which queries and assertions are exchanged among agents. Ontological commitments are agreements to use the shared vocabulary in a coherent and consistent manner. The agents sharing a vocabulary need not share a knowledge base; each knows things the other does not, and an agent that commits to an ontology is not required to answer all queries that can be formulated in the shared vocabulary. Thus, we can assume that 'concepts' are the key building blocks and that we manipulate these concepts with words. Ontologies are dependent on human language to represent the world. It is here that we face the first and perhaps the most significant challenge in order to achieve a shared understanding of the humanities.

With reference to ontologies, at the very outset, we are faced with two distinct issues – a problem of metaphysics and a problem of semiotics. The philosophical investigation of ontology seeks to find the necessary building blocks of the world, their properties and their inter-relationships. A starting point could be found in Brentano's notion of intentionality and 'objects of consciousness' (Brentano 1973: 127-128). An ontology must make clear what the nature, necessary conditions, and properties of these objects could be. Formal ontologies combine this goal with a use of logic that is intended to ensure rigour and axiomatizability of postulated results.

The general programme of ontology relies on it being possible to uncover properties that could not fail to be as they are for the world to be as it is. Existing ontologies have been concerned with the organisation and the structuring of human knowledge of reality rather than with reality itself. However, to engage with an ontology at a level deeper than this – with specific focus on the conceptual framework – it needs to be epistemologically adequate. Some form of accepted constraints on modelling decisions agreement over conceptual ontology construction is required. The main issue with creating these constraints is, of course, in defining the required ontological level. Since this level has to include accounts of basic objects and basic relations independently of our knowledge of them, it is necessary for the account to define how such objects and relations may be put together in order to reveal an understanding of the world. As argued by Heidegger (Heidegger, 1962 1927), and later by Schutz (1966: 82) Wittgenstein (Wittgenstein, 1953), and others, the world of human being is essentially committed and inter-subjective. That is, the world which human beings have access to is already organised ontologically in inter-subjective terms of human interest. Creating a committed view of the world from a 'God's eye-view' neutral perspective of necessity appears to be extremely difficult.

The semiotic problem (Bateman 1993: 5) is derived from a non-theoretical understanding of language that hinders an appropriate construction of ontologies. The underlying conception of language is that it places an emphasis on the world as a source of its decisions concerning ontology construction without a prior analysis of what is meant by the world. It compounds the problem by driving attention away from natural language as it is inadequate and restricted. The relationship between a sign and its meaning is only arbitrary for the most trivial of possible sign-types – that between linguistic form and phonetic substance. A semiotically richer view can capture the fact that more complex signs are strongly and non-arbitrarily related to their social purpose (Hodge and Kress 1988: 82).

With these considerations, we can posit that ontology or knowledge representation is a surrogate standing for the objects and relations outside in the world. The fidelity of the representation depends on what the ontology captures from the real thing and what it omits. Perfect fidelity is impossible. A simplistic view would say that an ontology is a model of the world which can be used to reason about it. One of the major claims made in favour of ontologies is that can facilitate the interchange of knowledge between agents, or the reuse in different systems. However, if each ontology is modelled around an imperfect universe, knowledge sharing would increase or compound errors which were not visible in the initial use of the ontology. Again, an ontology is a set of ontological commitments. The choice of ontology is also a ‘decision about how and what to see in the world’ (Davis et al., 1993). This is unavoidable when we consider that representations are imperfect; however, at the same time, the purpose-built ontology has its advantages as it focuses on what is relevant or interesting within the boundaries of the domain. These choices allow us to cope with the overwhelming complexity and detail of the world. Consequently, the content of the representation provides a particular perspective on the world. The way a knowledge representation is conceived reflects a particular insight or understanding in human reasoning. The selection of any of the available representational technologies commits one to the fundamental views on the nature of intelligent reasoning and consequently very different goals and definitions of successes. An ontology must allow for computational processing, and consequently issues of computational efficiency will inevitably arise. Since all ontologies depend on a propositional view of knowledge in order to begin to be computationally tractable, already a very restricted view of what it is possible to represent has arisen. All forms of knowledge representation including ontologies are both media of expression for human beings and ways for us to communicate with machines in order to tell them about the world.

The criticism levelled at ontologies focuses on the fact that they are unsuited to the world of applications once they get beyond a certain level of complexity. While some ontologies are acceptable there is always a trade-off between expressivity,

usability and accuracy. Further arguments can be made (on a more pragmatic level) about the difficulty of maintaining ontologies and reify a particular point of view of the domain knowledge. Ontologies can be seen to be struggling to keep pace with the dynamic, complex world of knowledge bodies and knowledge-sharing. One of the most basic issues facing the users and developers of ontologies is its degree of complexity. Folksonomies are comparatively easier to use and maintain while offering a flexible and personalised perspective; however, their use is limited due to two reasons – (a) their quality of concepts involved does not match that of ontologies and (b) their reliability cannot be compared to that of an ontology. On the other hand, formal ontologies, such as the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE), the General Formal Ontology (GFO), the Web Ontology Language (OWL), or the Resource Description Framework (RDF), require specialised knowledge to build and use them, and are more challenging to maintain. They are also more rigid than the ubiquitous folksonomies and thesauri, and less adaptable to changing applications and user perspectives (Brewster et al. 2007: 563-568).

The realisation of the obvious technical challenge of creating both human readable and machine readable data is heightened when we consider that over the last few years there has been a steady increase in localised digitisation projects by individuals often unaffiliated to archival and preservation institutions. The digitisation and curation of collections outside the walls of the archival institution bring with it its own set of challenges. Projects carried out without specific funding and done with consumer-level instruments, are often discouraged by the archival community. Katrina Dean (2014: 172), in a defence of traditional archival practices, contends that the practical problems of creating adequate descriptives for digital objects for the purposes of creating digital repositories are intensified by the digital economy. She argues that 'shifting content from public domains to commercial ownership, either through the domination of commercially generated content or the commercial ownership of and exploitation of third party and user-generated content' reflects poorly on the value of information and heightens its fragility. The act of publishing archival documents online, thus bringing them out of a controlled and moderated public sphere and into a digital economy brings with it issues of intellectual property, and consequently ownership of cultural property in general. The level of descriptive, technical, and administrative metadata required to manage, preserve, and discover digitised collections generally exceeds the level of metadata required to manage and make accessible their physical counterparts. Increased standards of explicit evidential value, and compliance required to bring archival collections online merely reflect on the inadequacies of meeting those standards in the traditional archive, in the first place. Traditional archives, for Dean, are about relationships; for their evidence and informational value to be fully explored, the objects must reveal relationships between contexts and records, and

among sources. While in the traditional archive, these relationships were implicit in the architecture of the buildings, in the storage configurations, within the old registers and the administrative files, and in the curatorial knowledge, its digital counterpart is still searching for those connections. 'Short of digitising whole collections and transposing these contexts into metadata, it seems unlikely that collections in their present configurations will be transmitted into the future knowledge economy' (172). While Dean speaks about digital repositories created by institutions and by individuals in the same breadth, there is a difference: the value of archival objects digital and curated to preservation standards lies in the expectation that they would stand the test of time. Projects done in individual capacities, mostly out of interest or curiosity, have a different approach and a different value. These projects accumulate private collections, objects that would otherwise have not entered the archival institutions. Most of these are freely available and resources that may be explored for research purposes. These collections often do not meet the required standards of preservation, or the quality that an institution would demand: their value lies singularly in their availability.

The creation of digital collections requires more than the availability of scanners, cameras, and a knowledge of metadata schemata. The primary problem of creating digital resources, specifically with historical artefacts, lies in the process of their curation. Beyond the technical challenges of metadata attribution, curatorial expertise is required to foster some understanding of the digital object. This is faced severely in the case of historical objects, specifically non-textual objects, which require a degree of curatorial more expertise to interpret. The ubiquitous nature of digital objects has made metadata a primary concern for all engaged in work in the digital space. To aid identification of the object, the machine requires textual markers; this become more necessary for non-textual objects where the machine cannot adequately look through the content of the object. The framework of metadata is vital to the creation of digital repositories. Beyond search and retrieval, metadata guides the conceptual understanding of an object within the collection.

## Bibliography

Appadurai, Arjun. *The Social Life of Things*. Cambridge: Cambridge University Press, 1986. Print.

Baca, Murtha, et al. *Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images*. Chicago: American Library Association, 2006. Print.

Baca, Murtha. 'A Picture Is Worth a Thousand Words: Metadata for Art Objects and Their Visual Surrogates.' *Cataloging the Web: Metadata, AACR, and Marc21.* Ed. Wayne Jones, Judith R. Ahronheim, and Josephine Crawford. 2001.

Barthes, Roland. *Camera Lucida: Reflections on Photography*. New York: Farrar, Straus, and Giroux. 1981. Print.

Batchen, G. *Photography's Objects*. Albuquerque: University of New Mexico Art Museum, 1997. Print.

Bateman, J. A. 'Ontology Construction and Natural Language.' *Proceedings of the International Workshop on Formal Ontology*. 1993. May 2012. PDF file.

Brentano, F. *Psychology from an Empirical Standpoint*. Trans. A. C. Rancurello, D. B. Terrell and L.L. McAlister. London: Routledge & Kegan Paul. 1973. Print.

Brewster, C. and K. O'Hara (2007). 'Knowledge representation with ontologies: Present challenges—Future possibilities', International Journal of Human-Computer Studies 65, Elsevier. 4 May 2012. PDF file.

Caplan, Priscilla. *Metadata Fundamentals for All Librarians*. Chicago: American Library Association. 2003. Print.

Carole Palmer, Oksana Zavalina, and Megan Mustafoff, 'Trends in Metadata Practices: A Longitudinal Study of Collection Federation.' *Seventh ACM/ IEES-CS Joint Conference on Digital Libraries, Vancouver, British Columbia, Canada, June 18–23, 2007.* http://hdl.handle.net/2142/8984. Web. Jan. 24, 2015.

Cole, Timothy W. and Muriel Foulonneau. *Using the Open Archives Initiative Protocol for Metadata Harvesting. Westport*, CT: Libraries Unlimited, 2007. Web. 21 February 2014.

Davis, R., H. Shrobe, and P. Szolovits. 'What is a knowledge Representation'. *AI Magazine* 14.1. 1993.

Dean, Katrina. 'Digitizing the Modern Archive.' *Archives and Manuscripts*. 2014. PDF file.

Edwards, Elizabeth and Janice Hart. 'Photographs as Objects'. *Photographs Objects Histories*. Ed. Elizabeth Edwards and Janice Hart. London and New York: Routledge, 2004. Kindle file.

Edwards, Elizabeth. 'Material beings: objecthood and ethnographic photographs.' *Visual Studies* 17.1: (2002) 67-75. PDF file.

Fear, Kathleen. 'User Understanding of Metadata in Digital Image Collections: Or, What Exactly Do You Mean by 'Coverage'?.' *The American Archivist* 26:(2010). JSTOR Journals. Web. 24 Jan. 2015.

Fineman, Mia. 'Kodak and the Rise of Amateur Photography.' *Heilbrunn Timeline of Art History*. New York: The Metropolitan Museum of Art, 2000. Web. 21 November 2013.

Genesereth, M. R. and N. Nilsson. Mateo, CA: Morgan Kaufmann Publishers: San. 1987. Print.

Harris, Brent. 'Photography in Colonial Discourse: The Making of "the Other" in Southern Africa.' *The Colonising Camera: Photographs in the Making of Namibian History*, 20-24. Ed. Wolfram Hartmann, Jeremy Silvester, and Patricia Hayes. South Africa: University of Cape Town Press Ltd., 1998.

Haynes, David. Metadata for Information Management and Retrieval. London: Library Association Publishing, 2004.

Heidegger, M. *Being and Time.* Trans. J. Macquarrie and E. Robinson, Oxford: Basil Blackwell. 1962.

Hillmann, Diane I. *Using Dublin Core. Dublin Core.* Web. 20 September 2015.

Hodge, R. and G. Kress *Social Semiotics.* Cambridge: Polity Press. 2004. Print.

Hughes, Lorna M. D*igitizing Collections: Strategic Issues for the Information Manager*. London: Facet, 2004. Print.

Jähne, Berndt. 'Pixels*.' Digital Image Processing: Concepts, Algorithms, and Scientific Applications.* 77-99. Berlin, Heidelberg: Springer. 1995.

Jane Barton, Sarah Currier, and Jessie M. N. Hey, 'Building Quality Assurance into Metadata Creation: An Analysis Based on the Learning Objects and E-Prints Communities of Practice' *Dublin Core Conference: Supporting Communities of Discourse and Practice—Metadata Research & Applications.* Seattle, Washington: eprints in library and information science. 2003*.*

Kenney, Anne R., and Oya Y. Rieger. *Moving Theory into Practice: Digital Imaging for Libraries and Archives*. Mountain View, CA: Research Libraries Group, 2000. Print.

Langford, Martha. *Suspended Conversations: The Afterlife of Memory in Photographic Albums*. Montreal: McGill University Press. 2001.

Legêne, Susan. 'Photographic playing cards and the colonial metaphor. *Photographs Objects Histories*. Ed. Elizabeth Edwards and Janice Hart. London and New York: Routledge, 2004. Kindle file.

Ma, Jin. 'Metadata.' *SPEC Kit 298*. Washington, DC: Association of Research Libraries, 2007. Print.

Miller, Daniel. *Material Culture and Mass Consumption*. Oxford: Basil Blackwell, 1987. Print.

Nordstrom, Alison. 'Making a Journey: The Tupper Scrapbooks and the Travel they Describe.' *Photographs Objects Histories.* Ed. Elizabeth Edwards and Janice Hart. London and New York: Routledge, 2004. Kindle file.

Park, Jung-ran, and Yuji Tosaka. 'Metadata creation practices in digital repositories and collections: Schemata, selection criteria, and interoperability.' *Information Technology and Libraries* 29.3 (2013): 104-116.

Roe, Kathleen D. *Arranging and Describing Archives and Manuscripts*. Chicago: Society of American Archivists, 2005. Print.

Rose, Gillian. *Doing Family Photography: The Domestic, The Public and The Politics of Sentiment*. Kent, UK: Ashgate. 2010. Print.

Sassoon, Joanna. 'Photographic materiality in the age of digital reproduction.' *Photographs Objects Histories.* Ed. Elizabeth Edwards and Janice Hart. London and New York: Routledge, 2004. Kindle file.

Schutz, A. 'The problem of intersubjectivity in Husserl'. *Collected Papers III: Studies in Phenomenological Philosophy*. Ed. Ilse Schutz. The Hague: Nijhoff. 1966. Print.

Schwartz, Joanne M. ‘"We make our tools and our tools make us': lessons for photographs from the practice, politics and poetics of diplomatics.’ *Archivaria*, 40: (1996). 40-74. Print.

Smith-Yoshimura, Karen. Rlg Programs Descriptive Metadata Practices Survey Results. Dublin, Ohio: OCLC Programs and Research, 2007. Internet resource.

Stielow, Frederick J. *Building Digital Archives, Descriptions, and Displays: A How-To-Do-It Manual for Librarians and Archivists*. New York: Neal-Schuman, 2003.

Tagg, John. *The Burden of Representation: Essays on Photographies and Histories.* Minneapolis: University of Minnesota Press. 1988. Print.

Taylor, Arlene G., with Daniel N. Joudrey. *The Organization of Information.*Westport, CT: Libraries Unlimited, 2008. Print.

Terras, M. 'Digital curiosities: resource creation via amateur digitization'. *Literary and Linguistic Computing* 25.4: (2010). 425-438.

Vinyals, Oriol, Alexander Toshev, Samy Bengio, and Dumitru Erhan. ‘Show and Tell: A Neural Image Caption Generator.’ *arXiv*: 1411.4555: (2014). Web. 12 June 2015.

Willumson, Glenn. ‘Alfred Hart: Photographer of the Central Pacific Railroad.’ *History of Photography.* Vol. 12, No. 1, 61-75. 1998.

Wittgenstein, L. *Philosophical Investigations*. Trans. G.E.M. Anscombe. Oxford: Basil Blackwell. 1953.

*Studia UBB Digitalia*, Volume 62 (LXII) 2017, June, Issue 1, 75-82

Published Online: 2017-06-30

DOI:10.24193/subbdigitalia.2017.1.05

# Media archaeology of institutional archives?

Gábor Palkó

Petőfi Literary Museum, Budapest, Hungary,
E-mail: luhpeg@gmail.com

**Abstract:** In his book entitled *What is media archaeology* Jussi Parikka suggests that it is possible to outline and develop a 'hardcore' media archaeology not only of the hard drive and software but their official use as a database, as information management systems as well. But is it possible at all to give a media archaeological account of the formal-institutional digital media practices, to provide a deep and material media archaeology of, say, a scholarly text archive as TextGrid or Deutsches Textarchiv? In my paper I try to check the ability of Parikka's argument to answer these questions.

**Keywords:** media archaeology; archive.

Jussi Parikka in his book entitled What is media archaeology starts the chapter focusing on digital heritage and software culture with a rather well-known statement in the discourse of media theory. Let's take a look at this statement and its possible contexts outside media archaeology.

"Immediacy is the shadow side of mediation" (Parikka 2012 113.) Or to quote the book entitled Remediation by Bolter and Grusin Parikka refers to here directly: "the logic of immediacy dictates that the medium itself should disappear and leave us in the presence of the thing represented." (Bolter and Grusin 5) To put it yet another way: the more effective a media does its job, the more invisible it gets, cherishing the illusion that content is flowing through the mediating channel untouched.

This may easily send us to Marshal McLuhan's famous statements – Bolter and Grusin in the above-mentioned book refer to Understanding Media by the title already: Understanding New Media. At the same time, the phrase 'shadow side' may also bring to mind Boris Groys and his brilliant essay on the sub-medial space of the archive (Groys) – all the more so since Parikka in the respective chapter mentions a media archaeological gaze on the archive and its invisible, hidden mediality.

I quote Parikka: “Media archaeology starts with the archive – the implicit starting point for so much historical research that it itself, as a place and a media form, has been neglected, become almost invisible.” (Parikka 2012 113) Arlette Farge, a scholar who has spent decades with historical research in the physical space of the archive sometimes feels the urge to reflect on it as a media form, in order to make the institute itself visible. Wolfgang Ernst, an emblematic figure of media archaeology, takes notices of Farge when discussing the archive in his Rumoren essay (Ernst 2002). Parikka neglects the wonderful, highly metaphorical and yet still very tangible lines of The Allure of the Archives (Farge) – and he does it not juts here.

But what does it mean that “Media archaeology starts with the archive?” Parikka, Erns and media archaeology in general refers to and expands Foucault’s notion of the archive, which is a modification that goes back to Kittler:

> The centrality of the archive for any cultural and media archaeology is to a large extent a follow-up to Foucault’s expansion of the concept from the concrete physical places of storage of cultural data to the discourses that govern modes of thinking, acting and expression. (Parikka 2012 113)

And:

> Kittler was adamant that we need to make sure that Foucault’s understanding of what governs our contemporary life—its archive—is not only about the statements and rules found in books and libraries. Instead, it is to be found in technological networks of machines and institutions, patterns of education and drilling: in the scientific-engineering complex that practices such forms of power that the traditional humanities theory is incapable of understanding or grasping if it continues to talk about hermeneutic meanings or persists to operate with traditional sociological concepts. (Parikka 2015 2)

The “methodology” of media archaeology for both Wolfgang Ernst and Jussi Parikka means a gesture towards bringing the shadow side of the medium into light, or to put it another way, towards shedding light on the invisible, sub-medial substrata of the medium by scrutinizing technological networks of machines and institutions. We do not risk too much when we state that this gesture or method is of vital importance in the case of digital heritage. If Manuel Castells is right and culture as such is delivered through the digital medium to the user of network society (this is what is described by Castells through the chiastic notion of real virtuality that overwrites and expands the term virtual reality), so if culture is digitally mediated today, then the gesture to expose the shadows of digital mediation is a prerequisite to understand present culture as such.

To cite ta famous paragraph from Castells: “All messages of all kinds become enclosed in the medium because the medium has become so comprehensive, so diversified, so malleable that it absorbs in the same multimedia text the whole of human experience, past, present, and future.” (Castells 404)

Parikka really goes into the shadowy depth of this medium of ‘multimedia text’ in his above-mentioned chapter on digital heritage. In my paper, I will try to show what models he uses to bring the dark side into the light.

Parikka states, that:

> the archive has been a key node in relaying and storing data of modern culture, and hence acted as a key medium in itself – very much connected to the bureaucratic mode of control alongside registering and manipulating data, primarily in offices and through office technologies: typewriters, calculators, spreadsheets, carbon copies and, later, databases, software-based applications, etc. In various accounts of media history, computers themselves are regarded as part of the lineage from papyrus to paper, to printing and the need for advanced information management systems to organize the massive amount of printed materials (…) the database is the primary form of organizing and expressing reality fits the same bill: instead of the narrative, the structural collections of data we call databases form new kinds of information realities enabled by computers. (Parikka 2012 114)

This quote triggers high expectations by referring to the media archaeological research of Cornelia Vismann who discloses in a breathtakingly meticulous manner the material factors behind bureaucratic control and governance, while situating them in a media historical context. The digital heritage, or as the ‘GLAM’ institutions (Galleries, Libraries, Archives, Museums) call them in an EU project phraseology: digital cultural heritage (DCH), as it misses materially oriented and media theoretical reflection, in its mediality stays in the shadow. Is it possible to crack the black box of DCH open? (Here I refer to Parikka’s words from his Geology-book: cracking open the black box of historical archives. (Parikka 2015 147, 150)).

Parikka’s argument (in this chapter and elsewhere as well) suggests that it is possible to outline and develop a Vismann-ian ‘hardcore’ media archaeology not only of the hard drive and software, but their official use as a database, as information management systems as well. The question arises: can we find any trace of such an effort in Vismann’s Law and Media Technology evoked here by Parikka, the book that focuses on analogue bureaucratic, archiving and governance practices? Yes and no. Although the whole initiative of dealing with the material practices of handling files is contextualized by the “ubiquitous promise of the paperless office” of electronic data

procession (Vismann 12), she only emphasizes the possible similarity between the analogue and the digital file management in her last short chapter. There is no doubt, that it is interesting and fruitful to state that: “The history of files therefore also contains a prehistory of the computer.” But from this perspective we certainly won’t expect the unveiling of medial specificity of the electronic data processing. And Vismann’s early death deprives us of the hope that she will augment her media archaeological analysis to the digital world. What is at stake in the sixth chapter of What is Media Archaeology is exactly this move from the analogue to the digital file. But before we try to understand in what way Parikka continues Vismann’s venture, let’s go back to Vismann and to her few remarks on the electronic medium. She tries to legitimize her focus on the material history of the bureaucratic file-handling by pointing to the process of electronic dematerialization. But does this mean, that the move from the material to the dematerialized world of the electronic files lets itself easily analyzed by media archaeology – beyond exhibiting historical context of commonly used terminology of the computer world? According to Vismann: “The very terminology of computer surfaces is designed to remind users seated before screens of the familiar world of files. The menu tab offering options like »list«, »format«, »thesaurus«, »table« and the instructions copy, delete, save turn users into virtual chanceries or chancellors.” (Vismann 163) Not the historical terminological background of such terms as “stacks, files, compiler, or registers” are of importance here, but the process Vismann draws our attention to, in which bureaucratic techniques sink from the state level to the individual one.

> By condensing an entire administrative office, the computer implements the basic law of bureaucracy according to which administrative techniques are transferred from the state to the individual: from the specialized governmental practices of early modern chanceries to the »common style«, from absolutist administrative centers to individual work desks, from the first mainframe computers in defense ministries to the desktop at home. (Vismann 163)

Parikka points to a radically different, maybe even inverse process to make archives comprehendible in the 21st century. Rather than tracing how official, central archiving practices preform everyday usage of the computer, in his first modelling gesture of DCH he compares archives to everyday social media practices.

> However, with the emergence of such new social media ‘archives’ as YouTube, Flickr, etc., the notion of the bureaucratic archive has changed. Modes of accessing and storing data have changed from centrally governed and walled spaces to distributed and software-based. (…) In addition to the bureaucratic techniques of offices, the new archives have to take into account formats,

> medium-specificity, as well as various software-related themes such as encoding. Similarly, despite the distributed nature, one can argue that power still resides in the archive, which is now embedded in architectures of software, and the political economy of social media platforms whose revenue streams are based on the fact of individual everyday contributions through activity: Facebook, YouTube, Google, etc., gathering data on user patterns, preferences and consumer desires, for further evaluation, reuse and reselling purposes.

Parikka – maybe surprisingly – connects the process of the official archives going digital – with social media. So, this is the first model he proposes to disclose the darkness of the digital archive. He does not state explicitly that digital archives today function as web 2.0 practices, he only states that both are software related and physically-technologically shared. Still notable is the fact, that concrete examples of medium specificity come mainly from social media practices in this main part of the argumentation. Even more interesting is that notions of "centrally governed" and "distributed" are placed on opposite sides of a binary opposition. This contradiction suggests that a physically distributed digital architecture (like a distributed, cloud based computer system) – that is not confined to a single physical place and in this respect, cannot be allocated to a concrete physical space – were never centrally governed.

No doubt that new cultural practices of network society has a lot to do with the digital archive. But here a distinction should have to be made. Parikka's argument rightly points to the practical and theoretically relevant problem that contemporary memory institutions face an almost unsolvable but surely unavoidable challenge when they seek to deal with these new practices. How to select, handle, curate or archive them? The problem has a theoretical and a practical aspect as well: memory institutions developed musealizing, archiving, curatorial practices in an analogue world which are incapable of dealing with time-critical, processual media. And on the practical side: it is very complicated, almost impossible to archive a 'born digital' culture. "the theoretical problems of recent media archaeologies of technical media and software along with a rethinking of the archive, go hand in hand with the practical challenges faced by cultural heritage institutions and professionals: how do you archive processes and culture which is based on both technical processes (software and networks) and social ones (participation and collaboration, as in massive online role-playing platforms as cultural forms)." (Parikka 2012 115) But does this rightly staged problem mean that the digital archive of cultural heritage may be modelled through YouTube and Google?

There is another level of Parikka's argumentation concerning the digital archive, a more media archaeological – and more Vismannian one if you like: turning attention to the materiality of the medium. Media archaeology tends to concentrate on the hardware:

> For us, media is approached through the concrete artifacts, design solutions, and various technological layers that range from hardware to software processes, each of which in its own way participates in the circulation of time and memory. The medium is an archive in the Foucauldian sense as a condition of knowledge, but also as a condition of perceptions, sensations, memory, and time. (…) a special emphasis is placed on hardware even if we do not wish to claim that it is the only aspect about media and obsolescence we should consider. (Parikka 2015 146)

Parikka – following Wendy Hui Kyong Chun's work on the enduring ephemeral (Chun), includes the decay, the vulnerability, of the digital medium and its material carrier into the argumentation from both symbolic and material aspects of archiving. He challenges the illusion that 'remediation' of the digital media, the so called "digitalisation" would serve as the final solution to avoid decay of material inscriptions of cultural memory. Museologists and archivists are facing the consequences of this illusion radically when CD-s – meant to be the utmost means of digital eternity – become inaccessible overnight. A productive aspect may indeed be to stage official digital archives as they fight and hide this vulnerability – but this step from the danger of decay to the archival mission of the official institutes is not done here.

Another point where Parikka sheds light to the material aspects of the digital archive is also rather Vismannian: staging digital inscription as the inheritor to analogue writing acts. And still, it is not obviously convincing to model the digital archives on the basis of a however inventive analysis of the writing act: the inscription of the hard drive. He evokes Matthew G. Kirschenbaum and his forensic methodology based argumentation and implies that screen essentialism – when the medium is understood solely by focusing on the visual interface – in the realm of the digital archive can be transcended by going "below", into the depth, down to the digital inscription on the hard drive. „Kirschenbaum focuses on inscription, but in doing so is able to question the dilemmas concerning the so often preferred screen-essentialism, (…). Indeed, the idea of forensics suits both practically and epistemologically, the media-archaeological idea of going deeper and deeper into the materiality of the informational systems." (Parikka 2012 127) No doubt, that traditional archives, museums and collectors are highly interested in the physical inscriptions on material platforms of born digital manuscripts, net art artworks or software. The Literaturarchiv Marbach invested countless creativity and manpower to archive and curate Friedrich Kittler's digital heritage (Enge and Kramski). But does this mean that official memory institutions or digital archives could be modelled focusing on the material inscription of their informational systems that actually provides technological and institutional framework for the whole process of archiving? To be cynical for a moment: as if the technical

and institutional machinery of the traditional archive could be made comprehendible through a focus on the process of the ink absorbed by paper stored in the repository in the basement.

Parikka's argumentation does not leave the space between the surfaces of the screen and the digital inscription that takes place in (or is hidden by) the digital archive empty, but from the practices he uses to enlighten this shadowy sub-medial space (artistic and social media practice of the network society) official digital archival mechanisms are strangely missing. Anything that is "centrally governed". Although these, rightly or not, often define and even legitimate themselves against the "ephemeral" and "chaotic" nature of the unofficial, like social media communication and also popular/commercial sharing services. There is of course overlapping of the official archive and the social media practice (one good example is crowdsourcing), but this does not mean that the difference in the machinery of the two could be by-passed.

I wouldn't say that it was easy to give a media archaeological account of the formal-institutional digital media practices. What I would only point to is that the aspects that make such an account deep enough are missing from Parikka's detailed analysis of digital heritage. Although if one takes a closer look they are not entirely missing: "the new archives have to take into account formats, medium-specificity, as well as various software-related themes such as encoding" (Parikka 2012 115) and Parikka also quotes a study that quotes Kirschenbaum:

"A bibliographic/textual approach calls upon us to emphasize precisely those aspects of electronic textuality that have thus far been neglected in the critical writing about the medium: platform, interface, data standards, file formats, operating systems, versions and distributions of code, patches, ports, and so forth. For that's the stuff electronic texts are made of." (Parikka 20123 85) But he does not have anything to say on how digital cultural heritage is conditioned by, say, data standards, modes of encoding or code versioning. This blind spot makes Parikka's argumentation on such special sub-fields of digital cultural heritage as digital philology especially vulnerable.

An average user won't get through the graphical user interface, so the chance to move beyond screen essentialism is rather low. Some text archives provide access to the encoded text, the text encoded in a mark-up language, but this language – designed to be machine readable and that is comprehendible only to trained digital humanists forms a rather high door-step. The transformations that are constitutional on the level of the archival machinery, that create the images present on the screen based on the encoded text, and the ones that connect the code to the material inscription on the storage surface, are hidden to digital humanists as well, either by authorization limitations or because of the lack of specific IT knowledge.

If so, is it possible at all to provide a deep and material media archaeology of, say, a scholarly textarchive as TextGrid or Deutsches Textarchiv? There are chances of a however partial insight into the dark space between the screen of the user interface and depth of the material inscription by reading specifications provided by the project team, by analyzing the used metadata scheme, the encoding principles and the implemented search query language. Sure, without reflecting upon this level of archival machinery the sub-medial won't let itself enlightened. In this respect Parikka fails to give a proper media archaeological account on official digital media practices. If it is possible at all, it remains an open question in a digital infrastructure where material inscription is moved from the material interface of a single hard drive to a cloud computing environment, where the writing act is hidden by a software architecture continuously evolving itself.

## Works Cited

Bolter, Jay David, and Richard Grusin. *Remediation. Understanding New Media*. The MIT Press, 1999.

Castells, Manuel. *The Rise of the Network Society*. Wiley-Blackwell, 2012.

Chun, Wendy Hui Kyong. "The Enduring Ephemeral or The Future Is a Memory." *Media Archaeology. Approaches, Applications, Implications*, edited by Erkki Huhtamo and Jussi Parikka, University of California Press, 184-203.

Enge, Jürgen, and Kramski, Heinz Werner. "Friedrich Kittler's Digital Legacy – PART I - Challenges, Insights and Problem-Solving Approaches in the Editing of Complex Digital Data Collections." *Digital Humanities Quarterly,* 2017.

Ernst, Wolfgang. *Das Rumoren der Archive. Ordnung aus Unordnung*. Merve Verlag, 2002.

Farge, Arlette. "The Allure of the Archives." *The Lewis Walpole Series in Eighteenth-Century Culture and History*. Translated by Thomas Scott-Railton, foreword by Natalie Zemon Davis, Yale UP, 2013.

Groys, Boris. *Unter Verdacht. Eine Phänomenologie der Medien*. Hanser, 2000.

Parikka, Jussi. "A Geology of Media." *Electronic Mediations*, Volume 46, Minnesota UP, 2015.

Parikka, Jussi. "Archive Dynamics: Software Culture and Digital Heritage." *What is Media Archaeology*. Polity, 2012, 113-135.

Vismann, Cornelia. *Files. Law and Media Technology*. Translated by Geoffrey Winthrop-Young, Stanford UP, 2008.

*Studia UBB Digitalia*, Volume 62 (LXII) 2017, June, Issue 1, 83-88
Published Online: 2017-06-30
DOI:10.24193/subbdigitalia.2017.1.06

# A low te(a)ch approach to digital humanities

Liviu Pop
DigiHUBB, Babeș-Bolyai University, Cluj-Napoca, Romania,
E-mail: liviu.o.pop@gmail.com

**Abstract:** Technological fatigue is spreading due to the increased arrays of technological branches and the rapid pace of development. Keeping up with the latest and newest might prove not only difficult and tiring, but also not necessarily recommended to beginners in digital humanities. A more "back to the basis" approach might be more useful for students that don't have technical backgrounds.

**Keywords:** low tech; minimal computing; technological fatigue.

There are few domains that faced such a fulminant evolution as the computer sciences, the only comparable other domain being genetics and bio-engineering - it took 35 years from isolating the DNA as the genetic material (Avery et al.) to cloning the first sheep, Dolly (Kellan), in 1997. If the prehistory of modern computing can be traced back to the automatic looms of the 18th century from France (Basil Buchon and later Jaques Laquard), the true birth of what we call today computers was the middle of the 20th century. It is mind-blowing what happened in the last half century in the computer science and computer industry. The computers evolved in only 40 years from being highly specialized machines that were calculating missile projections (Colossus), or coding messages (Enigma) during the Second World War, to personal computers affordable in every house (Apple II). Not even the most daring of the first people in the computer industry, except perhaps Gordon Moore (Moore), could have imagined the exponential growth of the computing machines.

> "The complexity for minimum component costs has increased at a rate of roughly a factor of two per year... Certainly over the short term this rate can be expected to continue, if not to increase. Over the longer term, the rate of increase is a bit

> more uncertain, although there is no reason to believe it will not remain nearly constant for at least 10 years. That means by 1975, the number of components per integrated circuit for minimum cost will be 65,000. I believe that such a large circuit can be built on a single wafer."

On the other side of the coin, there were assumptions that proved to be exceptionally wrong, like the anecdotic quote attributed to Bill Gates (Andrews, O'Toole) that we won't need more than 640 KB on a computer. He was referring, of course, to computer running MS-DOS, but none the less, this proves that even the most acclaimed visionaries are suffering sometimes from a form of myopia.

This myopia is what sometimes stops us from grasping the immense and accelerated advanced of the digital technologies that took place in the last decades. In the same time, it might sometimes act as a barrier for having a deep grasp of how they work and where to start in understanding them. One of the most recent example is the state of JavaScript in 2016 and 2017. There were several high-profile articles complaining about the extremely baroque way the JavaScript initiatives are overlapping and depend upon each other. The JavaScript fatigue was coined in 2015 by Eric Clemmons, in an article with the same name (Clemmons), followed up by a more humorous and viral take on the same issue that was published in 2016 by Jose Aguinaga (Aguinaga). Some even tried to create a history and to go deep for the roots of the contemporary state of affairs, getting back to the browsers wars from the 1990s (French-Owen). There were other reactions, of course, defending the diversity and the large array of options that JavaScripts provides, such as the opinion of the JavaScript cheerleader Eric Elliot (Elliot) or the analogies made by Quincy Larson (Larson).

This state of increased complexity and intricate layers build on top and across each other can be found in other parts of the digital landscape. As tutors for future digital humanists, how can we help them better trough the challenge of having a deep and useful understanding of how the contemporary digital technologies work? How can we help them navigate this landscape? My short answer is not to look over their shoulders as they explore each map, but to give them a compass and to teach them how to read a map. Going back to the basics.

One of the options would be to wait for the computers to adapt to the humans, not the other way around. We already see this happening with the increased usage of mobile devices. Mobile devices, the phones and the tablets that became mainstream over the last decade, are powerful little computer always available for us to interact with in our pockets. This lowered the barrier for many digital non-natives in interacting with the computing devices, the multitouch screen being a real facilitator of quicker and easier interactions. But behind the scenes a team made up of programmers, designers, UI an UX designers programmed what will happen on the screen when a certain

action of the user takes place. Programming is still an arcane endeavour and we are still up to achieve the state where the "computer is used as a human brain assistant instead of using the human brain as a computer assistant." (Rus)

But what can we do right now in the field of digital humanities in order to use the digital technologies for our benefit instead of being paralyzed by the myriad of options available? I would like to give a few examples of possible projects that might show an alternative approach. The first example comes from the Minimal Computing Working Group that "kickstarted itself into life with a workshop on July 8 at the DH2014 conference in Lausanne, Switzerland." (Minimal Computing)

The group put together a few "thought pieces" around this subject and thus loosely define what minimal computing might be. It might be an "unfinished stairway" (Oroza) concept:

> "Then, what is a stairway? How does one describe it? Could he build a structure in front of his doorway that looks nothing like a stairway but serves the same function? Maybe just objects stacked in such a way that one can climb and descend them? Or an object by Ettore Sottsass, a stack that includes all of Feijóo's books, a Franz West sculpture, anything?
> He decided on a conceptual shortcut: he built the stairway and waited to be fined. In this way, he gained time. The Law demanded that he cease building the stairway until the paperwork needed to divide the property was finalized.
> Years went by. He used the unfinished stairway.
> What's a finished stairway?"

The concept of unfinished stairway is used by Gil in order to induce the need to short-circuit the whole process of making scientific research available to the public, getting back to using basic tools in order to start "producing our own scholarship ourselves" (Gil).

But the quintessential example of minimal computing, in my opinion, is The Minimal Computing Lab at the Centre for Textual Studies at De Montfort University from Leicester, England. In an attempt to help their students deeply understand how the computers work and considering that "to introduce students to how computers work and how to program them, modern computers – even stripped-down ones such as the Raspberry Pi – are much too complicated" (Egan) they went further down the line of computing history and they setup a lab with Altair 8800 and Amstrad PCW. Now here comes the real minimal computing paradigm: "These machines have no hard drives, no built-in operating systems, and no Read-Only-Memory (ROM). When such a machine is switched on it has empty memory and a processor doing." (Egan)

Even more, learning how to program a very simple computer, by today's standards using BASIC is a less terrifying approach for students that have a background in arts and humanities. The students already have the necessary skills for understanding the

BASIC syntax which is not very far away from finding a paragraph in a scholarly edition. Using numbers as a way of navigating a text structure is something familiar and can be further developed into thinking more abstract code structures. Even more, by facing the limitations of the very simple computers, the students are forced to think in modules and to use more complex algorithms from start. It might come as a surprise, but “the principles of Minimal Computing are an excellent introduction, for those who wish to pursue the topic, to the principles of Supercomputing.” (Egan)

Another member of the group, Sayers, compiled a manifesto in disguise, a list of minimal definitions. The list of minimums and maximums there are an ideological template for achieving the most with the simpler and the most basic of the available technologies. The list of commands tries to answer the fundamental questions of what do/not need or want. As any manifesto, it is an inspiring call to action along the sets of principles that arise from answering those questions. While some of the principles are related to strictly technical choices (“Minimal Maintenance - Reduce dependencies and the use of features to decrease the labour of updating, moderating, and stewarding a project over time”, “Minimal Vulnerabilities - Reduce attack vectors (e.g., cross-site scripting, SQL injections, and directory traversals) of projects to decrease likelihood of hacks and harassment”), other have an assumed social and environmental impact (“Maximum Ephemerality - Reduce an impulse to inscribe, measure, or visualize with technologies in order to increase the likelihood of experimentation and collective participation”, “Maximum Justice -

Reduce the use of technological, cultural, social, and economic barriers to increase entry, access, participation, and self-representation in computing and to also build systems/projects premised on social justice and difference, not white supremacy and settler colonialism.”) (Sayers)

This ethos of minimal computing is something that should be further explored in the way digital humanities are thought to students. Inspired by the minimal computing manifesto we can create projects in which the computing part is hidden, less visible and the focus is more on the creative part and the creative input. A project in which this approach might prove successful is solution for returning immaterial culture back to the villages where it was captured.

The ethnological archive from the Faculty of Letters, Babes-Bolyai University of Cluj-Napoca was conceived as a means to complete the Literary Folklore course (Sava). The documents are kept on paper, being written down, dictated or following direct field observations, or they represent magnetic tape transcripts. We strongly believe that is a moral duty to bring the documents from the Archive of the Folklore Circle back into the villages where they were gathered and we expect an increase in the preservation of the traditional immaterial culture in those villages by learning and re-learning the old customs, stories and songs. The documents have been digitized and now we are

currently having something akin a seed-bank, seeds of a culture that is fading. By using the digital technologies, we would like to replant those seeds back to where they belong (Pop).

One of the way to achieve this is to install small memory boxes with the digital versions of the documents. Those boxes should feel familiar to the villagers and it shouldn't look too "technological", in order to decrease the barrier for interactions. Even though the box will be an installation containing a Raspberry PI loaded with sound files that could be accessed via RFID tags. The most important aspects shouldn't be the computer or the interface in itself, but the content loaded on the computer. In order to achieve this, we will create a custom-made interfaces: a box that will start playing songs when a wooden coin containing RFID tags will be placed on top of it. The originality of the concept doesn't consist in creating a robot that plays music using RFID tags, which was already done and used as an inspiration source1, but on the emphasis put on hiding the computer and the auxiliary interfaces in order to obtain an installation that is less inhibiting. The minimal computing consists in using wide spread technology like RFID for obtaining an installation that does a simple action: starts playing music.

Minimal computing ethos might prove an alternative to the technological fatigue created by a continuous need to be up-to-date with the latest technology. Having this approach for the students that are starting to learn digital humanities might prove fertile.

## Works cited

"About Minimal Computing." Minimal Computing. N.p., n.d. Web. 21 July 2017. http://go-dh.github.io/mincomp/.

Aguinaga, Jose. "How it feels to learn JavaScript in 2016 – Hacker Noon." Hacker Noon. Hacker Noon, 03 Oct. 2016. Web. 21 July 2017. https://medium.com/@ericclemmons/javascript-fatigue-48d4011b6fc4.

Andrews, Nancy. Windows: the official guide to Microsoft's operating environment. Harper & Row Publishers, Inc., 1986.Page 268, Microsoft Press, Redmond, Washington.

Avery, Oswald T., Colin M. MacLeod, and Maclyn McCarty. "Studies on the chemical nature of the substance inducing transformation of Pneumococcal types." Journal of experimental medicine 79.2 (1944): 137-158. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2135445

Clemmons, Eric. "Javascript Fatigue – Eric Clemmons – Medium." Medium. Medium, 26 Dec. 2015. Web. 21 July 2017. https://medium.com/@ericclemmons/javascript-fatigue-48d4011b6fc4.

[1] User *roaldh*, Raspberry Pi Based RFID Music Robot, available at http://www.instructables.com/id/Raspberry-Pi-based-RFID-Music-Robot/.

Egan, Gabriel. "Old Machines Running Old Languages." Minimal Computing. N.p., 03 Aug. 2015. Web. 21 July 2017. http://go-dh.github.io/mincomp/thoughts/2015/08/03/old-machines/.

Elliott, Eric. "Why I'm Thankful for JS Fatigue. I know you're sick of those words, but this is different." Medium. JavaScript Scene, 24 Nov. 2016. Web. 21 July 2017. https://medium.com/javascript-scene/why-im-thankful-for-js-fatigue-i-know-you-re-sick-of-those-words-but-this-is-different-296fae0c888f.

French-Owen, Calvin. "The Deep Roots of Javascript Fatigue." Segment. N.p., 15 Mar. 2016. Web. 21 July 2017. https://segment.com/blog/the-deep-roots-of-js-fatigue/.

Gil, Alex. "The User, the Learner and the Machines We Make." Minimal Computing. N.p., 21 May 2015. Web. 21 July 2017. http://go-dh.github.io/mincomp/thoughts/2015/05/21/user-vs-learner/.

Kellan, Ann. "A sheep cloning how-to, more or less." CNN. Cable News Network, 24 Feb. 1997. Web. 21 July 2017. http://edition.cnn.com/TECH/9702/24/cloning.explainer/index.html.

Larson, Quincy. "JavaScript Fatigue Fatigue – freeCodeCamp." FreeCodeCamp. FreeCodeCamp, 06 Oct. 2016. Web. 21 July 2017. https://medium.freecodecamp.org/javascript-fatigue-fatigue-66ffb619f6ce.

O'Toole, Garson. "Computer Memory: 640K Ought to be Enough for Anyone." Quote Investigator. N.p., 2012. Web. 21 July 2017. http://quoteinvestigator.com/2011/09/08/640k-enough/.

Oroza, Ernesto, and Penélope De Bozzi. Objets réinventés: La création populaire à Cuba. Paris: Alternative, 2002. Print. p. 107

Pop, Liviu. "Open Source Folklore Archives-Disseminating Small Memory Boxes", Review of the National Center for Digitization, 2015. 47-52. http://elib.mi.sanu.ac.rs/files/journals/ncd/26/ncdn26p47-52.pdf.

Roaldh, and Instructables. "Raspberry Pi Based RFID Music Robot." Instructables.com. Instructables, 12 May 2016. Web. 21 July 2017. http://www.instructables.com/id/Raspberry-Pi-based-RFID-Music-Robot/.

Rus, Teodor. "Milestones for Computing Future - (An answer to Vinton Cerf's call for Computing Milestones, Communications of the ACM 07/2015, 7)." N.p., n.d. Web. 21 July 2017. http://homepage.cs.uiowa.edu/~rus/milestones.pdf.

Sava, Eleonora, and Liviu Pop. "The ethnological archive: Memory and technology." Philobiblon: Transylvanian Journal of Multidisciplinary Research In Humanities, 15393 407 (2010). https://www.academia.edu/813668/The_Ethnological_Archive_Memory_and_Technology.

Sayers, Jentery. "Minimal Definitions (tl;dr version)." Minimal Computing. N.p., 03 Oct. 2016. Web. 21 July 2017. http://go-dh.github.io/mincomp/thoughts/2016/10/03/tldr/

*Studia UBB Digitalia*, Volume 62 (LXII) 2017, June, Issue 1, 89-99

Published Online: 2017-06-30

DOI:10.24193/subbdigitalia.2017.1.07

# The importance of data visualization: teaching art history through timelines

Voica Puşcaşiu

DigiHUBB, Babeş-Bolyai University, Cluj-Napoca, Romania,
E-mail: voicapuscasiu@gmail.com

**Abstract:** This research wishes to consider the venerable, but relatively marginal use of timelines as means of visualizing coexisting historical events and figures with special emphasis on Art History, a domain where it loses many of its inherent faults and manages to functional. The use of digital tools in humanities research is on the rise, yeilding interesting results as the timelines are now able to take on new forms. The benefits are thus increased and this chronological representation complements the traditional and established method, especially when it comes to teaching a subject that contains such ample amounts of data.

**Keywords:** timelines; data visualization; teaching; art history.

## Brief Introduction to Timeline Charts

Without even bearing in mind the fact that we are leaving in an increasingly visual world, it has always been clear that visualization is one human beings' most basic tool in understanding the surrounding environment (Brisson 257). It is no coincidence that graphic schemes have been used to translate some of the most complex ideas, as they help in creating a much clearer view on the entire matter. Of course nowadays, this translation has been taken to new heights if we are to consider infographics as the fastest, most efficient way to deliver large amounts of data in the Information Age that has changed the way we think and communicate (Lankow, Ritchie and Crooks), an age of immediate gratification, distributive attention, and much less patience for text-heavy material. So if we are to accept that graphic visualization is among the most useful means in the better understanding of information (Admiraal 53) it should be no surprise that this is also valid for the field of History which deals with such large amounts of data.

The visual method of presenting historical events through timeline charts came to be known in the 18th century as “chronography” from its combination of chronology and graphics (Admiraal 53), and it has largely stood the test of time, although it is surprisingly rarely used as a veritable teaching tool. Stemming from the genealogy and chronology charts of early 15th an 16th century, the most primitive shape they took was that of a table which contained the year on one column and the events on the second one, but from this, the timeline charts evolved in a multitude of styles, from the simple to the highly eccentric. One thing they do all have in common is the desire and intention to offer a global view of history of mankind as a whole. This is a seemingly impossible task, but it is made slightly easier through a graphic medium as a holistic view on such a large scale could not possibly be grasped otherwise. This desire to counteract the national discourses and the display of events grouped on geographical areas or certain time spans is beneficial for the realization that events and figures rarely stand on their own, but are interconnected in intricate ways.

The fact that complexity is often confusing is no less true for timeline charts, so it is obvious that most overly-zealous models are impractical due to the desire to include as much information as possible. This goes against the very idea of schematization and is one of the reasons some failed to get their point across, along with the biased that is sometimes present within them as they tend to give much more information about a specific era while ignoring others. The model that remained the most valuable due to its objectivity was given by Joseph Priestly in his New Chart of History (1769), which he dedicated to Benjamin Franklin (Jacobs 129), where he introduces a consistency of scale in his spatial representation of time that was not present before. The unvarying intervals that divide the timeline became the norm, as did the horizontal bands that indicate the lifespan of each person included in the chart, running from left to right in the direction of reading and it is this method that can be seen in the most basic linear chronologies that are present at the beginning of most history textbooks.

Chronology onto itself is one of the very first lessons introduced when learning about history (Păun 81), as placing an event in time is fundamental to its understanding, but this chronological approach underlined through the use of a simple timeline in not without its faults, mainly because it underlines a certain evolutionary aspect in history which is dangerous as it greatly diminishes the importance of later eras as compared to more recent ones. Perhaps even without setting out to do so, the focus on chronology creates the illusion of time passing from primitive societies all the way to the height of technology of the contemporary age and because of its schematic character it leaves no place for a more nuanced discussion. A second issue that might arise is the forceful implementation of a cause-and-effect principle (Stinespring and Steele), which although valid and of great importance for the study of history is not necessarily true in all situations, just because an event preceded another it does not automatically means

it also caused it so this possible reading needs to be taken into account whenever creating or analyzing a timeline chart. Thus the most useful graphics are not the ones that simply visually represent a succession of facts or figures, but the ones that are multi-layered, including multiple bands representing different geographical zones and if possible they should strive for interdisciplinarity in order to offer an ampler context to events and manifestations.

**Timelines in Art History**

Concerning the field of Art History, asides the "history-as-evolution" issue, many caveats of the timelines are lost, especially when comparing them to the real benefits this type of visualization can bring when dealing with very specific cause-and-effect scenarios that are inevitably present and important. But once again their presence is not as widespread as one would expect, despite being used by some of the most influential authors, mostly in the works that were designed for educational and popularization. Although not textbooks per se, the books containing timelines were used for this purpose and include some of the best-selling Art History volumes of all times such as H.W. Janson's History of Art (Janson) and Ernst Gombrich's The Story of Art (Gombrich) both of which have kept and refined the timelines section throughout their many editions. Besides offering a more global picture of the artistic manifestation, these graphic representations offer at a glance the artists and their contemporaries, which as Priestley said, gives "a peculiar kind of pleasure" when one imagines the possible conversations (Priestley qtd. in Jacobs 129). This instant emotional connection is incidentally another one of the visualization's benefits, which would be hard to achieve otherwise.

A problem that is specific for Art History, but finds a solution through the use of the timeline charts is the different speeds at which a certain style reached different geographical areas. The traditional and well established manner of writing about Art History is rarely more than a succession of styles, and even though they are sometimes differentiated by location this can prove both hard to understand and easy to forget. The most obvious case is that of the Renaissance whose ideals and manifestations are considered to have appeared with some delay in Northern Europe and kept a rather Gothic appearance for even longer than that, Janson himself created two separate timelines regarding Italy and the Northern Countries in order better to underline this issue (Bork 182). This however seems unsuitable, because if they were to be presented on the time scale, an even better appreciation of this delay could be achieved, as it would give the viewers the occasion to realize that Jan van Eyck, a pioneer of the portrait, was long dead before Leonardo da Vinci was even born despite the fact that in the usual Art History discourse van Eyck follows da Vinci. A different means of

reaching conclusions that would be hard to come by through traditional methods is if one would choose to point out on the timeline specific artworks and analyze their similarities and differences. A good example in this case would be Raphael's Crucifixion (1503) versus Matthias Grünewald's Crucifixion (1502), created just one year apart they have the same theme and a strikingly similar composition, and yet there is absolutely nothing of Raphael's serene color scheme and blissful figures in Grünewald's gruesome depiction of torment.

A different side of this issue is the customary division of the Baroque style based upon the different countries it reached, usually starting with Italy, continuing with Spain, The Netherlands, and so on. This flow of the discourse that seems to end each chapter before starting a new one is disruptive simply because it tends to overlook the concomitance of artists from different geographical locations and their possible interactions and influences on each other. What's more, the stylistic division in itself is problematic because it mostly overlooks the artists or works that are on the fringes of a certain current, or more frequently they get labeled into either one of the styles while ignoring the characteristics of the other. Through the use of timelines, these works could more easily show the way styles flowed and not abruptly ended in order to give way to the next one. This approach, based on concomitance and correspondence is notably used by Gombrich but is surprisingly rare afterwards, except in Julian Bell's book (Bell), and almost never in textbooks. Even rarer are the visual schemes that depart from the chronological timelines in order to show the influences one current has on others, or that others have had on it, but such examples do exist, an original and entertaining one being that of Alfred H. Barr Jr. concerning Cubism and modern abstract art (Jacobs 133). This goes to show that visual aids in Art History should not always be contained to Priestley's model in order to be effective, but that there are many more imaginative ways through which a valuable point could be made.

Besides Barr's suggestion of graphic representation that raises awareness on the many obscure artistic subdivisions that hardly ever get mentioned, even when it comes to timelines there are a number of different types that could be thought up in order to serve as many different purposes, depending on the discourse it is associated to, be it explicit or implicit. One of the ways in which timelines could be constructed implies following the artists' lives and pinpointing personal events that are thought to have influenced their work. This style could work if one starts with the presumption that personal events shape an artist's work, as some art historians do, and they would be accompanied by relevant examples from their oeuvre that prove this point. The events on the timeline could range from the death of a loved one, to happier events and from the artist's educational background to his moving to a different location and coming into contact with a different culture. This type of timelines would be instrumental in establishing the course of an artist's work and view the stylistic

and thematic changes it has gone through. As it is highly unlikely that a single event could provoke an instant change in an artist's style, these should only be seen as triggers towards change, while the timeline could also lead to a better appreciation of the lifelong artistic exploration of a certain theme of motif instead of just breaking up the entire volume of one's work into definite periods. Of course following the stylistic transformations of an artist's work does not necessarily have to be connected to his personal life or any other external events, a simpler version of this type of timeline could include just the works set up in a chronological manner, from which one could presumably get a better grasp on the nature of these transformations.

A different aspect of timelines could be used by those interested in establishing the iconographical changes (Szabo 159), through the ages by considering specific works that follow the same pattern, either through theme, style, compositions or the symbols they use. In this sense thematic timelines would pinpoint the most relevant artworks in order to offer a quick view of the entire history of one genre, while at the same time creating a clear connection between the artists who influenced each other. The possibilities for these timelines are endless and they could be as broad as the representations of a fruit bowl from Ancient Egypt to Matisse, or as narrow as the depictions of Napoleon in various stances. Even when tackling one genre in particular the examples can be broadened to all types of female nudes throughout history, regardless of the medium, or more focused on a precise characteristic such as the reclined nude. In this case a timeline could point out Giorgione's Sleeping Venus (1507), Titian's Venus of Urbino (1538), Velázquez's Rokeby Venus (1651), Ingres' Grand Oladissque (1814), Manet's Olympia (1863), Cézanne's A Modern Olympia (1873), and ending with the work of Guerilla Girls Do women have to be naked to get into the Met. Museum? (1989).

Much rarer, but much more useful are the timeline charts that while referring to Art History also choose to present things in a larger context, by offering a interdisciplinary view of a time section. Such examples can be found in Norbert Wolf's book on Romanticism (Wolf) that offers information on activities connected to art on a timeline that is placed at the bottom of each page and runs through the entire volume or David Gariff's endeavor (Gariff) of establishing artistic influences based solely on graphic representations of timelines. The utility of these charts is clearly seen if one is to consider the large effect the domains outside of art have on artist production throughout the ages all of which could be shown in a simple and efficient manner. Considering the fact that a great deal of art is not self-referential, but is inspired by various fields and events, it makes perfect sense to include these in order to gain a truly global view on Art History. By and large the most common influence comes from historical events, be they wars, inventions or relevant figures, all of which have prompted very specific manifestation in the visual arts, as well as in

literature and music. The problem is that these connections are more or less direct and they are not always easy to be touched upon in a coherent manner, thus a visualization of selected and pertinent data is preferable (Szabo 129). If in the case of history the influence upon art is unilateral, as significant events have rarely been precipitated by a work of art, when it comes to literature and music, the influences are mutual and there are many examples of this practice, so in order to fulfill a truly global purpose, layers including these fields could also be added to the timeline.

Accepting Arnold Hauser's thesis on the fact that all art is fundamentally connected to the society in which it was created (Hauser 6), it becomes obvious that not enough is being done in order to underline this causality. The fact that (Art) History is perceived as a series of epochs and (art) historians are specialized in a narrow section (Bork 179), creates the false impression of a clarity of boundaries which is detrimental for those who are not yet well-versed in the intricacies of these fields. It is for this reason that visualization could become a practical teaching instrument, that while covering an ample set of information, is not as dense and demanding, but rather "user-friendly". Interdisciplinary timelines are even more important because they are able to offer a cross-section of any given time period and paint a much more comprehensible picture of it. For example considering the time surrounding the French Revolution, a single visualization could mark the Storming of the Bastille (1789) but also what that has meant to for the arts: La Marseillaise (1792), Delacroix's Liberty Leading the People (1830), and Hugo's Les Miserables (1862), all four of which would have traditionally been discussed at different times, during the study of different subjects. What is even better is the fact that timelines are a universally applicable method that could be used for either for specific well-defined situations, or for a more general aspect, thus creating a sense of structure, also by containing less text but more detail they are superior to regular charts that are most often found in history textbooks (Stinespring and Steele 9).

Each of the methodologies for creating a timeline, briefly discussed above can have as a result a tool that could be used for either education or even just for broadening the general interest towards this field, it is however important to accept that as a schematic representation some things have to be left out (Admiraal 56), and mere existence in not a criteria for inclusion. As clutter is one of the most counterproductive aspects of a visual representation and history textbooks are notoriously crammed with too much data (Felezeu 107), it would be the timeline's task to right this wrong and not concentrate on isolated details. One of the most important didactic principles that is delivered through timeline charts is that of comparison (Felezeu 183), this refers to two different procedures, both of which could be achieved. First there is precisely the one discussed earlier in this paper concerning similar events of manifestations that happened at very different historical times and

the second one refers to simultaneous comparison, involving events and phenomena that happened in the same time span but under very different circumstances. A prime example for this could be the existence of a Baroque style in both Catholic and Protestant countries. Finally not only do timelines serve an educational purpose, but they are also visually attractive, which is of utmost importance especially nowadays when society is surrounded by glossy, colorful prints and images that captivate the eye and the mind.

**Digital Art History**

Data visualization was a term initially used by statisticians (Francis and Fuller 301), but has since become a modern term for a graphic representation of a set of data. It can be used in reference to a “hand-made” graphic display, but through recent developments, it has also become associated with software-generated images and charts that provide insight into scientific data, though traditionally used for engineering, medicine, and geology, this approach has slowly started to appear in humanist fields as well as of late. Despite the boom in technology that greatly affected the way graphic information in being presented, some traditional mediums have still remained viable and timelines are notably one of them (Admiraal 53). The timeline chart appears to have successfully passed the test of time and the translation to a digital environment which means that all its best traits are still sought after, underlining undeniable the fascination for its elements, be it the glimpse into simultaneous events, their simple and logical chronology, their application of the cause-and-effect principle, or just the fact they are intuitive to use. Whatever the reason it is quite clear that they are very much in use today and the existence of software such as Timeline JS just proves there is a market for it and that people are more than willing to create their own charts, often for educational or presentation purposes.

In order for generated timelines to be considered efficient it still needs to follow the same characteristics as before, it should be clear and appropriate for communicating the information as its success is mainly determined by the speed at which it manages to do just that. Even though its main goal still resides in conveying a message or showcasing patterns hidden in the data (Ramirez Gaviria 479), when compared to older, hand-made timelines, it becomes clear that new technology brought along a more aesthetic dimension into all visualizations, and it is through this aspect that innovative ways of presentation could eventually be reached, instead of relying on the same model that dates from the 1700s (Jacobs 133). What is more important is that even though it is still true that not all data could (or should) be contemplated at once, computer visualizations have the added bonus of being dynamic and interactive, so through filters, panning, and zooming the viewer can choose what

to see at a given time, according to his interest, while the rest of the data, though still present and available will be hidden from view. One can imagine in the case of Art History a timeline of an artists' life containing each and every one of his works in chronological order, but while a complete visual would be overwhelming, the viewer could choose to view only the works from a specific year, for example. This dynamic data visualization is considered to be one of the only genuinely new cultural forms enabled by computing (Manovich 3), the possibilities it offers are virtually endless and contemporary timelines have started to make use of these new prospects.

Without going into great detail, two timeline-based projects should be mention and even though many others exist, they are all only starting to scratch the surface of what could be done through this medium. The first and probably one of the most famous is the Heilbrunn Timeline of Art History hosted by the website of the Metropolitan Museum in New York. It was launched in 2000, spanning from 8000BCE until the present day and contains over 6000 photographed works of art from the museum's vast collection (Van Kirk 290). It is also accompanied by an interactive map of the world so that location can be used as search criteria, and it compiles 300 separate timelines that are stacked or layered for comparison (Knapp 28), and it is enhanced with so-called Thematic Essays, providing an engaging way to do research or just browse the collection. Data is always being added as well as practical information about which pieces are on display and where, and which are currently in storage, so it fulfills a secondary role in helping plan a visit to the museum. Of course this extensive and impressive project could only be upheld through an equally impressive funding in this case provided by Heilbrunn Foundation, New Tamarind Foundation, and Zodiac Fund, to which such a renowned institution is likely to have access.

Other examples are on a much smaller scale, but no less interesting and one of them is Giorgia Lupi's Visualizing Painters' Lives (Lupi) done in collaboration with illustrator Michela Buttignol. The project is made out of 10 separate timelines, each belonging to a painter, and the idea behind it was to build a "visual anthology" of some of the most recognizable painters of the modern era by borrowing specific pictorial elements from each artist in order to tell their story. Although each one marks events with unique symbols, the timelines are all structurally uniform, so they also function as a whole, as a series. What makes them particularly interesting is that they don't only contain details about artistic production in recognized stylistic periods of a painter's life, but also pinpoint his education, training, awards, as well as connections and influences. There is at the same time a biographical aspect to them in the information about the artists' habits, health, travels, and romantic connections, in order to offer a view of the man as well as the professional, all in one elegant, whimsical, and informational panel. This kind of projects push the boundaries of both Art History and data visualization, striving for both clearer information and better

design, thus software development could be combined with a comprehensive database, this should currently be priority that begs the collaboration of all institutions in order to digitize their precious collections.

In order to be able to speak of a proper Digital Art History, the developments in the field of Digital Humanities are a good start, but not nearly enough so Murtha Baca and Anne Helmreich have identified five steps, or phases, that need to be completed in order to go from digitizing to fully digital (Baca and Helmreich qtd. in Ross). The very first step is well on its way with many initiatives such as Artstor, the Google Art Project, and that of the Getty Research Institute, along with many other more focused ones like The William Blake Archive or Vincent van Gogh: The Letters, which not only make valuable resources digitally available but enhance the regular viewer experience by offering impeccable resolutions of the pictures that are now suitable for an whole range of analyses or even just for personal entertainment. The second step towards Digital Art History involved the creation of specific tools in order to provide the infrastructure for building a collection, accessible to the individual scholar, this way collaboration would be encouraged and students could easily become producers as well as consumers (Baca and Helmreich qtd. in Duggan). This step materialized with the appearance of software like Zotero and Omeka, both of which have been designed with scholars in mind in order to help with creating and sharing collections. Phase three supposedly involves the creation of new technologies for visualization, reconstruction, and virtual reality, which would mean that works could be understood in a very different manner that it was possible until now, and although there are projects in this direction, like Virtual Burnham Initiative, the process seems to have stalled as they most rely on the concept of curated collection. It is the fourth step however that would have a true implication for the future of this field as it would involve new means of scholarly publishing and peer review, which along with the fifth and final phase of using new models of scholarly research enabled by computational analyses, would really change the way in which Art History is experienced and thought while harnessing the full potential of computers for this field and its scholars.

The direction in which research is seems to be headed in our contemporary society could bring new and unexpected forms and scholars and educators should be able to adapt to this new state of things, especially when it comes to teaching generations that are “born digital” it is however obvious that the past still has worthy models that first need to be known and experienced in order to be reinterpreted and re-imagined through technology. Timelines have much to offer and it appears that they have not reached their full potential, so it is a direction worth following for the future of the Digital Humanities.

**Works Cited**

Admiraal, Ashley. “The History Timeline Chart”. Agora. Vol. 47. No. 1 (2012): 53-59.

Bell, Julian. Oglinda lumii. București: Vellant, 2007.

Bork, Robert. “Pros and Cons of Stratigraphic Models in Art History” RES: Anthropology and Aesthetics. No. 40 (Autumn 2011): 177-187.

Brisson, Harriet E. “Visualization in Art and Science”. Leonardo. Vol. 25. No.2/3. Visual Mathematics: Special Double Issue (1992): 257-262.

Duggan, Bob. “What would Digital Art History Look Like?”. April 21 2017. <http://bigthink.com/Picture-This/what-would-digital-art-history-look-like>.

Felezeu, Călin. Didactica istoriei. Presa Universitară Clujeană: Cluj-Napoca, 2000.

Francis, Brian and Mark Fuller. “Visualization of Event Histories”. Journal of the Royal Statistical Society, Series A (Statistics in Society). Vol. 159. No. 2 (1996): 301-308.

Gariff, David. Cei mai influenți pictori din lume și artiștii pe care i-au inspirat. Enciclopedia RAO: București, 2008.

Gombrich, Ernst. Istoria artei. București: Pro Editură și Tipografie, 2007.

Hauser, Arnold. The Philosophy of Art History. Cleveland and New York: Meridian Books, 1967.

Jacobs, Alan. “History as Wall Art”. The New Atlantis, A Journal of Technology & Society. (Fall 2010): 127-133.

Jansen, Leo, Hans Luijten, and Nienke Bakker (Eds.). Vincent van Gogh: The Letters. April 21 2017. <http://vangoghletters.org/vg/>.

Janson, H.W. and Anthony F. Janson. History of Art. Eight Edition. New York: HN Abrams, 1991.

Knapp, Jessica. “Clay Culture: Timeline of Art”. Ceramics Monthly. June/July/August (2012): 28.

Lankow, Jason, Josh Ritchie and Ross Crooks. Infographics: The Power of Visual Storytelling. New York: John Wiley & Sons, 2012.

Lev Manovich. “Data Visualization as New Abstraction and as Anti-Sublime”. Small Tech. The Culture of Digital Tools. edited by Byron Hawk, David D. Rieder, and Ollie Oviedo. Minneapolis: University of Minnesota Press, 2008: 3-9.

Păun, Ștefan. Didactica istoriei. București: Corint, 2001.

Ramirez Gaviria, Andres. “When is Information Visualization Art? Determining the Critical Criteria”, Leonardo. Vol. 41. No. 5 (2008): 479-482.

Ross, Nancy. “Teaching 20th Century Art History with Gender and Data Visualization”. The Journal of Interactive Technology and Pedagogy. Issue 4 (Fall 2003).

Stinespring, John A. and Brian D. Steele. “Teaching Art History: Getting Started”. Art Education. Vol. 46. No. 2. Art History (March 1993): 7-13.

Szabo, Victoria. “Transforming Art History Research with Database Analyses: Visualizing Art Markets”. Art Documentation: Journal of the Art Libraries Society of North America. Vol. 31 (Fall 2012): 158-175.

Van Kirk, Shannon. “Heilbrunn Timeline of Art History”. College and Research Libraries News. May (2012): 290-291.

Wolf, Norbet. Romanticism. Cologne: Taschen, 2007.

Giorgia Lupi. Visualizing Painters’s Lives. April 20 2017. <http://giorgialupi.com/work#/visualizing-painters-lives/>.

Timeline JS. April 20 2017. <https://timeline.knightlab.com/>.

Art Project - Google. April 21 2017. <https://www.google.com/culturalinstitute/project/art-project>.

Artstor. April 21 2017. <http://www.artstor.org/>.

Getty Research Institute, The. April 21 2017. <http://www.getty.edu/research/scholars/digital_art_history/index.html>.

Metropolitan Museum of Art, The. April 21 2017. <www.metmuseum.org/toah/>.

Omeka. April 21 2017. <http://omeka.org/>.

Virtual Burnham Initiative. April 21 2017. <http://vbi.lakeforest.edu/>.

William Blake Archives, The. April 21 2017. <http://www.blakearchive.org/blake/>.

Zotero. April 21 2017. <https://www.zotero.org/>.

*Studia UBB Digitalia*, Volume 62 (LXII) 2017, June, Issue 1, 100-111

Published Online: 2017-06-30

DOI:10.24193/subbdigitalia.2017.1.08

# eXist DB or Saxon/C in PHP. A comparison between two approaches for XSLT 2.0 based websites*

Christian Schwaderer

Eberhard Karls Universität Tübingen, Germany,
E-mail: c_schwaderer@hotmail.com

**Abstract:** XML and XSLT are very popular technologies among Digital Humanists. However, when it comes to deploying an XSLT 2.0 processor into a website infrastructure some difficulties arrise. This paper discusses two possible approaches for doing XSLT 2.0 transformations on the fly on a Web (Application) Server: Saxon as a PHP extension and Saxon within the popular XML database eXist. The conclusion shows, that both solutions have their pros and cons. In the end, however, the PHP solution wins.

**Keywords:** XML; XSLT; PHP; eXist.

There is no need to prove that XML is popular among Digital Humanists. While in recent times, JSON became probably the most important exchange format for structured data in commercial applications, it doesn't meet the needs of humanists, who often have to deal with semi-structured texts and thus with mixed content. Therefore, there is no real alternative to XML and none is in sight in the near future either.

This being the case, there is a need for practical ways to present XML structured data to end users in a web interface. While there are nowadays various ways to process and transform XML data, at least in a DH context, using XSLT seems to be among the most common methods.[1]

* I'd like to thank Christopher Johnson and O'Neil Delpratt for their valuable help and feedback. Of course, any remaining errors are mine.

[1] Take for example last year's programme of the most important summer school in Digital Humanities. In the workshop “Text to Tech” (http://dhoxss.humanities.ox.ac.uk/2015/text2tech.html) “a hands-on introduction” to ”XPath, and XSLT” was announced.

While deciding how input data should be displayed on the screen and writing XSLT code for that purpose, might be regarded as the main and major work process in building a digital humanities website, it is – unfortunately – not all that is to be done. By having sophisticated XSLT templates transforming input XML to pretty HTML pages, you are just half way there in setting up a web application. If you only want to transform data locally on your own machine, then you maybe an IDE like oXygen fits your needs best. However, once you decide to build a website:

1. You need to make sure that the user gets the appropriate HTML file for her request. In other words: You have to find solutions for session management.
2. You might want to give your users a site internal search.
3. Finally, you need some management solution for your XSLT processing. While you could start the XSLT processor of your choice from the command line on your local machine and copy the generated HTML files manually to the place you want to have them, this process becomes annoying if you are dealing with dozens of transformation processes and hundreds of result pages. If you want interactive components on your website with the help of XSLT, you have no choice anyway: You need a program for starting XSLT transformations on the server.

Probably, there are at least half a dozen solutions for the needs briefly outlined above. Many things depend on which XSLT processor you are using. However, if you don't want to miss the features of XSLT 2.0 and XPath 2.0 the Saxon XSLT processor family might be the choice for you. Thus, the number of methods narrows down. If you don't want to sue Saxon on the command line or write your web application in Java, there are not too many options on how to integrate Saxon into a website infrastructure which highly relies on using XSLT 2.0 for creating HTML files the end users get to see.

In this text, I want to outline my experience with two different approaches.

Roughly speaking, the first way (which I did for my Database of the Letters of Pope Gregory VII project[2]) is as follows:

- Run a web server like Apache on a server.
- Store your XML data on a server's file system.
- Install Saxon/C as a PHP extension.
- Develop PHP scripts for starting Saxon, storing the result HTML files on the file system and/or delivering them to the client.

---

In the “Editiones Electronicae Guelferbytanae”, the XSLT code itself is estimated as of being of such relevance as scholarly data that they even publish it together with the XML.

Of course, not all people like XSLT: “I don’t know about you, but I honestly can’t stand XSLT.” Jesse Alama, XSLTXT: A more compact form of XSL, April 20, 2015 http://goxrxyourself.com/2015/04/20/xsltxt-a-more-compact-form-of-xsl/.

[2] http://www.g7ldb.history.uni-tuebingen.de/.

☙ For the setting up a search engine on your website, build a PHP script which turns what user input in the search form into an XPath expression, then put this XPath into an XSLT script and process it.

The alternative (which is what I did in my edition of “Adelbert of Heidenheim”[3]) would be:

☙ Run eXist XML database on a server.
☙ Put your XML data into eXist. eXist will then store it internally as binary files.
☙ Develop XQuery scripts for starting Saxon, storing the HTML results within eXist and/or delivering them to the client. (As of now, Saxon already comes with eXist.)
☙ For the search, set up the eXist index options and develop an XQuery script which queries your data using the Lucene query engine deployed in eXist.

In the sections to follow, I'd like to compare both approaches and discuss their pro's and con's. The versions used are Saxon/C 0.3.1 beta and eXist db 3.0RC1.

**Installation and set-up**

eXist ships almost ready to use out of the box. All you have to do is run the installation jar and “answer a few questions” on the way through the installation process. Afterwards, in order to have eXist installed as a service, you need to run `$EXIST_HOME/tools/wrapper/bin/exist.sh install` – and you're done. (Unless something goes wrong like a port conflict with Tomcat or something like that.)

In order to use Saxon within PHP (and not via the command line) you have to install the Saxon/C PHP extension provided by Saxonica. And this is not so easy, since yet, there is no installation script for the various Linux distributions (I did not try it on Windows). So, you have to do a few steps by hand, like compiling, making, and setting things in PHP.ini. Personally, I had a lot of trouble and needed a lot of help from the Saxonica guys, before Saxon/C finally worked on openSUSE 13.1.

However, once the software runs, in the Saxon/C PHP extension scenario, you are done. There is pretty nothing to configure or set up. Either it runs or it doesn't. But if it does, there is nothing left to do.

Though you can start using eXist right after its installation, for XSLT users, there are some things to do. For example, in order minimize the white space problem (see below), in `$EXIST_HOME/conf.xml`, you have to ensure that the `preserve-whitespace-mixed-content` attribute is set to `yes`.

[3] http://www.forschungsdatenarchiv.escience.uni-tuebingen.de/adlils/

Also, there is a problematic default setting called “XSLT caching”. You can easily get rid of it by finding `<transformer class="net.sf.saxon.TransformerFactoryImpl".. ` element and setting `caching` to `no`. Otherwise importing or including XSLT files into other XSLT files (`<xsl:import href="...xsl"/>`) will cause you trouble, because the included/imported files will be taken from the cache, thus often being in an outdated state if you are currently working with them.

**File management and data handling**

We have already mentioned that eXist does not directly use XML documents on the file system but stores the XML iput in internally as binary data. While this is necessary for any XML database to work efficiently, it may affect the things you can do with your files and which you can't.

First of all, the eXist data directory consumes double the hard disk space that HTML files on the file system do. While in most cases, this might not be a problem at all, some other issues are.

If your XML data and your automatically created HTML files are stored on the file system you remain in full control over them. You can manipulate files with non XML-aware script languages like PHP or whatever. That's important in case you want to achieve something which is difficult in pure XSLT.

Take, for instance, the HTML5 doctype definition `<!DOCTYPE html>`. There is no “official” way to output it in XSLT 2.0.[4] However, you can easily write a small script which simply prepends the doctype as a string to all your HTML files. PHP doesn't care about the content of your files, you can tamper in any way you want with your files, regardless if you produce well formed XML or not. It's all up to you.

eXist on the other hand, is an XML database, so it cares if your data is well formed XML or not. It also cares about doctypes, putting them out according to its serialization settings (and removing them from the input). So, if you want to do your own thing, you probably will have a hard time. Additionally, XQuery is not a language designed and suitable for manipulating files and strings. (I didn't succeed in prepending `<!DOCTYPE html>` to all my HTML files stored in eXist.)

Concerning backup and version control, as far as I know there are no external tools to do this in eXist. Sure, any application that is able to deal with the XMLDB API[5], could make use of the data stored in eXist, but I don't know of any. (Additionally, even if you have an external backup of eXist's binary files there is no guaranty that you will be able to get them to work within eXist again, as I have experienced myself.)

[4] For some tricks, see http://www.whoop.ee/posts/2014/08/22/valid-html5-doctype-with-xslt.html

[5] See http://xmldb-org.sourceforge.net/xapi/ . As far as I know, there is not much support for this API anyway.

In the end, you have to rely on eXist's own built-in functions.

One really enervating issue is that eXist's doesn't exactly store XML documents the way you put them in. Not only does it remove the doctype definition (as mentioned forehand) it also tampers with white space.

For example, putting in

`<tei:rdg><tei:supplied>et</tei:supplied></tei:rdg>` will by eXist be turned into:

```
<tei:rdg>
   <tei:supplied>et</tei:supplied>
</tei:rdg>
```

Now imagine, you have a template for `tei:rdg` which wraps its contents in brackets and one for `tei:supplied` which puts quotation marks around the nodes inside. You probably expect a result like `(“et”)`. However, with the added white space by eXist you will get `( “et” )` which at least looks silly, is simply not what you want, and an absolutely unnecessary inconvenience. (In my case, I had developed two routines for my apparatus entries: One for single words and one for more than one word. The distinction was made by the existence of white space: If there was any the input was treated as two words. That worked perfectly till I began using eXist. Now, because of the added white space all single words were treated like they were two words. It took some days to figure out a solution here.)

### XSLT support and error handling

Despite Saxon being in both solutions the XSLT processor in the back end, there are some issues to point out in the handling and integration of XSLT.

When speaking of any code processor, compiler, and interpreter, one of the first things developers are interested in is the information you get if your code produces an error. Getting detailed error reports is absolutely crucial for debugging.

Using an XSLT script which is terminated by an XSL message (`<xsl:message terminate="yes">I am a terminate message!</xsl:message>`), we test the output we get. For both solutions, the most detailed information is to be found in the log files.

For Saxon/C in PHP we will see the following in the Apache “error_log” file:

```
I am a terminate message!
Error at xsl:message on line 15 of param.xsl:
  XTMM9000: Processing terminated by xsl:message at line 15 in param.xsl
Exception in thread "main" net.sf.saxon.s9api.SaxonApiException: Processing
terminated by xsl:message at line 15 in param.xsl
```

```
      at net.sf.saxon.s9api.XsltTransformer.transform(Unknown Source)
      at net.sf.saxon.option.cpp.XsltProcessorForCpp.xsltApplyStylesheet(Unknown
Source)
Caused by: net.sf.saxon.expr.instruct.TerminationException: Processing terminated
by xsl:message at line 15 in param.xsl
      at net.sf.saxon.expr.instruct.Message.processLeavingTail(Unknown Source)
```

(Four further lines of Java stack trace omitted.)

For other errors, you even get the number of the column:

```
Error at xsl:variable on line 14 column 69 of param.xsl: XTTE0570: Required
item type of value of variable $my_val is xs:integer; supplied value has
item type xs:string
```

In eXist's main log file “exist.log” the output is slightly different:

```
2015-11-12 15:04:57,243 [eXistThread-95] WARN  (Transform.java
[fatalError]:817) - XSL transform reports fatal error: Processing
terminated by xsl:message at line -1 in null
net.sf.saxon.expr.instruct.TerminationException: Processing terminated by
xsl:message at line -1 in null
      at net.sf.saxon.expr.instruct.Message.processLeavingTail(Message.java:223)
~[?:?]
      at net.sf.saxon.expr.instruct.Choose.processLeavingTail(Choose.java:796)
~[?:?]
      at net.sf.saxon.expr.instruct.Choose.processLeavingTail(Choose.java:796)
~[?:?]
      at net.sf.saxon.expr.instruct.Instruction.process(Instruction.java:131)
~[?:?]
```

And then you get about two hundred lines of Java stack trace which I omit here since it is of no use for debugging XSLT code.

You see, there are two main differences:

- In eXist, the actual message is not there. It's not logged in the main log file, but passed to the general system output, thus being either in the console window or in “EXIST_HOME/tools/wrapper/logs/wrapper.log”. That's a little bit inconvenient, but once you know where to look for it, it's just a few clicks extra.
- The line number of the XSLT snippet causing the problem is missing in eXist. That's really painful and makes debugging for a file with 5041 lines (my main XSLT file) almost impossible.

Still speaking of error handling, it is not only important what is put out to the log files but also how the wrapping PHP respectively XQuery script itself behaves in case there are some XSLT errors.

With Saxon/C in PHP, if calling `transformToString()` results in an XSLT error, nothing happens. You will just get an empty string as a result and the rest of the PHP script is executed as normal. (If Saxon/C itself throws an error like running out of memory that's another story. However, in version 0.3.1 I have never experienced something like that.)

After the execution of `transformToString()`, you can check the result and possibly access any error messages like so:

```
if($xslt_result_string == NULL) {
         $errCount = $xslt_processor_instance->getExceptionCount();
         if($errCount > 0 ) {
          for ($i = 0; $i < $errCount; $i++) {
             $errC = $xslt_processor_instance->getErrorCode(intval(i));
             $errMessage = $xslt_processor_instance-
>getErrorMessage(intval(i));
             if($errC != NULL) { echo 'Error: '. $errC.' :'.$errMessage;}
             $xslt_processor_instance->exceptionClear();
            }
          }
         }
```

eXist's XQuery function for XSLT transformation (namely `transform:transform()`) behaves slightly different. If the processor runs into an error while processing your XSLT, eXist will display an error page with the XSLT error and the rest of the XQuery script will be skipped. In order to keep the XQuery running even after it encountered an XSLT error you have to wrap the function call in a try-catch-block. If `catch *` is triggered you can put out the error messages by accessing the variables `$err:code`, `$err:description`, and `$err:value`.

However, the problem is: Not all XSLT errors are caught in eXist. Errors of the type: `exerr:ERROR XSL transform reported error:....` are treated as expected, while errors of the type `exerr:ERROR Exception while transforming node...` are not caught at all. In the latter case, the query puts out everything which is noted in the XQuery script before the block with the `transform:transform()` function call – and nothing more. The rest of the query seems not to be processed at all (including the error message).

Aside from error handling, any XSLT wrapper should provide a convenient way to pass parameters to the XSLT script. Generally, that works with both approaches pretty well, though not completely intuitive. In PHP, you first have to create an `XdmValue` before passing it to Saxon:

```
$xdmval_for_my_param = $xslt_processor_instance->createXdmValue("example str");
$xslt_processor_instance->setParameter('my_param', $xdmval_for_my_param);
```

(For reasons I don't not know you cannot directly pass `$xslt_processor_instance->createXdmValue("example string")` as the second parameter of `setParameter` when using a string. For integers, it works that way.)

In eXist, you have to wrap all your XSLT parameters in a `<parameters>` element like:

```
let $transform_parameters :=
 <parameters>
   <param name="my_param" value="'example str'"/>
 </parameters>
```

and then pass `$transform_parameters` as a parameter to `transform:transform()`.

However, the problem in eXist is that, as of now, you can only pass string values as parameters to Saxon. (That should change once the `fn:transform` function specified in XQuery 3.1 will be implemented in eXist.)

Not a real problem, but a fact one should be aware of when using eXist for doing XSLT, is that all the options set in `xsl:output` are ignored by eXist (instead, the settings you pass as serialization parameters are taken into account).

Furthermore, there seems to be a problem in eXist's XSLT handling when loading XML documents which are stored inside eXist into XSLT variables (like: `<xsl:variable name="my_var" select="doc('foo.xml')"/>`): An XSLT transformation started in eXist returns for the XSLT command `xsl:value-of` not only text nodes (as you would expect) but also the values of XML attribute nodes.[6]

Additionally, in order to get elements in the order they are stored in your document you need a silly XSLT line like `<xsl:sort select="(count(preceding::*) + count(ancestor::*))"/>`. And furthermore, using variables of these kind significantly slows down the performance. A transformation which took 20 seconds before using a variable document of this kind, took 10 minutes after adding `$my_var` into an expression like `<xsl:for-each select="//some_element">` (thus accessing the document in the variable instead of the input document).

Speaking of performance, aside from the phenomenon described above, I did not do any thorough tests. So, I cannot assess whether Saxon runs generally faster when started from within eXist or from within PHP. However, even without having done any real measuring I'd dare to say that the difference (if there is any) is rather small.[7]

[6] See https://github.com/eXist-db/exist/issues/791 for a description of this and other XSLT issues in eXist.

[7] For the performance of Saxon/C in general, see Michael Kay and Debbie Lockett, Benchmarking XSLT Performance. Presented at XML London 2014, June 7-8th, 2014. doi:10.14337/XMLLondon14.Kay01.

**Setting up a search**

Almost every website offers its users a full text search and there is no reason why XML/XSLT based ones should be an exception. However, realizing an on-the-fly querying of XML files with X Technologies requires a little bit of work to do. (I won't discuss other approaches for setting up a website search like Apache Solr, Google Custom Search, and so on.)

As of now, in Saxon/C you cannot query XML files with pure XPath expressions. (That will change in next version of Saxon/C, however). This being the case, there is only one way to search and query XML input based on user input: creating an XSLT or XQuery file with a random string inside a select expression, turning the user input to an XPath expression or an “XPath compatible” text snippet via PHP, replacing your random string with the XPath expression just created and then running the XSLT or XQuery as string. (In the lines to follow, I will only show XSLT, though since Saxon 0.3.1, XQuery is also an option and probably an easier one.)

For example, your XSLT could look like:

```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
version="2.0">
 <xsl:template match="/">
  <h1>Search results</h1>
  <xsl:for-each select="//some_element[contains(.,'Az9HpPC')]">
   <p><xsl:value-of select="."/></p>
  </xsl:for-each>
 </xsl:template>
</xsl:stylesheet>
```

In PHP, you have to do the following:

```
$search_term = $_GET["search_term"]
$xslt_raw = file_get_contents("my_xslt_file.xsl");
$xslt_with_actual_search_term =
str_replace('Az9HpPC',$search_term,$xslt_raw);
$xslt_processor_instance = new SaxonProcessor();
$xslt_processor_instance->setSourceFile('my_xml_file.xml');
$xslt_processor_instance-
>setStylesheetContent($xslt_with_actual_search_term);
$xslt_result_string = $xslt_processor_instance->transformToString();
```

While this is feasible, it gets a little bit more complicated, if you want to offer your users an “advanced” search, i.e. combining different input fields with checkboxes, dropdown inputs, and so on. In this case, you have to put together a more complex XPath expression. In PHP, those XPath expressions will be treated as strings, so during development you will never get any hints on XPath syntax errors you may produce.

For advanced full text search options like similarity search, truncation and boolean searching you either have to build your own XPath functions or find existing libraries appropriate for your use case. Either way, there is some work to do.

In eXist, of course you can also use your own XPath functions for offering a full text search. However, there is no need to do it, since eXist has a great built-in solution for this. So, eXist can play to its strength in this category. eXist fully integrates Lucene, thus providing you and your users with a full developed query syntax. Also, there are built-in functions for ordering search results and highlighting a user's search term within a search result text block.

Configuring the indexes for the Lucene full text search in eXist is not completely intuitive, but requires a little bit of reading through the docs. It is not complicated, though.

Since your search management script will be written in XQuery, you won’t have any trouble building complex XPath expressions. You have full XPath support while writing an XQuery search script in the XML IDE of your choice.

**Long term maintenance**

Especially for humanities projects, long term availability matters. We all hope, that even in decades (not to talk about centuries) our web applications will still be there in the open web and still be running. As of now, nobody knows whether this will be the case and who should be in charge for keeping projects up. However, one thing is for sure: The faster and easier it is to keep a web application running and a website alive, the more likely it is that someone will actually do it – be it a member of the original project team or a somebody of library or computing service stuff.

So, the question here is: Which solution is more convenient when it comes to maintenance and keeping things alive?

Basically, there are two cases:

- The current software on which a web application relies needs an update (for whatever reason).
- The current software does not run anymore and is to be replaced by another software with similar functionality.

eXist is definitely more common than Saxon/C is. This being the case and keeping in mind what we have said about the installation process, eXist wins in the first scenario: It is more likely that you will get support and the necessary information on how to replace one eXist version with another than you will receive on Saxon/C. Triggering a full backup, uninstalling the current eXist version, installing a new one and restoring the backup will in most cases be rather smooth and won't cause too much trouble – unless you use eXist features which will be removed in newer versions.

However, in the second scenario, things are different. eXist has created its own XQuery dialect and thus offers much functionality which is not part of the XQuery standard and which is not even among the things for which XQuery originally was designed for, I.e. things like `request:get-cookie-value()`. It is very unlikely that there will ever be any XML database with the very same XQuery functions. So, in most cases, the end of eXist will be the end of most web applications running in it – even if the data itself will be rescued.

However, even if there will be a similar XML database with similar XQuery functions, then still, there will be a lot of manual maintenance work to do. Someone has to check not only all internal paths (like `/db/apps/foo/`) inside a web application but also every single function call, replacing eXist's function names with those of the new database and – more complicated – adapting the XQuery function parameters to the new XML database.

On the other hand, chances are, that in 30 years, there will still be web servers with PHP processors. Replacing Apache HTTP Server should be easy anyway (there are millions of instances, one can be sure that as long as there are still servers one will find instructions how to replace Apache with another web server). In case you still find an XSLT 2.0 processor which runs as a PHP extension then after installing it your job will be almost done. You do not have to go through the whole code, finding functions of a special programming language dialect like for eXist's XQuery functions. All you have to do is replacing one single block of code with another.

```
$xslt_processor_instance = new SaxonProcessor();
$xslt_processor_instance->setSourceFile(...);
$xslt_processor_instance->setStylesheetFile(...);
$xslt_processor_instance->setParameter(...);
$xslt_processor_instance->transformToString();
```

Those five lines are everything which is specific to Saxon in PHP, everything else you need for a “normal” web application in PHP is just part of the PHP standard.

So, the PHP solution wins in the second scenario.

## Conclusion

First of all, I'd like to point out that my comparison might be unfair, since eXist is much more than just a wrapper for XSLT. It offers many features in addition to what is outlined above, e.g. URL rewriting/mapping.

However, when trying to summarise the aspects discussed we get the synopsis below:

| | **eXist-db and XQuery** | **Saxon/C as PHP extension** |
|---|---|---|
| **Installation and set-up** | + Easy<br>A few things to do in the configuration file. | - Difficult and complicated.<br>+ Nothing to do in configuration. |
| **File management and data handling** | - Complete dependence on eXist's built-in tools for manipulating data, backup, and version control.<br>- Adds white space in XML data. | + Full control over files.<br>External tools may be used for version control, backup, and manipulating files. |
| **XSLT support and error handling** | - Only strings as parameters.<br>- Bug in handling documents in variables.<br>- No line numbers in error messages.<br>- Not all XSLT errors catchable. | + Full detailed error message directly in Apache's error_log.<br>+ All types of parameters.<br>+ PHP scripts never stop because of an XSLT error. |
| **Setting up a search** | + Full developed search engine with advanced syntax.<br>+ built-in functions for search result highlighting and ordering.<br>+ Native XPath support since XQuery is used. | - You have to develop your own XPath functions for search and result highlighting.<br>- No syntax help while building XPath expressions in PHP. |
| **Long term maintenance** | + More widely used.<br>- Hard to replace. | - Probably rarely used.<br>+ Easier to replace completely. |

As you can see, the PHP solution wins in the most important category: XSLT support. Among the rest, the pluses and minuses are distributed rather evenly. Depending on what you are willing to do, convenient built-in functions for setting up a search or full control over your input files might be more important for you. Personally, I'm in favour of Saxon/C as a PHP extension.

*Studia UBB Digitalia*, Volume 62 (LXII) 2017, June, Issue 1, 112-132

Published Online: 2017-06-30

DOI:10.24193/subbdigitalia.2017.1.09

# The digitisation – measuring physical and virtual presence

Liana Stanca*, Cristina Felea,
Ramona Lacurezeanu, Christian Schuster

Babeș-Bolyai University, Cluj-Napoca, Romania,
*Corresponding author: liana.stanca@econ.ubbcluj.ro

**Abstract:** The aim of the present study is to create the profile of persons who have become aware of their need to have a specific tool to move intelligently in an urban space. The research is conducted within the frame of smart cities literature and approaches cities from the perspective of the presence of Information and Communication Technologies in all areas of social, economic and political life. In this context, smart people are considered to be persons who use the advantages of technology to obtain information-knowledge and are capable to identify the need of a smart tool to assist them in various domains. A concept map is created that is interpreted as a tool to measure the physical and virtual presence of senior undergraduate students in Economic Informatics in places significant for academic information and knowledge transfer. The authors try to demonstrate the importance of combining data mining and social network analysis methods so as to carry out a complex analysis of the problem and validate by their findings important research by Fortunato, Girvan, Papadopoulos, Adedoyin-Olowe et al.

**Keywords:** smart people; behaviour pattern; concept map; knowledge; information flow; social network analysis; data mining.

## 1. Introduction

In the past decade, urban development has undergone major transformations due to the presence of information and communication technologies (ICT-s). The impact of technology on various aspects of urban development and the resulting new structures and relationships have been subject to a labelling process that attach to city

determinants such as 'informational', 'intelligent', 'hi-tech', 'tele-', 'online', 'digital', 'smart', etc. In his ground-breaking 2008 article, Hollands approaches critically the phenomenon of what he calls 'urban labelling' stating that there is no scientifically founded definition of 'smart' cities, due to the complex variables that enter discussion. The transition from 'intelligent' to 'smart' cities, he claims, should be based on more than information and communication technologies used. As a conclusion, he offers his own provisional definition whereby "smart cities can be understood as a high-tech variation of the 'entrepreneurial city'" and speculates on some general principles that would make smart cities more "progressive and inclusive". Various definitions of the concept have been proposed. For instance, Kominos, Ionita and Ilie, Paskaleva mention the concepts of "Telecities", "Cities On Line", "Intelligent cities", "Digital cities" that are used in the European Union for projects and local programs that stimulate the development of information technology at the level of online cities.

The European Smart Cities Project (started in 2007) provides rich data from a comparative analysis of 70+ cities' economic, social and geographical features to provide a hierarchy of the best places to conduct particular activities. Researchers proposed six major dimensions for generating a consistent ranking: Smart Economy, Smart People, Smart Governance, Smart Mobility, Smart Environment, and Smart Living (Giffinger et al.) focused on medium-sized cities, in an attempt to prove that these are engaged in efforts to become smart in order to compete with large metropolises. Within the study, Luxembourg is ranked first in the general hierarchy and Timisoara, Romania, is ranked 12 in the Smart People category, with position 51 in the general classification, labelled as smart city (medium smart in terms of smart mobility and people). In their paper, (Androniceanu, Ivan) drew on this project (Giffinger et al.) and tried to add more criteria needed in the further analysis of smart cities, such as transparency, responsibility, flexibility, durability, competitiveness, and benchmarking. The authors also mention that, if analysed according to the grid of Giffinger et al., Cluj-Napoca, Romania meets the conditions for a smart city, namely the number of inhabitants, public and private universities, a dynamic economy, and an IT and software centre. In their turn, Kominos, Shapiro, and Soom all consider that the six dimensions mentioned in the European Smart Cities Project report are a result/continuation of traditional regional and neoclassical urban and economic development theories such as regional competitiveness, transport and ICT economy, natural resources, human and social capital, quality of life and civic participation to city management. Lombardi considers smart cities to include participation, security/safety, and cultural heritage. Consequently, in all mentioned studies, the following major characteristics or clusters of features are mentioned: smart governance (participation), smart human capital (smart people), smart environment (natural resources), smart life conditions (quality of life), smart economy (competitiveness). In the call for papers for the 2015 Media City 5$^{th}$ International Conference and Exhibition (Plymouth University,

UK) dedicated to urban cultural studies, the organisers propose two major paradigms for approaching virtual cities. On the one hand, emphasis is put on a "data centred" change of the urban space in relation to technology, with focus on the optimisation of urban infrastructure/systems/technology to the benefit of citizens. In this view, the city should change and offer people customised/personalised services. Urban ICT-s are expected to generate new spatial-temporal participative patterns and will diminish people's participation, gradually eliminating their physical presence in favour of the virtual, which will be measured by "likes". On the other hand, a "citizen driven" approach is proposed, with unplanned, informal practices generating "counter-cultural" scenarios in which citizens are increasingly present in the transformations.

Within our research, we adopt the idea according to which the term smart city can be used only in the context of information society and its role is twofold: it supports people to perform information exchange and facilitates that exchange by ensuring the information flow infrastructure. We also draw on the European Smart Cities Project report, in an attempt to identify an instrument that would transform citizens into smart people. To start with, an investigation is made to find out if citizens have become aware of the need to have an instrument specific to a smart urban space.

The smart instrument used in this study is a concept map with urban locations, local academic and cultural poles, where events are organised for dissemination of knowledge. In our view, smart people are persons who use technology to simplify their life and assimilate knowledge and are reluctant to technology replacing human decision-making. Such citizens are open to challenges introduced by ICT-s in all areas of life, extend their presence to the virtual space and at the same time maintain their presence in the physical space.

An instrument was designed to measure the physical presence of people in a society increasingly influenced by Web 2.0 technology with the aim of detecting emergent smart features. The social features of 21$^{st}$ century Romanian academic culture are targeted by means of measuring the intensity of individuals' virtual and physical presence during a time interval that pre-dates the smart era.

Our study emphasises that there are people who are able to identify and become aware of the lack of instruments needed to minimise the time for information transmission/exchange; they are also adepts of technology used to simplify their life instead of depending entirely on it. An example of the way smart governance can react to and solve customers' needs is described in the 2010 report on Customer Journey Mapping, during which the manner in which customers are treated during the interaction process attached to various services is recounted. The Journey Mapping report proves to be an extremely useful tool for the description of the customers' experience of services, including thought processes and reactions. Based on this unsupervised learning experience of the studied group, company customer services can be improved, and

isolated, badly integrated, or duplicated systems can be identified and corrected. In another research study by Lindström, the concept of personas is used as an instrument to identify the requirements and needs of users by collecting data and creating loyalty programmes. According to Cooper et al., in a smart city "a persona is a user model that is represented as a specific set of characteristics for named individuals. These personas are not actual people, but are archetypes: a set of characteristics that have been constructed based on direct observations of the needs and wants of real people." In such studies, a person encapsulates a distinct set of behaviour patterns in relation to a certain product. These patterns are identified by means of interview data and further supported, if the case, by additional quantitative data. The persona concept plays a crucial role in identification of target groups with similar objectives and expectations towards a particular product or service (Lindström). In our study, we will use the concepts of personas and Journey Mapping to measure students' one-to-one interaction to the locations where academic culture events take place and where they are supposed to acquire and develop their academic culture skills. Thus, the paper is written within the frame of smart city oriented towards smart people capable to act "beyond the 'like' button and to take responsibility for the future shape of the city", more particularly university students. It aims to create students' profiles based on their physical and virtual presence, which allows for the computation of the "importance"- "centrality" of the locations-spaces-buildings where they get their academic knowledge and improve their academic skills.

Local interest and preoccupation for the transition to smart cities dates back to 2012-2013, when academics researched and proposed to the private/public sectors the implementation of ICT-s to increase traffic safety, drivers' comfort and increased mobility due to assistive programmes on smart mobile devices (Varga and Nedevschi, Brehar and Nedevschi, Petrovai et al.). Our current research takes the initiative a bit further because it aims to provide an answer as to the readiness of individuals, namely university students, to take a further step towards a smart urban space. The need to map physical presence and information/knowledge collecting flows prompted the creation of a concept map tool. In accordance to the standards of Palmer et al., the instrument can be considered

1. an adequate descriptive model for determining polarisation in smart cities;
2. a prototype exercise of mapping locations and routes destined to interaction with information flows within physical and virtual space. Based on this prototype, a platform can be set to highlight the relevant links between city cultural locations, information/knowledge and events, suggesting key poles for economic culture promoters;
3. to allow for the identification of the deficiencies of the existing standard initiatives;
4. to clarify the economic data, information and knowledge flows between the major pillars/promoting cultural centres in a large university centre.

## 2. Material and methods

### *2.1. Research Objective and Variables*

The main objective of the study is to investigate if persons with medium to high education are ready to identify the need for an intelligent tool meant to guide them in exploring physical locations to obtain information flows in a short period of time.

The derived objectives of the study are: (1) to create a prototype of the persons who identified the existing need; these are persons whose movement in a physical and virtual space can be measured in a social and professional context; (2) to validate a concept map tool by using a complex combination of statistical methods and data mining, which then allow for application of social network analysis to research patterns of behaviour towards the topics of interest; (3) to validate the findings in the studies by Fortunato, Girvan / Newman, Papadopoulos et al., and Adedoyin-Olowe et al., by investigating if the length of the inter-node pairs can be calculated based on the measurements done in a recommendation system that can be used in hierarchical clustering, more particularly we focus on the network connections shared by two nodes.

### *2.2. Data Collection*

Data was collected for a number of 43 senior undergraduate students enrolled in the course of Economic Informatics. Firstly, a brainstorming session was conducted with 10 students to make clear the information needed for the experiment. The answers confirmed the initial research idea, namely that students would be highly interested in concept mapping guidance to reach the locations and events where they acquire academic information-knowledge. The following information was obtained:

1. novice students would appreciate/benefit from having physical locations and routes mapped;
2. the mapping should be done by experts, namely academics and faculty;
3. the available online maps did not offer the necessary information, resulting in waste of time and confusing information;
4. their presence in the virtual world consists of social networking to stay in touch with friends and colleagues, the Moodle platform to access learning materials and virtual libraries subscribed by the university to do their tasks;
5. the average time spent daily to collect specialist information is about 2-3 hours and the average time spent daily on the social networks is about 4-6 hours.

The second step consisted of designing a questionnaire to collect data on the type of information a student can obtain by physical participation to events taking place in the most important academic cultural centres in Cluj. The main objective was to verify whether (1) students in Economics in their final year identify correctly the economic

academic culture centres and (2) the type of presence (physical versus virtual) student prefer in order to acquire knowledge and develop their academic knowledge. Their answers were expected to configure a concept map of cultural centres of vital importance for students' academic development. The questionnaire format was adopted in view of eliminating the stress factor generated by the unknown, in this case the mapping activity.

Finally, the questionnaire-map was emailed by Facebook to 100 students asked to answer within a day. The response rate was 43%. This procedure was also meant to collect data on the virtual presence of students (on the social networking site) and diminish potential distortion of answers caused by teacher's presence. The evaluation consisted of [0,1] values.

### *2.3. The Analysis Tool*

The current research is situated within the framework of empirical research and scientific reflection aimed at finding ways to measure the degree of readiness of smart citizens to correctly adopt the new organization of the real world so that they use technology in their benefit and are not being "used" by technology.

As mentioned earlier, the preliminary data collection was organised as a survey. The questionnaire consisted of two sections: one with demographic data and one for selection of locations as academic culture poles (University, faculties, university and county libraries, etc). For each location, a list of events (training, internships, workshops, etc.) and type of information-knowledge (economic, sociological, psychological and educational, etc) to be acquired were attached. Each person had to choose a value ranging between [0,1]. As this was a dichotomous data processing, the statistical method used followed the steps described below:

1. The data statistical analysis was applied to highlight the features of the study participants.

2. A test followed to confirm if the items of the physical presence measuring tool have one dimension, namely one latent factor. This basic rule is known as unidimensionality (Dan) and is connected to another assumption, namely the item local independence (Dan). Consistent with the proposition that a subject's performance for an item can be predicted and explained by the existence of a latent factor or trait, the ensuing relationship between this performance and the latent factors can be described by a monotone increasing function called item response curve or item characteristic curve (Hambleton et al.). This type of analysis estimates the reliability of the instrument by measuring the elements' internal consistency when they correlate well. The reliability was calculated in accordance to van Soom's theory, which states that a measuring

instrument is perfectly reliable if it has a perfect positive effect (r = 1). The interpretation of results was done in compliance with specialist literature (Popa et al.) whereby a scale/test is more reliable when the reliability score is higher than 0.80 (Dan). This was the state when we decided if our instrument has a high degree of reliability/consistency so that the results can be maintained in time and it can be used for similar scenarios in the future.

3. The next step was aimed at finding out the correlating variables. In specialist literature, Hambleton and Jones state that unidimensionality refers to the existence not of only one dimension but of a dominant dimension which can influence test performance, which is called ability or, more generally, latent factor coverage (Dan). In the case of a test measuring presence, we cannot expect the items to cover the pure presence.

4. The principal component analysis (PCA) is conducted to reveal pre-existent structures in a multitude of multivariate data (Culic). These structures are generally expressed by means of variance and covariance of variables and of similarities and dissimilarities between objects. In this sub-stage, we identified the values and vectors of a square matrix obtained by multiplying matrix data (objects) with matrix anti image, a relatively common procedure in matrix calculus. The null hypothesis is that there is no limited set of factors that can determine the presence of citizens in a smart city. The alternate hypothesis is: there is a limited set of factors that can determine the presence of citizens in a smart city. PCA was used to build a decision model based on the answers to the questionnaire/concept map. By correlating these data, a pattern can be formulated to reveal the limited factors that explain the variation of the studied variables. Within this method, the unrepresentative correlations (<0.3 from correlation matrix) were eliminated. Other variables were eliminated in the next step based on Kaiser-Meyer-Olkin measurements (the values of the covariation indices from the principal diagonal of the anti image >0.5).

5. The hierarchical method was applied to determine the number of clusters, followed by MacQueen's k-means improved by E. Diday's dynamic clouds method (Enachescu). The k-means algorithm was used to generate cluster with objects (persons) who share the same behaviour within the research. The hypothesis was: there are differences between the persons' preferences to visit certain locations to get knowledge. We validated the result by applying the hierarchical k-means, which is a top-down hierarchical clustering method using k-means iteratively with k=2. This decision was taken in order to (1) determine if similar results can be obtained by using the two methods and (2) determine the existence of potential communities who, related to space, would allow for the idea of graph, in other words would allow us to perform social network analysis.

6. In order to reach our purpose, we set on to determine a pattern of present persons by means of applying classification algorithms characteristic to the studied context. The Receiver Operating Characteristic (ROC) Regression analysis was used to validate the result given by the classification algorithms (Enachescu). The analysis model completed aimed to outline the profile of active persons. We employed ROC classifier to test the validation of the developed classification model. Up to this phase, we created a statistical analysis oriented towards the individual behaviour that was meant to demonstrate that a community was created based on our instrument. The ROC classifier helped in the measuring of the smart persons' physical presence in space and similar behaviour patterns.

7. In the final part of the analysis, we set out to determine the places that have the highest concentration of smart persons by means of the social network analysis, particularly the calculation of centrality. Centrality is considered the most often used conceptual tool for exploring the actors' (locations) roles within social networks. The centrality degree of a node in an undirected graph is defined as the number of nodes connected to that node. Consistent with (Wasserman, Faust), "central locations have to be more active in the sense that they have the highest number of connections with other actors-locations in the network." According to the same authors, closeness centrality measures the closeness of a node to the other. Consistent with (Freeman), "a node is the sum of geodesics distance from all other nodes, which is defined as the length of the shortest path from one node to another". Consequently, closeness centrality can be used to determine the degree of influence a node has in the network. Another interpretation belongs to (Borgatti), who states that betweenness gives "the share of times that a node j needs a node t (whose centrality is being measured) in order to reach j via the shortest path". In this research, we adopted Everett and Borgatti's theory of nodes in a network, according to which (Everett, Borgatti)

- Group Centrality: the centrality degree of a group is the number of actors outside the group that are connected to the member of that group. Different ties to the same actors by different group members are only counted once.
- Group closeness centrality is defined as the normalized inverse sum of distances from the group to all nodes outside the group.
- Group betweenness centrality shows the "proportion of geodesics connecting pairs of non-group members that pass through the group. (Chaoqun et al.)

These three centrality indicators allowed us to measure the role of locations in the personas' participation to certain events in order to obtain knowledge. The results of our research will meet the results of (Chaoqun et al).

### 3. Study Findings

The present study is conducted along two dimensions, with their respective hypotheses. The first hypotheses for the first dimension are:

H0. The subjects are not ready to be interpreted as smart persons.

H1. The subjects are ready to be interpreted as smart persons.

For the second dimension, the hypotheses are as follows:

H0. The behaviour towards the locations does not generate communities that can act as networks. More exactly, by using hierarchical clustering, we will not obtain communities which, related to the physical space, form a graph, namely a network.

H1. The behaviour towards the locations generates communities that can act as networks. More exactly, by using hierarchical clustering, we will obtain communities which, related to the physical space, form a graph, namely a network.

The descriptive analysis allowed us to create a profile of the participants, with the following characteristics: age: 18-24 years; urban: 90.7%; rural: 9.7%; gender: 72.1% females; 27.9 males; concept map knowledge: 46.5 know about it; 53.5 do not know about it.

Next a reliability-validity analysis was performed to obtain an instrument to measure the extent to which students in economics get involved in academic culture life. This was done in a context in which certain activities (for example, learning, information seeking, training, etc) performed traditionally in a physical environment are increasingly replaced by their counterparts in the virtual environment, resulting in a diminished participation to real world events. The instrument measures students' presence in physical locations at events that occur regularly in the life of a person studying economics in Cluj-Napoca and have been traditionally associated to major sources of academic culture. Within this context, the instrument was used to measure if the locations traditionally considered as poles of student life and the associated culture generating events have maintained their place in a society increasingly governed by the virtual dimension. The results of the analysis are:

1. there are differences between the evaluations of the need for students' physical presence in certain location and at events designed for their field of study;

2. the instrument for the analysis of persons' presence in the physical space needs not be reorganised;

3. the instrument items used to measure the physical presence/absence are highly reliable. Following the application of the test, we obtained that the test is sensitive to the measured characteristics: alpha Cronbach =0.915. Mean=2.13663 Std.Dv.=3.29576, Number of items in scale is 14, Valid N:344. The alpha value obtained is significant and suggests that the instrument is adequate to the purpose. It follows that

the test is unidimensional. Within the process of determining the alpha coefficient, the values of the test item discrimination coefficient are determined. They are presented and interpreted as follows: the students (specialists) manifested a moderate to high agreement in appreciation of the presence/absence of the persons to local academic culture life. Average Inter-Item Correlation: 0.697.

The analysis continued with the measuring of the test internal consistency, based on the split-half method. The results are good: Cronbach alpha, full scale: 0.915 Corr. 1st & 2nd half: 0.948 Split-half reliability: 0.973 Guttman split-half: 0.94. Notice should be made that test reliability is easily modifiable. Thus, in accordance with reliability coefficient calculation, the number of items is adequate, so the number of questions need not be increased. Yet, if this were the case, the reliability would not be affected significantly. In conclusion, the test has a good reliability and that means that the questions inclined to be a whole, meaning that the halves of the instrument (locations and events) could be used on their own to measure students' physical presence. Because of the alpha Cronbach =0.915, it is recommendable that the instrument be used as it was designed. In a parallel test developed with similar elements, students' relative scores would show a slight difference.

Based on the ANOVA test, whose results are presented in the table below, the study hypothesis is unilateral and namely there are differences between items generated by students' answers, meaning that the test can differentiate students in accordance to their physical presence to the respective locations and events.

| Analysis of Variance | | | | | |
|---|---|---|---|---|---|
| | **Sums of** | **Df** | **Mean** | **F** | **p** |
| **Between Subjects** | 266.1204 | 343 | 0.775861 | | |
| **Within Subjects** | 356.7066 | 4472 | 0.079764 | | |
| **Between Items** | 62.7137 | 13 | 4.824128 | 73.16769 | 0.00 |
| **Residual** | 293.9930 | 4459 | 0.065932 | | |
| **Total** | 622.8270 | 4815 | | | |

*Table 1. Anova test results*

The analysis continued with Principal Component Analysis (PCA) by which we tested if our instrument has latent factors, more precisely if there are central and marginal items or they form a whole that can be measured along the same dimension. As known, PCA provides various results such as descriptive analysis, the correlation matrix, KMO and Bartlett test, communality, total variance, screen plot image, component matrix and rotated component matrix.

The first principal eigenvalue captures 28.93% of the variability in the data.

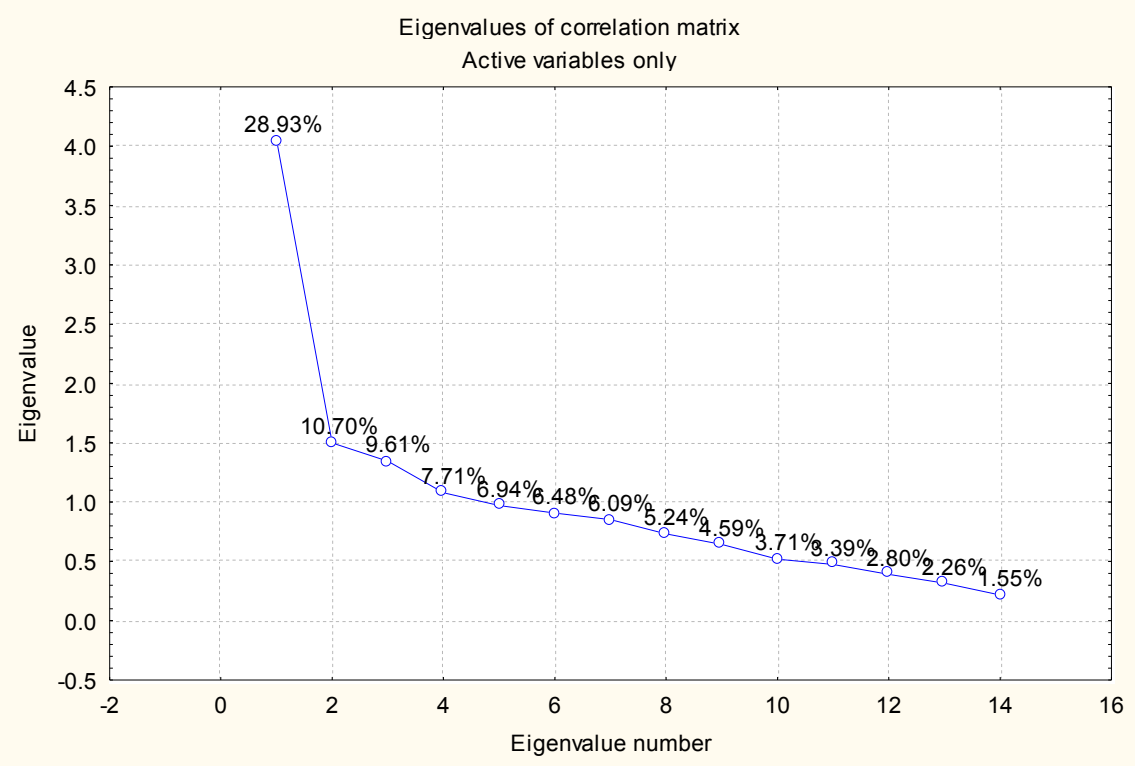



*Figure 1. Screen Plot*

It is known that coefficients with values higher than 0.6 indicate a close connection between factors and variables. Thus:

1. the first factor may suggest students' interest for self-development and includes the variables: Event CourseLab (0.687); Training (0.569), Workshops (0.405), C1.economic (0.758), C1.linguistic (0.652), C1.psychological and educational (0.550), C1.communication 0.819.

2. the second factor may suggest students' interest for professional development the variables: C1.internships (0.581); C1.political (0.518), C1.sociological (0.490).

According to communality results and calculation of the square cosines, values are obtained between [0.268, 0.822], close to 1, so we can conclude that the information is well-preserved in the studied physical space.

This method was used to eliminate items with correlations less than 0.3 but such items did not exist. Factor analysis is adequate for the current study, which is a decision taken based on the results of the Kaiser-Meyer-Olkin test (used to measure the extent to which data variation is caused by the instrument). The value of the Kaiser-Meyer-Olkin test is 0.745, which shows that factor analysis is indicated in this case. The results of the Bartlett test p=0.000 confirm that the analysis is useful on the given data. Consistent with Kaiser's criterion (communality values have to be higher than

0.4), (Culic) the latent root consists of studied items. Consequently, the null hypothesis is rejected and the alternate is accepted, according to which there is a limited set of items which determine the validity of the instrument in a ratio of 50.76%. Additionally, the second null hypothesis is rejected and the alternative accepted, according to which in the study there are moderately correlated items.

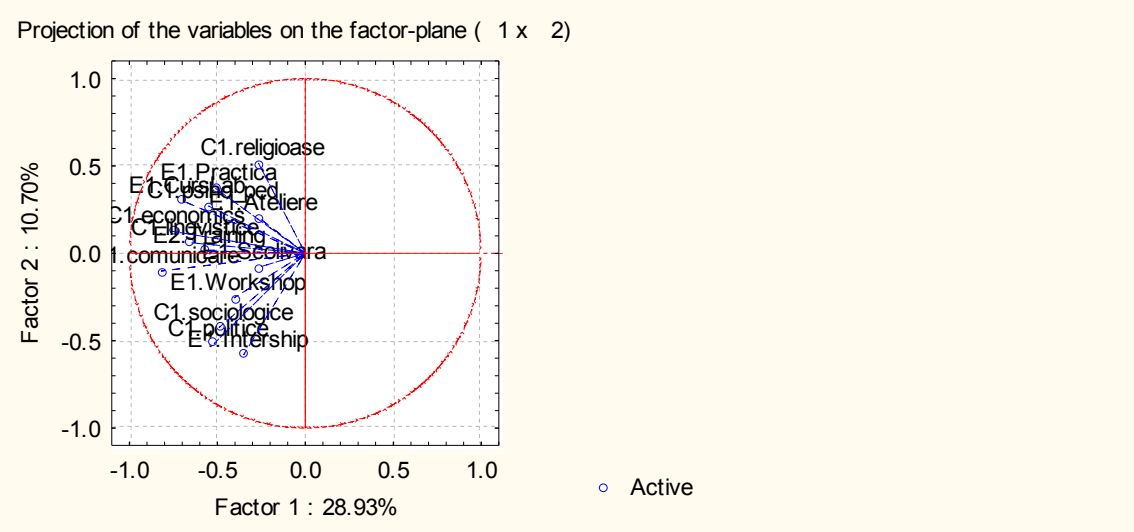



*Figure 2. Projection of the variabiles*

Next, we applied hierarchical clustering. The hypothesis for this step is that there are differences between personas' evaluation as recognition of the importance of their presence in the physical spaces to obtain information/knowledge. The results of applying the k-means algorithm show that students can be grouped into two clusters characterised by the fact that the value of inter-class inertia is significantly higher than the value of intra-class inertia. The clusters resulted from the way students indicated their presence in the specific locations at the events and for the purpose of certain knowledge transfer activities.

Cluster 1 includes the preferences of students with a low presence and no high fluctuations in their physical presence at academic culture activities intended for self-development (F=65.442; p=0.000).

Cluster 2 includes the preferences of students with a high presence and a fluctuating behaviour in activities intended for their professional development (F= 39.869; p=0.000).

Next, we applied the Manova test to assess the following hypothesis:

H0: There is no difference between the factors studied for the 2 groups of students distributed according to their participation in the specific locations.

H1: There are differences between the factors studied for the 2 groups of students distributed according to their participation in the specific locations.

As a result of the Manova test, the null hypothesis was rejected and the alternative hypothesis was accepted: Wilks' Lambda (F=1.512, p=0.000). Cluster analysis produced student clusters which show inherently similar behaviour patterns. The students with good physical presence are grouped in a cluster, being considered a main reference in the process of improving the characteristics of the other students in the same cluster. The graph of the two clusters is presented below:

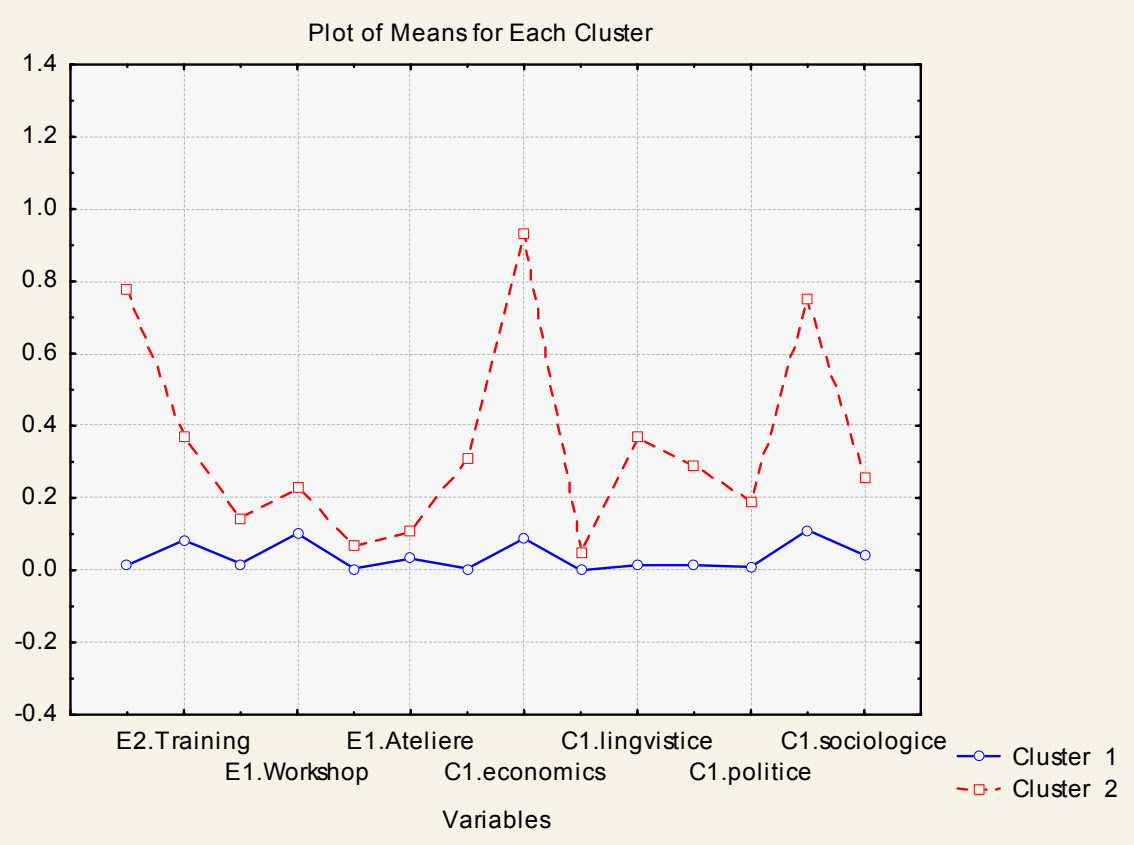



*Figure 3. Physical presence clusters*

The study continued by applying ROC curve to find out if the studied attributes are predictive for the model of determining the students' physical presence in the academic culture life of the city.

The area below the curve is 0.986, p=0.001; 95% CI (0.978;0.999). This is a discriminating model, eliminating the false negatives and the false positives for 92% of the cases.

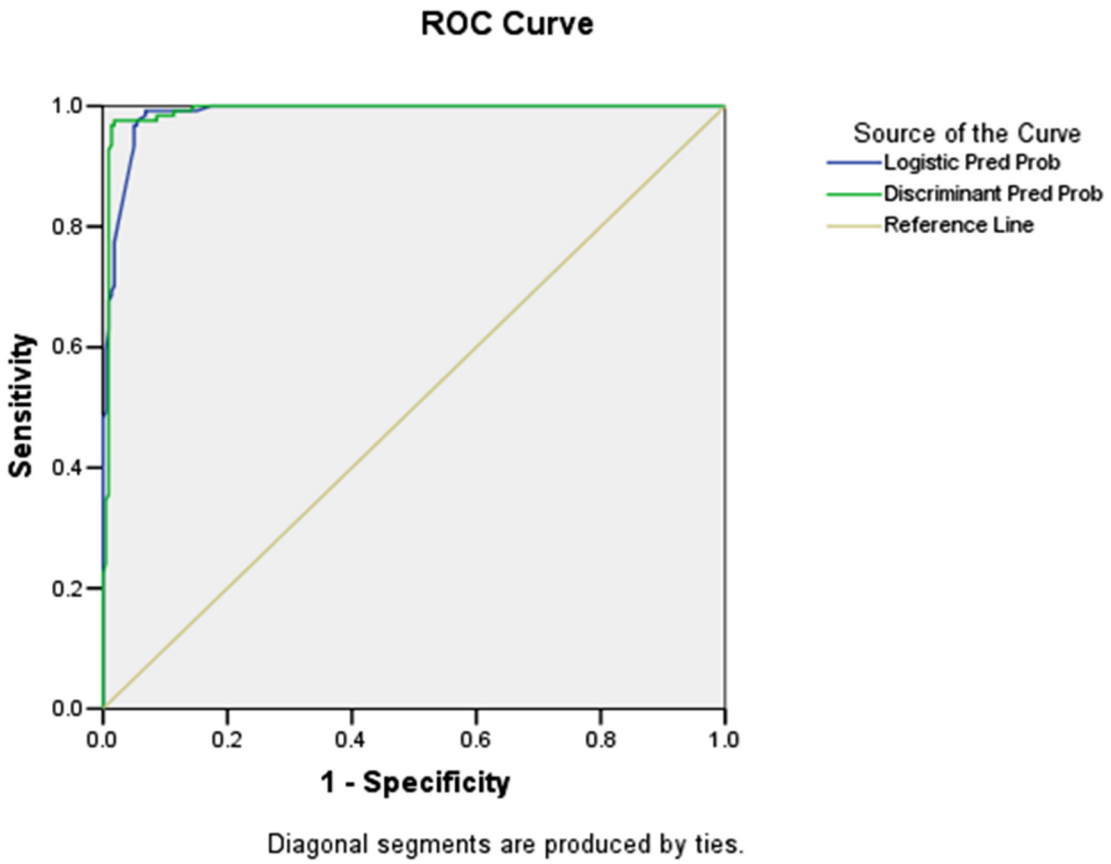



*Figure 4. ROC Curve*

Thus, the null hypothesis of the study is rejected and the alternate hypothesis is accepted, according to which the subjects under study are ready to be interpreted as smart persons.

Moreover, two behaviour clusters were obtained, which may indicate that these persons can also be interpreted as same behaviour communities, representing a graph. Consequently, social network analysis can be applied to determine what the most important locations are, namely the poles where knowledge transfer is made by means of events.

In order to establish the presence/absence of the poles, we set out to determine the values of the graph nodes to obtain a classification in accordance with the frequency of use of the respective locations. In the specialist literature, there are two approaches to measuring the “importance” of a node: – the importance of the nodes is established in accordance to the importance level of the information flow that passes through the nodes (Bonacich); – the nodes represent a concentration of information groups with the corollary that the network is an image of the data groupings based on features studied in the research (Borgatti).

Our approach draws on the second author (Borgatti). This means that we focussed on the term “importance” with the aim of determining the role presence and localisation have in information-knowledge transfer. In order to carry out this stage, we used the classical topological measuring methods, more specifically we calculated the degree of centrality (Nieminen), (Freeman), the closeness centrality (Freeman), (Sabidussi), (Wasserman, Faust), the betweenness centrality (Freeman) and the eigenvector centrality (Latora, Marchiori). The theoretical background for the indicators is presented in detail in (Nieminen), (Freeman), (Sabidussi), (Carrington, Scott, Wasserman) and (Latora, Marchiori). Based on these indicators (Cadini, Petrescu), data were processed to obtain a topological network of importance.

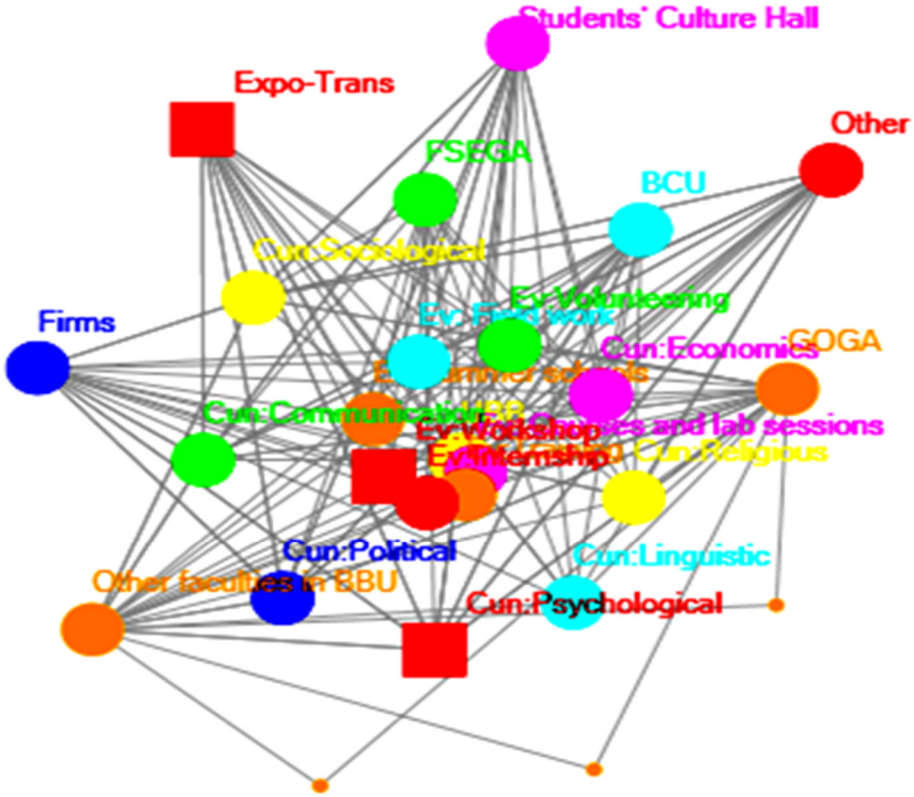



*Figure 5. Closeness Centrality of location*

Each “edge” represents a link between two locations and events, respectively information/knowledge acquired.

The edges represent different types of connections that can be created during the information/knowledge transfer occurring at those specific places. A “tie” edge is created in the moment when a user visits the location to be subject to information/ knowledge transfer during an event.

| Vertex | Degree | Betweenness Centrality | Closeness Centrality | Eigenvector Centrality |
|---|---|---|---|---|
| BBU | 21 | 1.000 | 1.276 | 0.285 |
| Ev: Courses and lab sessions | 6 | 0.027 | 1.897 | 0.084 |
| Ev:Training | 6 | 0.027 | 1.897 | 0.084 |
| Ev:Internship | 9 | 0.081 | 1.690 | 0.123 |
| Ev:Workshop | 9 | 0.081 | 1.690 | 0.123 |
| Ev:Summer schools | 9 | 0.081 | 1.690 | 0.123 |
| Ev: Field work | 9 | 0.081 | 1.690 | 0.123 |
| Ev:Volunteering | 9 | 0.081 | 1.690 | 0.123 |
| Know:Economics | 9 | 0.081 | 1.690 | 0.123 |
| Know:Religious | 9 | 0.081 | 1.690 | 0.123 |
| Know:Linguistic | 9 | 0.081 | 1.690 | 0.123 |
| Know:Psychological | 9 | 0.081 | 1.690 | 0.123 |
| Know:Political | 9 | 0.081 | 1.690 | 0.123 |
| Know:Communication | 9 | 0.081 | 1.690 | 0.123 |
| Know:Sociological | 9 | 0.081 | 1.690 | 0.123 |
| ……. | …. | … | … | .. |

*Table 2: Topological and Reliability Centrality Measures*

Centrality is considered one of the most important and frequently used conceptual tools for assigning roles to the subjects studied in social networks. In the current study, between's centrality is approached as measure of the degree of connection between locations in a network for information-knowledge transfer in physical space. The findings indicate that the studied locations act as ties between events and information-knowledge.

Finally, based on the centrality measuring, we explored the place of each location in relation to information-knowledge transfer within specific events. The values Graph Density - 0.42, Maximum Geodesic Distance (Diameter) - 3, Average Geodesic Distance - 1.59, show that not all nodes are interconnected. The basic structure is non-hierarchical, within which all locations play an equal role in information-knowledge transfer taking place during specific events. The resulting network facilitates common comprehension, a knowledge centre and support for students' information-knowledge acquisition.

## 4. Discussion

The current paper proposes a prototype tool for measuring the presence of individuals in the physical and virtual space. The objective was to look into the theory according to which in smart cities, people's physical presence will diminish and be

replaced by a ‘like’. The participants to the study were senior undergraduate students of Economic Informatics. The need for a concept map was validated, then a survey based on its elements collected the data. The main null hypotheses of the study were rejected and the alternate are accepted. According to these:

1. the participants in the study are ready to be interpreted as smart persons, more exactly as persons who use technology to simplify their life and to obtain an adequate information flow in a short time.

2. Behaviour towards the specific locations generates communities which in relation to the physical space create a graph, a network and, by extrapolation, a social network. The latter can be analysed using specific methods to determine poles that attract concentrations of persons. Consequently, problems in which the variables studied in relation to space can be demonstrated to generate graphs may be solved based on data mining and social network analysis so as to obtain more rigorous results.

Consequently, the profile of the smart person obtained can be argued to deny the theory according to which urban ICT-s may diminish people’s physical presence in favour of virtual presence measured by ‘like’-s. In our view, smart persons are much more focussed on using both sides of reality to achieve their purpose.

3. The results of the final part of our research validates the findings of(Freeman), (Stanca and Felea), (Palmer, Alfi and Dam). Two clusters were obtained that could be interpreted as communities sharing the same behaviour. By their physical presence at the specific locations, the members of the two communities can be represented as in (Adedoyin-Olowe, Gaber and Stahl) by a graph. The hierarchical clustering methods helped us to identify the poles of academic culture by the number of persons visiting them to acquire information-knowledge. These poles were validated in relation to real physical space. Thus, findings indicate that the length of the inter-node pairs can be calculated based on the measurements done in a recommendation system that can be used in hierarchical clustering, more particularly by focussing on the network connections shared by two nodes.

The intermediate results that lead to major results are:

1. The concept map is a valid tool and can be used to guide students through the most important routes for extracting academic information-knowledge. It is also a useful tool to reduce the time for offering and finding information. If generalised at the level of student population, it could also be used for obtaining information about the student interactions with the physical environment or about their routes.

2. The findings of multivariate analysis suggest that the proximity of association between locations, events and type of information-knowledge transfer reveal that the respective individuals have chosen to participate socially to the economic space

created. Values also suggest that students are still adepts of physical participation to the events organised with the aim of economic information-knowledge transfer within academic environment.

3. The study suggests that locations are visited/chosen as a consequence of the type of events and information-knowledge transfer taking place; this may suggest students' interest in the contents of the events and probable benefit for their self and professional development.

4. The findings confirm that there is a significant group of students who have intense physical presence but there is also a group who show increased virtual presence and diminished physical presence.

## 5. Conclusions

The study demonstrates that, in order to become a smart city, an important East-European city needs people who are ready to undertake the transition but tools need to be devised to measure readiness so as to avoid issues caused by cultural shock. The current findings indicate that 18 to 26 year old people (students) still need to be physically present in locations where information-knowledge transfer is performed. Moreover, in spite of being in their final undergraduate year, these students still need guidance to define this physical presence for their personal and professional development. On the other hand, students' virtual presence has been noted on social networks and e-learning platforms. However, the current study confirms the findings of previous research (started in 2009) where students that use ICT-s as a subsidiary were noticed to use learning platforms only under constraint. Currently, though, mature students knowledgeable in economics and ICT-s are aware of technology utility and it is our opinion that citizens should be supported in the process of transition so as a too abrupt a change to a smart city should not generate passive citizens adepts of "like"-s.

Future research will extend the study to junior undergraduates in economic studies to examine if the education level plays a role in the transformation in smart people. Experiments could be approached from the angle of dynamic decisions theory to determine if learning the environment has an effect on completing a route. In practical terms, the authors intend to design a virtual application to assist the students to discover the cultural space in Cluj so that they may extract the information-knowledge they need in the shortest time possible.

## Works Cited

A. Androniceanu, M. Ivan. "„Smart City" - A Challenge For The Development Of The Cooperation Mechanism Between European Cities." Proceedings of Administration and Public Management International Conference, Issue: 8 (2012), pp. 333-346.

M. Adedoyin-Olowe, M. M. Gaber, F. Stahl. “A Survey of Data Mining Techniques for Social Network Analysis.” jdmdh.episciences.org/18.pdf, last accessed 2014.

P. Bonacich. "Power and Centrality: A Family of Measures". American Journal of Sociology (University of Chicago Press) 92:(1987), pp.1170–1182

S. P. Borgatti. “Centrality and network flow.” Social Networks, 27(1) (2005), pp. 55-71.

R. Brehar, S. Nedevschi. “Scan Window Based Pedestrian Recognition Methods Improvement by Search Space and Scale Reduction.” Proceedings Intelligent Vehicles Symposium, Dearborn, USA, 2014, pp. 529-534.

F. Cadini, E. Zio, C-A. Petrescu. “Using centrality measures to rank the importance of the components of a complex network infrastructure.” Critical information infrastructure security, Springer Berlin Heidelberg, 2009, pp. 155-167.

P. J. Carrington, J. Scott, S. Wasserman. “Models and Methods in Social Network Analysis.” New York: Cambridge University Press, 2005

A. Cooper, R. Reimann, D. Cronin. “About Face 3: The Essentials of Interaction Design.” John Wiley & Sons, 2007.

I. Culic. “Metode avansate în cercetarea social - Analiza multivariată de interdependență.” Editura Polirom, 2004.

C. O. Dan. “Caracteristici ale modelelor de evaluare psihologică bazate pe teoria răspunsului la Item.” www.academia.edu/5218682/Characteristics_of_psychological_assessment_models_based_of_Item_Response_Theory_Romanian, 2012, last accessed 2014.

D. Enachescu. “Data maining, Metode și aplicații.” Editura Academiei Române, București, 2009.

M. Everett, S. P. Borgatti. “Extending centrality, Models and Methods in Social Network Analysis.” Cambridge University Press, Cambridge, UK, 2005, pp. 57–76.

C. Felea, L. Stanca. “Wikis in Teaching English for Academic Purposes – Engaging Romanian Students in a 21st Century Learning Environment. eLSE, elearning and Software for Education.” April 28-29, Bucharest. Editura Universitară, (2011) 460-467.

C. Felea, L. Stanca. “Wiki Tools in Teaching English for Academic Purposes - An Analysis of Student Learning Behaviour Pattern.” “CAROL I” National Defence University Publishing House. Vol. 1, (2013) 82-91.

C. Felea, L. Stanca. “Wiki Tools in Teaching English for Academic Purposes - IT and Language Proficiency as Predictors of Online Participation.” “CAROL I” National Defence University Publishing House, Vol 3, (2013) 312-321.

C. Felea, L. Stanca. “Wiki Tools and English for Academic Purposes-Fostering Collaborative and Autonomous Learning in Higher Education. Revista de Informatică Socială, No. 14, (2010) 55-65.

S. Fortunato. “Community detection in graphs.” Physics Reports 486(3) (2010), pp. 75-174.

L.C. Freeman. “Centrality in social networks conceptual clarification.” Social Networks, 1(3) (1979), pp. 215-239.

M. Girvan, M.E. Newman. “Community structure in social and biological networks.” Proceedings of the National Academy of Sciences, 99(12) 2002, pp. 7821-7826.

R. Giffinger, C. Fertner, H. Kramar, R. Kalasek, N. Pichler-Milanovic, E. Meijers. “Smart Cities Final Report.” Vienne: Centre of Regional Science. Retrieved 12 May, 2012, from: www.smart-cities.eu/download/smart_cities_final _report.pdf, last accessed 2014.

R. Hambleton, H. Swaminathan, J.H. Rogers. “Fundamentals of Item Response Theory.” London: Sage Publications Inc. 1991.

R. K. Hambleton, R. W., Jones. “Comparison of classical test theory and item response theory and their applications to test development.” Educational Measurement: Issues and Practice, 12(3) (1993), pp. 38-47.

R. G. Hollands. “Will the real smart city please stand up?” City, Taylor & Francis Journals, vol. 12(3) (2008), pp. 303-320.

N. Kominos. “Intelligent cities: innovation, knowledge, systems and digital spaces.” London. Spon Press. 2002.

A. Ionita, R. Ilie. “Conceptul de Oras On-Line.” on-line: rria.ici.ro/ria2000_2 /art05.html, last accessed 2014.

S. Wasserman, K. Faust. “Social network analysis: Methods and applications.” New York: Cambridge University Press, 1994.

V. Latora, M. Marchiori, “A Measure of Centrality Based on the Network Efficiency”, New Journal of Physics 9, 2007,pp.188-197

M. Lindström. “Report-Making customer group real-using Personas.” Smart Cities Brief, No.10, 31 January 2011, www.smartcities.info/files/365_42759_Smart_Cities_Research_Brief_Using_Personas.pdf accessed November 25, 2014. last accessed 2014

P. Lombardi. “New challenges in the evaluation of Smart Cities.” Multi-sector, Network Industries Quarterly, vol. 13, no. 3, 2011, http://newsletter.epfl.ch/mir/index.php?module=epflfiles&func=getFile&fid=241&inline=1, last accessed 2014.

J. Nieminen, “On Centrality in a Graph. Scandinavian”, Journal of Psychology, n.15, 1974, pp. 322-336.

Chaoqun Ni, C. R. Sugimoto, J. Jiang. “Degree, Closeness, and Betweenness: Application of group centrality measurements to explore macro-disciplinary evolution diachronically.”
elektrokomponenten.ch/media/files/09e4150bd20e58c913000000.pdf, last accessed 2014.

D. Palmer, S. Alfi, K. van Dam. “Mapping Smart City Standards- Based on a data flow model.” www.bsigroup.com/LocalFiles/en-GB/smart-cities/resources/BSI-smart-cities-report-Mapping-Smart-City-Standards-UK-EN.pdf, last accessed 2014.

S. Papadopoulos, Y. Kompatsiaris, A. Vakali, P. Spyridonos. “Community detection in social media.” Data Mining and Knowledge Discovery. 24(3) (2012), pp. 515-554.

K. Paskaleva. “Enabling the smart city: The progress of e-city governance in Europe.” International Journal of Innovation and Regional Development. No. 1 (2009), pp. 405-422.

A. Petrovai, A. Costea, F. Oniga, S. Nedevschi. “Obstacle detection using stereovision for Android-based mobile devices.” IEEE International Conference on Intelligent Computer Communication and Processing, Romania, (2014), pp. 141-148.

N. Popa, L. Antonesei, A. V. Labăr. “Ghid pentru cercetarea educației, Editura Polirom-Collegium, 2009.

G. Sabidussi. “The centrality index of a graph.” Psychometrika 31: (1966), pp. 581-603.

J. M. Shapiro. “Smart cities: quality of life, productivity, and the growth effects of human capital.” The Review of Economics and Statistics, 88 (2) (2008), pp. 324-335.

E. van Soom. “Measuring levels of supply and demand for e-services and e-government: a toolkit for cities.” Smart Cities Research Brief, N. 3:2009. <http://www.smartcities.info/files/Smart_Cities_Research_Brief_Measuring_levels_of_supply_and_demand_for_eservices_and_egov.pdf>, accessed November 25, 2014, last accessed 2014.

C. Spearman. “General Intelligence.” American Journal of Psychology.15 (1904), pp. 201-293.

R. Varga, S. Nedevschi. “Label transfer by Measuring Compactness.” IEEE Transactions on Image Processing, Vol.22, 2013, pp. 201-210.

Report, “Customer Journey Mapping.” Smart Cities Brief, No.12, 2010 http://www.smartcities.info/files/Smart_Cities_Brief_Guide_to_Customer_Journey_Mapping.pdf>, last accessed 2014.

*Studia UBB* Digitalia, Volume 62 (LXII) 2017, June, Issue 1, 133-137

Published Online: 2017-06-30

Review

# Bringing Bosch alive through multimedia

Voica Puşcaşiu

DigiHUBB, Babeş-Bolyai University, Cluj-Napoca, Romania,
E-mail: voicapuscasiu@gmail.com

When one talks about "digital art" there are a great number of very different manifestations that could potentially fall into this category. One of the first things to keep in mind is the fact that as a term "digital art" does not refer to any particular style or aesthetic, thus the great variety. Much rather it is a medium through which artists choose to present their ideas. However as with all things digital, this can take on a lot of different facets, the digital turn has provided not only the opportunity for art archival and dissemination, but also as an extensive creative tool whose only limit lies in the imagination and/or technical skill of the author. To put it simply, digital art can range from works created entirely with the use of a computer and software, to alteration done with the help of the digital medium. It can include appropriation of other classical works that receive new life and meaning through this approach.

Included in this last category there has been a sharp increase in the number of truly delightful projects, one of which includes a mesmerizing stroll through Salvador Dali's surrealist paintings. In *Dreams of Dali*, the viewers get to be in a virtual reality and "walk" around the deserted and bizarre landscape, passing underneath the spindly legs of Dali's signature elephants. The full virtual reality experience can be accessed at the Dali museum of Saint Petersburg, Florida, or is downloadable as an app for the users of HTC Vive or Oculus Rift for the modicum price of $2.99, but a linear 360º version – a panoramic video – is available online for anyone.[1] This new manner of interacting with the works of the famous surrealist engages the viewer in an entirely different way – and dare I say perhaps even a more enjoyable one?

[1] [http://thedali.org/exhibit/dreams-vr/]. Accessed June 1st 2017.

Seeing as though surrealist works hold a constant fascination in the public's eyes and mind, it is not entirely surprising that a similarly oriented project put the spiritual godfather of the current into focus. His name is Hieronymus Bosch, and few have not heard of him, who despite living and working around the 1550s, continues to intrigue and amaze people through the ages with his otherworldly and mysterious iconography. His paintings need no introduction, from the ghoulish imaginary creatures to the foul acts performed by mankind in nightmarish landscapes. He captivates the imagination perhaps even more so since the meaning behind his works has never quite been clarified. Arguably a pessimist and a moralist, Bosch shows the vile and hedonistic world of man and the horrifying repercussions that await in this realm or the next. What is more, the medieval imagery of Bosch's paintings seems to possess that rare quality of timelessness, by not giving true indications of specifically where or when it was created.

The year 2016 marked a full 500 years since the death of the painter, and as the year of his birth is but speculation, this was considered to be the proper time to pay an homage to this very peculiar artist. As the Year of Bosch unfolded in many locations around Europe the multimedia show Bosch. Visions Alive was opened in Moscow and was then moved to Berlin, which is the subject of this present chronicle. Such was the success of the show in Germany that even though it was supposed to stay open from July 6th 2016 to October 30th 2016, it actually got extended by popular demand to June 2017.

The project was created by Artplay Media, which according to their website[2] is an international team of professionals specializing in exhibitions, art, technology, sound, design, architecture, marketing, and public relations. Their quickly expanding portfolio is based on a patented technology titled "new media experience" which combines multimedia and interactive hardware, cinematography, and motion design in creating a synergy between education and entertainment, which they dubbed "edutainment". Not surprisingly this approach to art in a new format manages to bring in crowds of visitors. Artplay Media statistics boast over 1.000.000 visitors in 10 cities in 5 countries. Their itinerant shows including *Michelangelo*. *The Creation*, and *The Great Modernists*. *Revolution in Art* can be hosted as long as all the conditions (venue, expenses, etc.) are met, but they also offer solutions for permanent installations. As a side thought it is worth mentioning that Artplay Media are not the only ones involved in setting up this type of shows. Grande Exhibitions for example are true veterans in this field, and also hold the patent for the new media platform called Sensory4™[3], which they developed themselves. Their offer an even larger number of travelling shows such as *Van Gogh Alive – The Experience*, *Monet to Cézanne – The French Impressionists*, various da Vinci exhibitions,

[2] [http://www.artplaymedia.com/]. Accessed June 1st 2017.
[3] [http://grandeexhibitions.com]. Accessed June 1st 2017.

but also shows unrelated to art, like *Ottoman Empire – The Experience*, *Planet Shark: Predator or Prey* or *Leaders: Speeches That Changed the World*.

The show itself was hosted at Alte Münze, a typical Berlin space, repurposed from what was originally a coin minting centre and eventually transformed into a venue for various events. The entire exhibition occupied a total of 4 rooms, with an ample space that provided a one of a kind viewing experience. The very first room hosts not only the ticket office and the small gift shop which contains all the regular items one might expect, but also a couple of tables where refreshments can be purchased, or even alcoholic beverages for those who look to further enhance or alter their viewing. However, from a curatorial point of view, before passing through the dark curtain into the show per se, it is interesting to stop and take note of the slide show which features the original works that were used in the production of the multimedia feature. Since the show consists of fragments and characters from different works of the artist, that are made to interact, it is quite thoughtful to see the original, static sources and familiarize yourself with them beforehand. Hieronymus Bosch has a small number of accredited paintings, so it is obviously that some of the most intriguing and specific ones have been chosen for this project. Among them were the world-famous *Garden of Earthly Delights*, *The Last Judgment*, *The Temptation of Saint Anthony*, and *The Seven Deadly Sins and the Four Last Things*.

Going through the curtain the visitor is transported into a room brimming from floor to ceiling with an animated collage of Bosch's characters and landscapes. The entire video is played in an approximately 40 minute loop, and even though it is difficult if not impossible to control at which point of the video one would enter the room, the ending and the beginning sections of the loop are quite distinguishable through their rhythm. In a rather obvious manner, the starting scene was consecrated to the passing of time as indicated by a complex projection of rotating cog wheels, accompanied by the specific sound of ticking clockworks. Immediately after, starting from a liquid droplet, one was revealed Bosch's vision of the Creation as depicted on the closed panels of his *Garden of Earthly Delights* triptych.

However from then on the entire narrative was much less clear and basically featured a random flow of nightmarish scenes from the aforementioned paintings, which were brought to life through the means of very clever, albeit minimal, animation. Various parts of Bosch's vast painterly universe made an appearance from a giant knife cutting through a solitary human ear, to mermaids, faceless knights, flying fish and other animals – be they real or fantastic beasts, to highly detailed berries of all sorts, the latter of which were greatly enlarged and rained down like an eccentric storm all around the chamber. The only utterly unexplainable sequence that had the potential to ruin the entire visual experience, was that of a man' animated silhouette dancing to the instantly recognizable moves of Michael Jackson's Thriller, that is in the company of medieval musical instruments taken straight out of Bosch's works.

As for the space itself and how the video was shown it must firstly be mentioned that all four walls were covered in imagery from floor to ceiling, but none showed the exact same picture simultaneously. Instead there were slight variations be it in size or orientation. Some were zoomed in, while others were flipped on a vertical axis, so that each viewer could choose which "wall" to follow or even move between them. As a viewing option the space was also interesting in the sense that it provided benches and a cluster of bean bags in the middle of the room, but also all along the walls. This decision made for a cozy and comfortable stay and much less stiff than your usual art exhibition or even movie theater. The music, all of it exclusively instrumental, only added to the atmosphere, without being too loud or distracting for the moving visuals.

Moving along from the main room of the show, there was yet another, but much smaller room with less available wall space for projections seeing as though it has two doors. There was still a bean bag or two, but the greatest asset of this particular room came from the fact that it offered to possibility to chance to view the images up close and to revel in all their wonderful details and to truly appreciate the astonishing quality of the resolution.

Last but not least, there was yet another room free of any moving image, but which focused on the file and times of Hieronymus Bosch. The centerpiece of the room was a chronological timeline that marked every single year of the artist's life starting from 1450 onward. A total of 66 years, each featuring a relevant thumbnail image and a narrow column of text which pointed out a wide array of facts – such as the birth of Barbarossa in 1475 or the death of Vlad the Impaler in 1477, but also mentioned particularly draughty years, or the creation of important and relevant works of art or political acts. All in all without following a specific discourse it seemed aimed to paint a lively, eclectic picture of more than half a century in European history and it absolutely managed to do just that.

This room also provided a tribute to Bosch's artistic influence throughout Art History by linking all the artists that proudly cited him as their source of inspiration. The names rage from Pieter Bruegel the Elder, his contemporary to the already mentioned Salvador Dali and other representatives of the artistic avant-garde like Marcel Duchamp, Max Ernst, Francis Picabia, Otto Dix, Paul Eluard, Paul Klee, and George Grosz, all of which dabbled in surrealist imagery. The curatorial choices present in this room enriches the show in a way that makes viewers realize it is much more than an entertaining production through digitization and video art, but also provides historical context into the life of this strange character. The connections with other artists also try to connect the loosely arranged dots in Art History, while at the same time underlining Bosch and making him stand out as a singularity.

Speaking of which, other than the 500$^{th}$ anniversary of his death, Bosch is already thought of as a highly interesting artist because of the mystery aura surrounding his paintings. Due to the fact that his works are brimming with small and intricate details

that could always benefit from closer inspection, it is quite natural that they were used in projects meant to showcase the wonders of new technologies regarding digitization. The Google Art Project is already involved in creating extremely high-resolution images of masterpieces, but Bosch's painting was particularly selected in yet another separate project. The *Garden of Earthly Delights* was turned into an online interactive adventure[4] in which you could wander through the painting and discover its many stories through audio commentary or text – with a children's version also available, but only in Dutch. This goes to show the continuous fascination that this artist still holds on the imagination of contemporary art lovers, and story lovers of all kinds.

It is safe to say that the creators of this exhibition have successfully tapped into a well-received genre, which is also easily seen by the success of similar shows. It is hard to decide if this type of viewing is ultimately able to replace the experience of seeing the actual paintings, or even if it ends up providing something more, whatever that may be. Sincerely I consider that as a non-issue, it is simply a different experience and that is important in itself. It is an entirely new way of viewing art, brought on by the ever-evolving digital means and in this age of democratization of art, it is quite undeniable that this kind of show has the potential to bring Bosch (as well as others) to the attention of a different group of people than the ones who visit museums. And what is more, it can truly influence their perception of art and increase their overall interest for this subject. The educational potential brought on by the digital turn should never be underestimated, perhaps it can even make people want to see and know more about a painter's life and work. And even without this very optimistic note, I must admit that I do look forward to many more of these reinterpretations, if only for the sheer visual and sensorial pleasure they provide.

[4] [https://tuinderlusten-jheronimusbosch.ntr.nl/en]. Accessed on June 1st 2017.

*Studia UBB* Digitalia, Volume 62 (LXII) 2017, June, Issue 1, 138-142

Published Online: 2017-06-30

Review

# Viorella Manolache, *Signs and designs of the virtual(izing) E@St*, introduction Ian Browne, instead of an afterword Cristi Pantelimon, LAP Lambert Academic Publishing, Saarbrücken, 2013, 124 p.

Valentin Trifescu

Babeș-Bolyai University Cluj-Napoca / Sapientia University Miercurea Ciuc,
E-mail: valentintrifescu@yahoo.fr

Înțeles în sens larg, fără pretenția unor redefiniri sau reinterpretări teoretice, postmodernismul a avut în spațiul cultural românesc manifestări vizibile mai ales în domeniul literar, muzical ori al artelor vizuale. În ceea ce privește cercetarea științifică din domeniul umanioarelor, puțini au fost cei care au avut vocația sau curajul de a fi contemporani-cu-epoca-lor în stil dar și în conținut. Cu alte cuvinte, cercetarea științifică românească a rămas cantonată în vechile tipare metodologice, tematice și stilistice, păstrând „toți nasturii închiși la cămașă" și constrângerea unei cravate deja demult demodate, purtate pe post de marcă a seriozității academice. Cu toate acestea, câțiva cercetători cu o deschidere interdisciplinară mai amplă – asemenea unor temerari experimentatori ai „stilului neastâmpărat" în cercetare sau a unora care știu-să-se-joace-serios și să facă din actul lecturii academice o plăcere nu numai pentru creier, dar și pentru urechi sau ochi, cum sunt (pentru a intra în *logica lui Trei*[1]) clujeanul François Bréda, sibianul Gheorghe Manolache ori bucureșteanul Bogdan Ghiu – au abordat subiecte marginale ori metisate într-o scriitură cu certe valori estetice, nefăcând rabat, în același timp, de la rigorile științifice. Importantă pentru acești cercetători a fost realizarea unor texte polivalente, structurate pe mai multe paliere de înțelegere, cu trimiteri mereu subversive „pour les connaisseurs", dar care, nu în ultimul rând, trebuie să producă plăcere. De cele mai multe ori, tehnica (de codare) folosită pentru lansarea mesajului ideatic apelează la: jocuri de cuvinte, întreruperi, pauze, puncte de suspensie,

[1] Vezi Viorella Manolache, *Philosophical-Political Hecate-isms. The Rule of Three*, Cambridge Scholars Publishing, 2016.

paranteze, slashuri, sublinieri, ruperi de ritm, fragmente, aluzii, linkuri, travestiuri, mixaje, combinații neașteptate ori kitschuri de toate felurile (de limbaj, bibliografic, tematic, discursiv etc.).

Într-un mod cu totul particular, Viorella Manolache, căci despre ea este vorba în rândurile următoare, se remarcă drept unul dintre puținii intelectuali români care fac cu adevărat o cercetare postmodernă, adaptată la ultimele achiziții științifice ale momentului. Pentru a fi *cool*, putem vorbi și de o formă de manifestare a hipsterității în cercetare. Viorella Manolache continuă preocupările sale anterioare, deja centrate pe marginal și pe inedit[2], adaptându-le tehnologiei și erei digitale. Cartografierea spațiilor identitare este transferată de această dată în domeniul *virtual*iza(n)t, cartea *Signs and Designs of the Virtual(izing) E@St* reprezentând un exercițiu metodologic prin care au fost interpretate, într-o cheie *soft-ware* dar și *soft power* „gruparea intervențiilor (cu aproximație) săptămânale (e)lansate în *jurnalul și blogul Estica*, texte deschise-la-vedere, care ilustrează corelarea preocupărilor de cercetare academic-științifică cu investigațiile de radiografiere a realității local-globaliza(n)te în desfășurare; de fapt, este vorba despre o aplicare a tehnicii *dezarticulatoare*, cu miză în fragmente separate și updatate prin enunțare" (p. 9).

Autoarea declară deschis, încă de la început, că „Demersul de față nu se poate distanța de un anume efect de *cantabilitate* – înțeles, deopotrivă, cu sensul de artisticitate și de produs artizanal – recunoscând *platformei Estica* statutul de *server textual*, autoasamblaj, mod declanșator, de proces și distribuitor de informație. Cu o primă avertizare, aceea că prezentul studiu nu intenționează să schimbe/adnoteze stilul/structura și *așezarea formatată* a intervențiilor, urmând/respectând cerințele tiparului electronic și dimensiunile spațiului indicat (număr de semne/cuvinte) – mod condensa(n)t de redactare a textului și exercițiu de poziționare a acestuia înlăuntrul unor formule/forme/tipare prestabilite. // În acest sens, nu putem să nu amintim, cu referire la deghizările *semnificațiilor pline* în *mesaje diluate*, opinia lui Bogdan Ghiu[3], cel care considera că *punctul* dinlăuntrul denumirilor adreselor de internet ar reprezenta un *marcaj* de ruptură/întrerupere, treaptă istorică, prag discret de trecere, de fapt, ar însemna traducerea *limbii* culturii în *limbaj - net*, ca *mod basic* simplificat. // În același registru, pentru Jean-Pau Nerrière[4], dincolo de fragmentările de ordin lingvistic, ar exista «o formă autentică, corectă de *angloricană»*, limitată la un «număr de doar 1500 de cuvinte deja cunoscute», cu o sintaxă elementară și completată cu/de procedee practice de formulare, o limbă generală și uniformă, formă dominantă de «*English light*» – limbajul *Globish*, expresie a filosofiei modului planetar de viață" (p. 9).

[2] Vezi Viorella Manolache, Henrieta Anișoara Şerban, *Cartografierea marginalității*, Editura Institutului de Ştiințe Politice și Relații Internaționale, București, 2010.

[3] Bogdan Ghiu, *Globish - comunicarea (II)*, în *Atelier LiterNet*, 20.08.2006, http://atelier.liternet.ro/articol/3725/Bogdan-Ghiu/Globish-comunicarea-II.html, accesat la 1 iulie 2013.

[4] Jean-Paul Nerrière, *Don't speak English, parlez globish*, Groupe Eyrolles, 2006.

În primul mare capitol, intitulat *E@ST'S Positioning of the Virtual Map* (*Așezarea Est@ului pe harta virtuală*), autoarea realizează o topografie virtuală a Estului european, pornind de la afirmațiile lui Alfred Korzybski, potrivit cărora „*harta nu este teritoriu*" precum și că „*tot ceea ce este, există undeva (într-un anume loc)*" (p. 13), fomule care au fost cuplate, prin intermediul lui Augustin Ioan, la marea experiență antică premonitorie a lui Platon și Archytas din Tarentum. În această ordine de idei, Est-ului, așezat pe harta vituală, i se va mai adăuga un „@" pe post de marcă, cratimă sau punct-de-trecere/link de acces spre lumea virtuală; Est@ul geografico-cultural european devenind astfel un *spațiu convertit* într-o *realitate digitală*, postmodernă prin excelență. „De aici și translatarea *locului*, prin elasticități multiple, până la *spațiu imaginat – loc vizualizat*, *spațiu imaginar – loc virtualiza(n)t*" ( p. 13).

Inevitabil, se aduce în discuție „un altfel de spațiu", *cyberspațiul*, care – legând spațiul public de cel privat, realitatea de virtual – ridică o serie de probleme tehnice, filosofice sau etice. Totodată, este nevoie de o clarificare definițională a conceptului. Astfel, „Depășind definițiile canonizante potrivit cărora «realitatea virtuală este o simulare generată pe calculator a unui mediu tridimensional, pe care utilizatorul are posibilitatea să-l cerceteze și să-l manipuleze ca și cum ar fi real»[5] sau potrivit cărora, spațiul mediului virtual ar aluneca între/printre imperativul de stocare și cel de reînnoire de informații, de simulare și redare, de generare și navigație, de interacționare coincidențială *lume virtuală - spațiu public*, prezentul demers va accepta *spațiul virtual* drept *loc-surogat*, prelungire tentaculară a *realului* în *arealul simulat*. // În acest sens, nu vom opera echivalențele riscante ale spațiului virtual cu ale celui utopico-distopic, cu toate tentațiile/capcanele (sur)prinderii «codificărilor și standardizărilor» spațiului alternativ – imaginat-dar-încă-neprodus[6], constatând că, mai ales în cazul *cyberspațiului*, acesta împrumută coordonatele unui *spațiu mental,* non-fizic și imaginar. Afirmația nu este deloc surprinzătoare în măsura în care conceptul de *hartă mentală* se reîncarcă de la zonele spațiului virtual, corelând *zona centrală*, cu volum redus de informații, cu accent pe valori și credințe – cu *zona periferică*, de stocare a unui (a)flux lărgit de informații, chiar dacă asumate inconștient, dar care au tendința de a se schimba în timp" (p. 14).

Extrem de densă în conținut, analiza *Est@ului pe o hartă virtual(izant)ă*, realizată de Viorella Manolache, aduce în discuție o mulțime de alte concepte-cheie, care ramifică în mai multe direcții subiectul inițial. Într-o abordare topocritică sau geocritică apar: *toposul virtual*, *cyberpunk-ul*, *nonlocul*, *plusspațiul*, *spațiul fetișizat*, *harta mentală*, *harta stereotipală*, *geografia revizuibilă*, *societatea în rețea*, *comunitatea virtuală*, *logosul virtual*, *identitatea digitală*, *non-prezența* etc.

[5] D. Popescu, D. Şendrescu, *Realitatea virtuală*, Editura Universitaria, Craiova, 2002, p. 5.
[6] J.-J. Wunenburger, *Utopia sau criza imaginarului*, Editura Dacia, Cluj-Napoca, 2001, p.19.

Mizele locale și regionale sunt dezvoltate și transferate în capitolul *A Virtual(izing) Europe Project* (*Proiectul unei Europe virtualiza(n)te*) asupra întregului spațiu european. În acest fel, „Deloc imun la asaltul virtualului asupra realității în forma sa canonizantă, modelul european, declarându-și drept finalitate umplerea lacunelor dinlăuntrul spațiului interdisciplinar al studiilor despre integrare, rezonează la mutațiile virtuale prin platforma de cercetare CVCE. Aceasta favorizează accesul publicațiilor digitalizate, drept *corpora* de cercetare, cu tot ceea ce presupune platforma particulară a fișierelor pe subiect, a hărților interactive, a diagramelor sau a cronologiilor generale" (p. 36). Cu această ocazie, metodele și rezultatele de analiză ale materialului digital vor căpăta, într-un mod firesc, denumirile de: *e-cercetare*, *e-umanisme*, *e-publicații* sau *cyberinfrastructuri*.

Inevitabil, sunt atinse limitele democrației și ale culturii în *epoca maselor* (anunțate profetic de *José* Ortega y Gasset), ale accesului la informații și ale libertății de a alege. În acest sens: „Sub semnul autoironiei și al autorelativizării, ca strategie de comunicare, Sascha Lobo[7] denunța depășirea democrației scrise prin adaptarea ei la mediul virtual, ai cărui explicatori formează un parlament online, cu toate deschiderile transparentizante înspre modelul unui canal de *feedback. Intervalul poartă însemnele la vedere ale tranziției curioase*, subsumate coincidențelor virtual-politice și ancorate în ceea ce Sascha Lobo condamna prin «*faptul că, dacă oamenii citesc doar primele 10 rezultate ale căutării pe Google, nu este problema Google*»" (p. 38).

În capitolul Estica: Journal and Blog (Estica: jurnal și blog), Viorella Manolache abordează, într-o manieră deosebit de fragmentară o serie de subiecte de interes românesc, cum ar fi: localismul creator, feminismul, biopolitica, și nenumărate alte tematici cu implicații politice. Toate acestea pornind de la ideea că: „Acest subcapitol își propune să comenteze modificările lansate de coordonatele virtuale cu impact în posibilele interogații: În ce măsură hărțile virtuale pot fi aplicate oricăror mesaje? Virtualul sincronizează Estul cu modelele tari ale Vestului? Reprezintă virtualul sfârșitul spațialităților tradiționale? // Ipoteza generală are în vedere evidența că platforma Estica funcționează ca o hartă, cu multiple trasee, recunoscând istoria comună: texte filosofico-politice, texte culturale, semnale literar-culturale și filosofico-politice, localizate [...] înlăuntrul perimetrului Estic românesc, confirmând axioma că șablonul virtual poate fi aplicat oricărui mesaj" (p. 48).

Viorella Manolache ne propune o carte unică în bibliografia de specialitate din România, care reprezintă un prag important de trecere spre noi discipline (interdisciplinare prin definiție) care racordează științele umanistice la era digitală. Putem spune că autoarea face o pledoarie savantă pentru științele umanistice digitale. Intuind, de la bun început, că mediul virtual implică o strânsă legătură cu mediul real, precum și faptul că acesta presupune o evidentă dimensiune spațială care se pretează la o cartografiere specifică,

[7] Sascha Lobo, „Sălbaticul războiului cibernetic", în *Die Zeit*, 22 mai 2012.

geocritică, în care spațiile, locurile, simbolurile, semnele, grafiile și textele sunt decodate (ba chiar traversate prin intermediul sistemelor GIS) într-o cheie virtual(izant)ă adresată de la bun început unui destinatar exterior, din spațiul real, dar intrat în logica digitală (p. 106). În cele din urmă, după cum afirmă deschis autoarea, „De fapt, lucrarea demonstrează că spațiul Estic, în general, și cel românesc, în particular, nu suferă de complexe în a-și construi virtual rampe de lansare, neputând nega însă întreținerea statutului de culoar/ terminal[8] dintre/între continuitate și fractură; toate acestea fără a eluda modelul preluării, al racordărilor la modă, al tentației de a fi în flux, cu lansarea, pe nerăsuflate, în asumarea și pe cale virtuală a unei cetățenii universale (e-cetățean). // Fără a acorda câștig de cauză realului sau virtualului, în sens hassan-ian, considerând că marca postmodernismului este aceea de a recurge la o serie de tehnicisme, în demersul nostru ne-am propus să demonstrăm că nu este vorba despre un fenomen de înlocuire, ci despre instaurarea unei relații democratice (parteneriale) între real-virtual, de coproprietate sau de încorporare a osemintelor moderniste în lutul modelator al postmodernismului; sau de ce nu, doar de o reconvertire/refolosire a conceptelor modernismului târziu" (p. 109).

Acestea fiind spuse, nu putem decât să felicităm Autoarea pentru munca depusă și să ne bucurăm de rezultatele acestei cărți-manifest!

[8] Cu aluzie la termenul de *platformă*, *terminalul* poate fi înțeles (aici) ca aerogară urbană care servește drept punct de sosire sau de plecare a pasagerilor și, deopotrivă, dispozitiv de intrare sau de ieșire, cuplat cu calculatorul; element de conectare a componentelor electronice periferice ale unui calculator, organ de intrare sau de ieșire a datelor. Nu putem omite, în armonizarea celor două note definiționale, imaginea livrată de un film precum *Terminalul* (2004), în sensul denunțării, deopotrivă, a unui spațiu Estic de nelocalizat - „Krakozia" - și a unui "no man's land", cu statut de „home"- aeroportul JFK din New York, *platformă* care descărcată de orice realism politic ușor de suprapus legendei urbane, care i-a servit drept inspirație lui Spielberg, reîncarcă profilul apatridului cu desemnele celui care ajunge să domine, în sens „colegial", spațiul, trasându-i sau, din contră, ajustându-i, prin deschidere, regulile. Perspectiva poate fi direcționată spre acel sens badiou-ian al lumilor susținute de un contra-punct, drept formulă (de)teritorializantă, având ca unică existență *reprezentația*.