

# TEHNOLOGII DE MARKUP PENTRU CREAREA DE EDIȚII DIGITALE

NICOLAIE CONSTANTINESCU

Universitatea din București

SILVIU BORȘ

Universitatea „Lucian Blaga” din Sibiu

## Markup Technologies for the Creation of Digital Editions

### ABSTRACT

Today, more than ever, it is essential to develop workflows for creating digital editions that are tailored to the specifics of the content being processed (in other words, migrating works that exist only in print format to digital format where this is necessary for various reasons). In this article, we explore three core technologies that enable digital editions for works that need to be converted to digital format. These core technologies are Extensible Markup Language (XML), Extensible Stylesheet Language Transformations (XSLT) and the Text Encoding Initiative (TEI). The article discusses these technologies in the context of other initiatives and projects, and illustrates the results that can be achieved through efforts to digitise and transform works in traditional formats. The aim of the research is to establish some robust benchmarks for the validation of these technologies, thereby contributing to the development of a framework for advancing the practice of digital humanities among specialists dedicated to the management and valorisation of cultural heritage.

**Cuvinte-cheie:** digitalizare, carte veche, editare, internet, baze de date

**Key words:** digitalization, old book, editing, internet, data bases

### Introducere

Fiecare ediție a unei opere scrise contribuie nu numai la comunicare, ci și la conservarea unor repere ale gândirii într-un context istoric specific, reflectate în exemplarele distribuite la un anumit moment. Edițiile Digitale Științifice (Digital Scholarly Editions) reprezintă o manifestare a experienței și tehnicilor dezvoltate de specialiștii din domeniul Umanioarelor Digitale. Acest domeniu este definit în moduri diverse<sup>1</sup>, dar, pentru acest articol, ne vom raporta la definiția oferită de Profesorul Patrik Shale de la Universitatea Bergische din Wuppertal, Germania, în prezentarea domeniului Umanioarelor Digitale. Conform acestei definiții, Umanioarele Digitale sunt o punte, un punct de intersecție și o meta-disciplină ce integrează umanioarele, științele informării și tehnologia informației într-un context aplicativ<sup>2</sup>.

---

<sup>1</sup> Terras, Nyhan, Vanhoutte 2013, p. 6.

<sup>2</sup> <https://www.geschichte.uni-wuppertal.de/de/lehrgebiete/digital-humanities> (accesat la 05. 06. 2024).

Puntea cu tehnologia informației este și cea care permite prelucrarea textelor sub o formă nouă, capabilă să fie manipulată digital. Conținuturile edițiilor actuale sunt create direct în format digital, iar cele realizate anterior în formate tradiționale sunt convertite în ediții digitale prin procesul de digitizare. Printre beneficiile digitizării, enumerăm accesibilitatea îmbunătățită a materialelor, flexibilitatea oferită de conținutul digital și capacitatea de a face conținutul colecțiilor mai interactiv și accesibil la nivel global<sup>3</sup>. Digitizarea permite căutarea, prelucrarea și realizarea de referințe, care în formatele tradiționale se realizează prin indexare manuală, neasistată de tehnologii informatice. Indiferent de format, textul joacă rolul principal, fiind și materia primă a cercetătorului. În faza de publicare, toate tehnologiile converg, servind obișnuințelor de lucru proprii fiecăruia cu următoarele posibile repere: consultarea online a unei ediții digitale (în format HTML sau ePUB), descărcarea sub formă de date (prin API – Application Programming Interface), sau transformarea în formate potrivite tiparului, precum PDF, fără a depinde de o platformă informatică sau alta.

Una dintre mizele importante ale întregului efort de realizare a unei ediții digitale este aceea de a avea acces la materia primă sub formă de date, care să fie supusă analizei, folosind instrumente automatizate sau semiautomatizate în contextul diverselor scenarii de investigare. XML-ul se dovedește a fi tehnologia standardizată care nu depinde de soluții software. O serie explicativă pentru un prim contact de inițiere în înțelegerea tehnologiei este pusă la dispoziție de *Saxonica*<sup>4</sup>. În *Catalogul edițiilor digitale științifice* editat de Patrick Sahle și compania, pus la dispoziție de Universitatea Wuppertal din Germania<sup>5</sup>, avem o definiție de lucru importantă din punctul de vedere al analizei istorice: *o ediție științifică este reprezentarea critică a documentelor istorice*. Această definiție este lămurită într-un context mai larg într-o lucrare de referință pentru domeniu<sup>6</sup>, asupra căreia vom reveni mai jos. Datele *Catalogului* sunt realizate folosind Text Encoding Initiative (TEI), fiind instrumentul specializat TEIPublisher<sup>7</sup>. Institutul pentru Documentologie și Editare Științifică (Institut für Dokumentologie und Editorik – IDE) publică cu succes *Review Journal for Digital Editions and Resources (RIDE)*<sup>8</sup>, o revistă electronică în regim open access, cu articole care evaluează și promovează calitatea și inovația edițiilor

---

<sup>3</sup> Cohen, Rosenzweig 2006.

<sup>4</sup> <https://www.saxonica.com/technology/technology.xml> (accesat la 05.06.2024).

<sup>5</sup> <https://www.digitale-edition.de/> (accesat la 05.06.2024).

<sup>6</sup> Driscoll, Pierazzo 2016, p. 19.

<sup>7</sup> disponibil la <https://teipublisher.com/index.html> (accesat la 05.06.2024) cu sprijinul *e-editions* de la <https://www.e-editiones.org/> (accesat la 05.06.2024).

<sup>8</sup> <https://ride.i-d-e.de/> (accesat la 05.06.2024).

digitale care reprezintă colecții critice de texte sau materiale primare<sup>9</sup>. În *Criteriile pentru evaluarea Edițiilor Științifice Digitale*, versiunea 1.1<sup>10</sup>, găsim o definiție foarte utilă a edițiilor digitale științifice, care oferă o perspectivă clară asupra înțelegerii acestora, inclusiv din perspectiva tehnologiilor implicate: *acestea nu sunt doar publicații în formă digitală; mai degrabă, sunt sisteme informaționale care urmează o metodologie determinată de o paradigmă digitală, la fel cum edițiile tipărite tradiționale urmează o metodologie determinată de paradigmele culturii tipărite*. Realizarea unei ediții digitale presupune câteva etape pe care le vom trece în revistă. O parte dintre acestea reflectă *Criteriile pentru evaluarea Edițiilor Digitale Științifice*. Joris van Zundert, cercetător la Institutul Huygens pentru Istoria Olandei, indică metodele de abordare a textului ca *obiect digital* prin: materialitatea performativă, sociologia textelor și filozofie reprezentatională. Ceea ce este interesant sunt perspectivele funcționale pe care le expun edițiile digitale științifice din perspectiva aceluiași autor la nivelul practicii. Acestea sunt *reconstrucția* (a textului, bineînțeles), *metodele de cercetare, ediția/arhiva (referințe și publicările textelor ori dovezile că acel text există), ediția critică, ediția genetică (capturarea procesului prin care a fost creat textul), site-urile de cunoaștere (textul însoțit de alte texte care îl servesc, îl explică, îl completează) și textul multidimensional (reprezentare ca graf)*<sup>11</sup>.

### **Etape pentru valorificarea textului ca obiect digital**

Pentru a avea la îndemână un set de date utile, colectivul care are în misiune valorificarea unui corp de texte cu ajutorul mijloacelor de prelucrare digitală va avea drept obiective primare crearea de date structurate și de markup relevant scopurilor. Cel mai adesea se pornește de la operele în format tradițional, care sunt supuse unui flux de prelucrare în mai multe etape. Prima etapă include optical character recognition (OCR), optical layout recognition (OLR sau segmentarea de conținut) și/sau Handwritten Text Recognition (HTR). O soluție tehnică o oferă OCR4all<sup>12</sup>, care expune un flux complet automatizat de OCR-izare, având la bază un software cu sursă deschisă<sup>13</sup>. XML-ul este tehnologia folosită pentru descrierea paginilor, formatul PAGE XML fiind cel propus pentru a surprinde și marca cu acuratețe dispunerea elementelor unei pagini de text (incunabule etc.)<sup>14</sup>.

---

<sup>9</sup> Neuber, Henny-Krahmer, Assmann, Scholger 2024, resursă digitală.

<sup>10</sup> Sahle, Vogeler 2024.

<sup>11</sup> van Zundert 2024, minutul 28:36.

<sup>12</sup> Reul, Christ, Hartelt, Balbach, Wehner, Springmann, Wick, Grundig, Büttner, Puppe 2019.

<sup>13</sup> Vezi mai multe detalii la <https://www.ocr4all.org/> (accesat la 05.06.2024).

<sup>14</sup> Pletschacher, Antonacopoulos 2010, p. 257-258.

Un alt standard XML pentru detectarea aranjamentului paginii care este folosit adesea în proiectele de digitizare este ALTO (Analyzed Layout and Text Object), pe care Biblioteca Congresului American îl dezvoltă<sup>15</sup>. Dacă este posibil, este necesară parcurgerea unei etape în care vor fi *marcate* diferite fragmente din text care identifică persoane, locuri sau evenimente. Această etapă este cunoscută drept *named-entity recognition*. Formatarea rezultatului se va face folosind tehnologii din familia XML precum SLT, XSL-FO, XQuery sau chiar CSS. Acestea ar asigura o independență de soluțiile software. Scripturile necesare execuției oricăror transformări ale documentului deja codat, ar trebui să fie disponibile în asociere cu sursa de bază. Vorbim despre XML Schema-uri, XSLT-uri sau DTD-uri. Tehnologiile de markup sunt cele care asigură structura datelor și pot oferi prin prelucrare ulterioară și o versiune hypertext. Când privește schimbul de metadata, preferabil este ca datele și metadatale să fie compatibile cu protocolul OAI, ceea ce înseamnă că vor trebui să aibă un minim de metadata conform Dublin Core. Atribuirea de metadata folosește de cele mai multe ori tot markup XML conform unei anumite scheme. O ultimă etapă într-un proces complex este planificarea și execuția etapelor privind prezervarea digitală a datelor. În ajutor avem un alt standard bazat pe scheme XML: PREMIS (Preservation Metadata: Implementation Strategies) Data Dictionary for Preservation Metadata. Transformarea lucrărilor existente în format tradițional în obiecte digitale constituie o parte importantă care așază cercetătorul în poziția zero a propriilor investigații asistate de instrumente digitale. Următorul pas esențial este jalonarea entităților din text folosind limbaje de markup.

### **Tehnologii de markup**

După prelucrarea textelor, rezultatul final reflectă un proces de *datificare* a textului (în limba engleză *datafication*), adică transformarea acestuia în date cuantificabile și analizabile, adaptate mediilor digitale. Scopul acestui proces este de a facilita căutarea și de a permite integrarea sursei în fluxuri complexe de prelucrare. Principalul instrument pentru a realiza codificarea este cel mai folosit limbaj de markup: Extensible Markup Language (XML), așa cum am punctat deja. Motivul pentru care este folosit XML-ul este legat de faptul că un document este reprezentat folosind un arbore de noduri care are unul central (*root element*), cu rol de părinte pentru restul. În ceea ce privește modul de procesare a textelor, dacă acestea sunt codificate utilizând XML, ele se transformă în grafuri nedirecționate în care două *noduri* sunt legate printr-o *muchie* conform teoriei. Reprezentarea textului sub forma unui graf nedirecționat face posibilă o analiză a acestuia mai ușor de gestionat, dezvăluind tipare, relații și structuri, care, altfel, ar fi greu de observat, cum ar fi relații

---

<sup>15</sup> Alto 2024.

tematice între termeni (asocieri semantice), structura relațiilor interpersonale întâlnite la personajele din operă, identificarea de rețele semantice între termeni, stabilirea de teme sau de idei recurente și interconectarea acestora, analiza structurii unui text academic, identificând referințele comune sau secțiunile cu legături tematice<sup>16</sup>.

### **XML, XSL și XPath**

Aceste tehnologii sunt fundamentale pentru a putea pași pe calea realizării Edițiilor Digitale. XML (Extensible Markup Language) și familia limbajelor complementare care sunt numite generic Extensible Stylesheet Language (XSL) constituie setul fundamental de instrumente digitale necesare realizării de ediții digitale științifice.

XML Path Language (XPath) este o Recomandare W3C<sup>17</sup>, cu ajutorul căreia poți selecta noduri dintr-un document XML. Această tehnologie poate fi folosită pentru a lucra cu noduri ale arborilor XML și JSON în scopul căutării și/sau transformării documentelor XML.

XSL Transformation (XSLT – Extensible Stylesheet Language Transformations) este un limbaj cu ajutorul căruia pot fi transformate documentele XML în alte documente, fie acestea XML, XHTML, fie HTML, folosind un document care indică cum vor fi făcute acele transformări (stylesheet). Versiunea curentă (2022) este 3.0, care în 2017 devine o Recomandare a W3C<sup>18</sup>.

### **Text Encoding Initiative**

Formatele XML, aranjamentul și regulile de codare sunt cele care oferă avantaje cercetătorului. Cel mai cunoscut este Text Encoding Initiative. Text Encoding Initiative este o inițiativă de standardizare a practicilor care privesc codarea textelor folosind Extensible Markup Language, care a început în 1988. În 2014 s-a constituit TEI Consortium care are drept țintă crearea unei scheme de codare pentru orice tip de text, materializată în TEI Guidelines<sup>19</sup>. TEI nu este util doar pentru a coda textul la nivel structural, începând cu TEI Header, ci este util și tuturor elementelor descriptive suplimentare care sunt specifice unui anumit document, cum ar fi filigranul sau marginaliile în contextul codicologiei.

---

<sup>16</sup> Malliaros, Vazirgiannis 2017, p. 10-14.

<sup>17</sup> Disponibil la <https://www.w3.org/TR/xpath-31/> (accesat la 05.06.2024).

<sup>18</sup> Accesibil de la <https://www.w3.org/TR/xslt-30/> (accesat la 05.06.2024); XSLT folosește XPath prin procesorul dedicat Saxon XSLT.

<sup>19</sup> Există o listă extinsă de proiecte care folosesc TEI, fiind accesibilă de la următorul link <https://tei-c.org/activities/projects/> (accesat la 05.06.2024). Pentru practicieni există și o revistă specială, intitulată *Journal of the Text Encoding Initiative*, disponibilă la <https://journals.openedition.org/jtei/index.html> (accesat la 05.06.2024).

Un exemplu de realizare a unui *site de cunoaștere* este Inscriptions of Roman Tripolitania 2021<sup>20</sup>. Un alt exemplu este însăși revista Digital Humanities Quarterly, ale cărei articole au o versiune TEI/XML; pentru articolul *Something Called "Digital Humanities"* avem o reprezentare în format XML: <http://www.digitalhumanities.org/dhq/vol/2/1/000020.xml>.

Aceste tehnologii nu pot fi separate de necesitatea de instruire care presupune dobândirea de noi abilități tehnice și organizatorice. Practicile care privesc tehnologiile de markup sunt legate foarte strâns de o rețea de inițiative și organizații care găzduiesc activitatea. Următorul pas este să investigăm câteva repere necesare înțelegerii contextului mai larg.

### **Rețele, infrastructuri și proiecte de referință**

DARIAH (Digital Research Infrastructure for the Arts and the Humanities) este o infrastructură paneuropeană a cercetătorilor din domeniul umanioarelor și artelor. Dincolo de proiectele pe care le desfășoară partenerii<sup>21</sup>, un punct de interes deosebit îl constituie resursele de învățare puse la dispoziție în DARIAH Campus<sup>22</sup>. Multe dintre aceste materiale sunt dezvoltate folosind platforma [programminghistorian.org](http://programminghistorian.org). Materialele de studiu disponibile pot constitui tot atâtea puncte de pornire pentru conceperea unor planuri de valorificare a colecțiilor de materiale text. Sunt disponibile soluții care privesc scanarea și OCR-izarea operelor în format tradițional, precum și promovarea și expunerea acestora folosind interfețe adaptate. European Association for Digital Humanities (EADH)<sup>23</sup> este punctul cel mai important de unde se pot explora cele mai bune practici la nivel european.

### **Bune practici**

Începând cu anul 2020, Fundația Germană pentru Cercetare (Deutsche Forschungsgemeinschaft – DFG) desfășoară un *Program de Finanțare pentru Digitizare și Indexare* ('Digitisation and Indexing Funding Programme', 2024), care privește *obiectele legate de cercetare de orice fel*, printre care documente scrise de orice fel, dar și documente nativ digitale (email-uri etc.). Este vizată și posibilitatea de a realiza conectarea între colecții, precum și integrarea de vocabulare controlate utile în identificarea conceptelor și a entităților. Aceeași instituție a publicat în anul 2023 *DFG Practical Guidelines on Digitisation*, versiunea 2022<sup>24</sup>, care este o actualizare a îndrumărilor deja existente. Acest document prezintă un interes deosebit pentru că reflectă experiența a zeci de

<sup>20</sup> <https://irt2021.inslib.kcl.ac.uk/en/> (accesat la 05.06.2024).

<sup>21</sup> <https://www.dariah.eu/activities/projects-list/> (accesat la 05.06.2024.)

<sup>22</sup> <https://campus.dariah.eu/> (accesat la 05.06.2024.)

<sup>23</sup> <https://eadh.org/> (accesat la 05.06.2024.)

<sup>24</sup> Altenhöner, Berger, Bracht, Klimpel, Meyer, Neuburger, Stäcker, Stein 2023.

specialiști din cele mai importante instituții de cercetare și de memorie din întreaga Germanie. Metadatele atribuite obiectelor digitale obținute în urma digitizării vor fi codate folosind XML-ul. Metadatele descriptive pentru obiectele digitale care reprezintă resurse textuale vor fi în format Text Encoding Initiative (TEI). În ceea ce privește afișarea metadatelor folosind DFG Viewer, metadatele vor fi în format Metadata Encoding and Transmission Standard / Metadata Object Description Schema (METS/MODS) pentru documentele text și documentele de arhivă, iar pentru manuscrise se va folosi Metadata Encoding and Transmission Standard / Text Encoding Initiative (METS/TEI).

Academia de Științe a Austriei (ÖAW) organizează activitatea Centrului Austriac pentru Umanioare Digitale și Patrimoniul Cultural (ACDH), având o bogată activitate de cercetare. Unul dintre cele mai interesante proiecte desfășurate a fost prezentat la DARIAH Annual Event, fiind implicate etape de transformare a unor colecții existente în diferite formate XML care răspundeau diferitelor obiective de cercetare. În prelucrarea datelor au fost utilizate scripturi XSLT care au permis exportul datelor din Transkribus (software folosit cu succes pentru culegerea textului din diverse surse digitizate) în TEI. Un alt script XSLT a fost folosit pentru a transforma datele XML în format METS, în format TEI. Întregul flux de transformări a fost gestionat prin Github Actions<sup>25</sup>. La sfârșit, rezultatul poate fi generat ca o publicație în format ePub, utilizând scriptul disponibil online<sup>26</sup> (Arthur Schnitzler und ich), care convertește sursele XML în XHTML. Tot în Austria, Ministerul Federal al Educației, Științei și Cercetării finanțează dezvoltarea unei platforme dedicate creării edițiilor digitale prin Digital Edition Competence Network (KONDE – Kompetenznetzwerk Digitale Edition – digitale-edition.at), găzduită de Centrul Austriac pentru Umanioare Digitale din cadrul Universității Karl-Franzens din Graz<sup>27</sup>. Carta albă a site-ului *digitale-edition.at* – KONDE White Paper oferă recomandări valoroase pentru proiectarea și realizarea edițiilor digitale.

### Text+

Text+ este o infrastructură dezvoltată de Consorțiul NFDI Text+ în colaborare cu Autoritatea de Control pentru Patrimoniul Cultural (Gemeinsame Normdatei). Nationale Forschungsdateninfrastruktur (NFDI), Infrastructura Națională pentru Date de Cercetare a Germaniei<sup>28</sup>, se concentrează asupra conservării și accesibilizării datelor textuale în trei domenii principale: colecții, resurse lexicale și ediții<sup>29</sup>. În domeniul edițiilor, sistemul de catalogare al

<sup>25</sup> Untner, Andorfer 2024.

<sup>26</sup> <https://github.com/arthur-schnitzler/pollaczek-epub> (accesat la 05.06.2024).

<sup>27</sup> [digital-humanities.uni-graz.at](https://digital-humanities.uni-graz.at) (accesat la 05.06.2024).

<sup>28</sup> [nfdi.de](https://nfdi.de) (accesat la 05.06.2024).

<sup>29</sup> [Text-plus.org](https://text-plus.org) (accesat la data de 05.06.2024).

Registrului Text+<sup>30</sup> oferă suport pentru organizarea și integrarea informațiilor din diverse surse, fiecare cu niveluri diferite de descriere. Un obiectiv esențial este crearea unui model de date care să permită descrierea diverselor tipuri de ediții.

### **De ce sunt utile edițiile digitale din perspectiva datelor**

În acest moment, datele sunt considerate a fi o resursă valoroasă a transformărilor pe care Deceniul Digital le impune. Instituțiile de memorie sunt generatoare de date foarte utile în procese complexe de creație, unde pot fi reutilizate sau pot reprezenta surse de inspirație.

Una dintre politicile Uniunii Europene privind datele are drept țintă crearea de Spații Comune pentru Datele Europene (*Common European Data Spaces*)<sup>31</sup>. În 2021, a fost propusă pentru prima dată crearea unui Spațiu de Date al Culturii. În documentul de politică a Comisiei Europene *Common European Data Spaces* sunt menționate cinci spații de date pentru cultură: Europeana pro, Eureka3D, 5Dculture, DE-BIAS, AI4Europeana<sup>32</sup>. Singurul spațiu de date dedicat culturii scrise este cel oferit de platforma Europeana Pro, ale cărei date pot fi accesate prin API-uri disponibile la următorul link: <https://pro.europeana.eu/discover-the-data/apis>. Conform Europeana, spațiul de date în domeniul culturii reprezintă *o infrastructură care permite partajarea de date de încredere între participanții la spațiul de date. Spațiile de date urmăresc să fie descentralizate și bazate pe standarde și să funcționeze cu o abordare deschisă, interoperabilă și durabilă. Spațiile de date pot fi specifice scopului sau sectorului, sau intersectoriale.*

Eforturile notabile ale Europeana sunt îndreptate către expunerea prin intermediul Colecțiilor Europeana a contribuțiilor pe care agregatorii de la nivel național le trimit organizației. O astfel de contribuție, de cele mai multe ori, este o reprezentare a unei opere digitizate. De exemplu, *Scrisoare, N. Batzaria către Marcu Beza, București, 12/25 august 1913* este prezentă în portalul național Culturalia.ro<sup>33</sup>, fiind preluat de Europeana<sup>34</sup>. Aceste eforturi sunt laudabile pentru expunerea în format digital a materiei prime pentru cercetători, mediul de educație și publicul general interesat.

Domeniul de cercetare al Umanităților Digitale este în acest moment unul vast și implică adesea fluxuri de lucru cuprinzând etape de prelucrare a datelor în diverse formate caracteristice unor variate formate de metadate. Lucrul

<sup>30</sup> [Registry.text-plus.org](https://registry.text-plus.org) (accesat la data de 05.06.2024).

<sup>31</sup> Common European Data Spaces | Shaping Europe's Digital Future 2024.

<sup>32</sup> Commission proposes a common European data space for cultural heritage | Shaping Europe's digital future 2021.

<sup>33</sup> <https://culturalia.ro/search/c293ba62-407a-4430-9bc6-5f2719809db4/view> (accesat la 05.06.2024).

<sup>34</sup> [https://www.europeana.eu/en/item/952/Culturalia\\_c293ba62\\_407a\\_4430\\_9bc6\\_5f2719809db4](https://www.europeana.eu/en/item/952/Culturalia_c293ba62_407a_4430_9bc6_5f2719809db4) (accesat la 05.06.2024).

cu date în formate standardizate conduce la posibilitatea de a reutiliza datele prin mijlocirea pe care metadatele o oferă. Pentru patrimoniul cultural digitizat, cel mai adesea, reutilizarea este caracterizată de aplicațiile, vizualizările și realizarea de materiale educaționale bogate în media, care reflectă realitatea științifică la zi.

Una dintre țintele importante ale tuturor inițiativelor politice europene este interoperabilitatea, iar datele nu fac excepție. Datele din domeniul cultural în general și cele care constituie materia primă a domeniului Umanioarelor Digitale în special sunt parte a spațiului de date din domeniul patrimoniului cultural. Edițiile digitale pot fi considerate seturi de date care prezintă un grad ridicat de interoperabilitate și standardizare.

### Concluzii

Hans Walter Gabler, în Cuvântul-înainte al lucrării *Digital Scholarly Editing: Theories and Practices*, spune că o ediție digitală științifică concepută ca un site de cercetare va fi astfel, în contextul Umanioarelor Digitale, un exemplu pentru ceea ce informatica se străduiește în prezent prin interacțiunea umană cu computerul<sup>35</sup>. În aceeași lucrare, Patrick Sahle pune în centru obiectul de lucru primar sub denumirea de „artefact cultural” (texte, obiecte sau imagini) și propune și o definiție de lucru pe care am amintit-o mai sus, în care este introdus termenul *reprezentare*. Acesta conduce către nivelul tehnic de prelucrare, care, în final, construiește contextul îmbogățit al *artefactului cultural* prin aranjamentul/codarea realizată la nivel digital (numită *paradigma digitală*), obținându-se o Ediție Digitală Științifică (*Scholarly Digital edition*). Sahle face o distincție importantă, menționând faptul că digitizarea unei ediții nu este echivalentă cu realizarea unei ediții digitale, dar dintr-o ediție digitală se pot deriva alte opere (*fluid publication*), cum ar fi o variantă facsimilată, o transcriere diplomatică și chiar variante care pot fi tipărite sau pot fi adaptate necesităților editoriale adaptate unor evenimente. În ceea ce privește recunoașterea eforturilor pentru realizarea edițiilor digitale, există opinia că fluxurile tradiționale de publicare ignoră această dimensiune a studiilor umanistice și, din această cauză, toți cei implicați în realizarea lor, de la editori până la programatori, nu au o direcție clară către recunoașterea meritelor pentru eforturile depuse (*10 Years of RIDE – still enjoying the ride 2024*)<sup>36</sup>. Realizarea de astfel de *sisteme informaționale* este, în mare măsură, rezultatul eforturilor și stăruințelor unor promotori care și-au asumat pionieratul, dar și al unor instituții și/sau consorții raliat în realizarea obiectivelor unor proiecte sau ale unor politici la nivel de stat, prin organismele de resort, în colaborare cu mediul academic.

---

<sup>35</sup> Driscoll, Pierazzo 2016.

<sup>36</sup> <https://www.i-d-e.de/10-years-of-ride> (accesat la 24.10.2024).

Edițiile digitale par a fi o prolemă legată de utilizarea tehnologiilor, dacă privim din perspectiva instrumentelor care netezesc experiența de lucru cu produsele finale, cum ar fi EVT (Edition Visualization Technology<sup>37</sup> sau TEI Publisher<sup>38</sup>. Trebuie să arătăm însă că acest lucru implică înțelegerea și utilizarea tehnologiilor dintr-un ecosistem digital complex, al cărui scop este extragerea de noi relații, înțelesuri și perspective din textele prelucrate. În ceea ce privește strict edițiile digitale, există deja materiale de instruire și familiarizare, așa cum sunt cele puse la dispoziție de DARIAH-CAMPUS prin Open Educational Resources for Digital Arts & Humanities. Foarte util este Digital Scholarly Editions: Manuscripts, Texts and TEI Encoding<sup>39</sup>, pentru un prim contact. Există și inițiative care au fost orientate către instruire, așa cum este DiXiT (Digital Scholarly Editions Initial Training Network)<sup>40</sup>, fiind o rețea internațională de instruire a cărei preocupare principală sunt edițiile digitale științifice (*Digital Scholarly Editions*).

Tehnologiile XML, în general, precum și aplicarea în lucrul cu TEI, în special, cer abilități suplimentare de la cercetătorii din domeniul Umanioarelor Digitale. Astfel că, înțelegerea tehnologiilor din ecosistemul XML este obligatorie pentru specialistul expus noilor oportunități de prelucrare și explorare a resurselor de text și nu numai. Există eforturi care țintesc educația, dar mai ales instruirea. Unul dintre cele mai bine organizate sub aspect pedagogic este programminghistorian.org, un exemplu notabil util în înțelegerea etapelor de prelucrare folosind tehnologiile XML fiind *Transforming Data for Reuse and Re-publication with XML and XSL*<sup>41</sup>.

România are câteva inițiative notabile pentru dezvoltarea Umanioarelor Digitale. Pentru relevanța inițiativelor ce vor urma și pentru relevanța pe plan național a domeniului, vom puncta câteva. Cea mai notabilă este crearea în 2021, drept unitate de cercetare în cadrul Universității Babeș-Bolyai, a Centrului de Științe Umaniste Digitale DigiHUBB (Transylvania Digital Humanities Centre), inițiat în 2015. A fost publicată și o revistă specializată intitulată *Digitalia*, dedicată domeniului Umanioarelor Digitale, inițiată în 2017, dar care a publicat ultimul număr în 2020<sup>42</sup>. Apoi, în 2022, a fost creat Centrul Digital Arts & Humanities în cadrul Universității Babeș-Bolyai (UBB), ca parte a proiectului *Rețeaua academică de cercetare de excelență STAR UBB*. Tot la UBB, există chiar un program postuniversitar în cadrul Facultății de Istorie și Filozofie, având două module, intitulat *Cultură umanistă și mediul digital*, sub

<sup>37</sup> <http://evt.labcd.unipi.it/> (accesat la 05.06.2024).

<sup>38</sup> <https://teipublisher.com/exist/apps/tei-publisher-home/index.html> (accesat la 05.06.2024).

<sup>39</sup> <https://teach.dariah.eu/course/view.php?id=32> (accesat la 05.06.2024).

<sup>40</sup> <https://dixit.uni-koeln.de/> (accesat la 05.06.2024).

<sup>41</sup> <http://programminghistorian.org/en/lessons/transforming-xml-with-xsl> (accesat la 05.06.2024).

<sup>42</sup> <https://dighubb.centre.ubbcluj.ro/journal/index.php/digitalia> (accesat la 05.06.2024).

coordonarea lectorului asociat Rada Varga<sup>43</sup>. Universitatea din București, prin Institutul de Cercetare al Universității din București (ICUB), desfășoară activități în domeniul Umanioarelor Digitale prin Institutul pentru Cercetare în Umanioare<sup>44</sup>. Activitatea poate fi urmărită online<sup>45</sup>. Universitatea din București are, începând cu anul 2019, și un program de masterat numit *Digitalizarea în științe umaniste / Digital Humanities*, în cadrul Facultății de Limbi și Literaturi Străine (domeniul fundamental Filologie), coordonat de lect. univ. dr. Anca Dinu. Universitatea de Vest din Timișoara are un centru de cercetare CODHUS (Centre for Corpus Related Digital Approaches to Humanities, Centrul de Cercetare și Implementare a Metodelor Digitale Bazate pe Lingvistica de Corpus cu Aplicabilitate în Științele Umaniste)<sup>46</sup>. Intenția cercetătorilor din acest centru este să creeze o „casă a codurilor digitale pentru umaniști”. Universitatea din București împreună cu Universitatea de Vest din Timișoara (UVT) organizează o Școală de vară sub egida CIVIS, intitulată generic *Cultural Heritage and Digital Humanities*, începând cu anul 2021, dar și o conferință intitulată *Recent Advances in Digital Humanities*, aflată la a doua ediție, având drept gazdă CODHUS. CODHUS este și un centru de instruire dedicat competențelor digitale.

Încheiem prin a sublinia necesitatea de a privi competențele digitale de lucru cu tehnologiile de markup ca pe un vehicul care ar garanta suportul instituțional necesar în cadrul instituțiilor de învățământ superior și al institutelor de cercetare din domeniul umanioarelor<sup>47</sup>.

#### Abrevieri bibliografice:

- Altenhöner, Berger, Bracht, Klimpel, Meyer, Neuburger, Stäcker, Stein 2023 – R. Altenhöner, A. Berger, C. Bracht, P. Klimpel, S. Meyer, A. Neuburger, T. Stäcker, R. Stein, *DFG Practical Guidelines on Digitisation. Updated Version 2022*, Zenodo, disponibil la: <https://doi.org/10.5281/zenodo.7561148>
- Alto 2024 – *ALTO: Technical Metadata for Layout and Text Objects* (Standards, Library of Congress) [online] (2024) disponibil la: <https://www.loc.gov/standards/alto/>
- Chitez, Rogobete, Foițoș 2020 – M. Chitez, R. Rogobete, A. Foițoș, "Digital Humanities as an Incentive for Digitalisation Strategies in Eastern European HEIs: A Case Study of Romania", în A. Curaj, L. Deca, R. Pricopie (eds.), *European Higher Education Area: Challenges for a New Decade*, Cham, Springer International Publishing, p. 545-564, disponibil la: [https://doi.org/10.1007/978-3-030-56316-5\\_34](https://doi.org/10.1007/978-3-030-56316-5_34)

<sup>43</sup> <https://hiphi.ubbcluj.ro/studii/postuniversitar.html> (accesat la 05.06.2024).

<sup>44</sup> <https://irhunibuc.wordpress.com/digital-humanities/> (accesat la 05.06.2024).

<sup>45</sup> <https://icub.unibuc.ro/research/research-seminars/digital-humanities> (accesat la 05.06.2024).

<sup>46</sup> <https://codhus.projects.uvt.ro/?lang=ro> (accesat la 05.06.2024).

<sup>47</sup> Chitez, Rogobete, Foițoș 2020.

- Cohen, Rosenzweig 2006 – D. J. Cohen, R. Rosenzweig, *Digital history: A guide to gathering, preserving, and presenting the past on the web*, University of Pennsylvania Press, disponibil la: <https://chnm.gmu.edu/digitalhistory/digitizing/index.html>
- Commission proposes a common European data space for cultural heritage | Shaping Europe's digital future 2021 – Commission Proposes a Common European Data Space for Cultural Heritage | Shaping Europe's Digital Future, 2021, disponibil la: <https://digital-strategy.ec.europa.eu/en/news/commission-proposes-common-european-data-space-cultural-heritage>
- Common European Data Spaces | Shaping Europe's Digital Future 2024 – Common European Data Spaces | Shaping Europe's Digital Future, 2024, disponibil la: <https://digital-strategy.ec.europa.eu/en/policies/data-spaces>
- Driscoll, Pierazzo 2016 – M. J. Driscoll, E. Pierazzo, *Digital Scholarly Editing: Theories and Practices*, 2016, disponibil la: <https://doi.org/10.11647/obp.0095>
- Malliaros, Vazirgiannis 2017 – F. D. Malliaros, M. Vazirgiannis, "Graph-based text representations: Boosting text mining, NLP, and information retrieval with graphs", in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing: Tutorial Abstracts*, Copenhagen, 2017, disponibil la [https://fragiskos.me/projects/graph\\_text\\_tutorial/Graph-based\\_Text\\_Representations\\_Tutorial\\_EMNLP\\_2017.pdf](https://fragiskos.me/projects/graph_text_tutorial/Graph-based_Text_Representations_Tutorial_EMNLP_2017.pdf)
- Neuber, Henny-Krahmer, Assmann, Scholger 2024 – F. Neuber, U. Henny-Krahmer, B. Assmann, M. Scholger, "I-d-e/ride: Datasets of RIDE (issues 1-18)", 2024, disponibil la: <https://doi.org/10.5281/zenodo.12794244>
- Pletschacher, Antonacopoulos 2010 – S. Pletschacher, A. Antonacopoulos, "The PAGE (Page Analysis and Ground-Truth Elements) Format Framework", in *2010 20th International Conference on Pattern Recognition*, p. 257–260, disponibil la: <https://doi.org/10.1109/ICPR.2010.72>
- Reul, Christ, Hartelt, Balbach, Wehner, Springmann, Wick, Grundig, Büttner, Puppe 2019 – C. Reul, D. Christ, A. Hartelt, N. Balbach, M. Wehner, U. Springmann, C. Wick, C. Grundig, A. Büttner, F. Puppe, "OCR4all – An Open-Source Tool Providing a (Semi-)Automatic OCR Workflow for Historical Printings", in *Applied Sciences*, 9 (22), 4853, 2019, disponibil la: <https://doi.org/10.3390/app9224853>
- Sahle, Vogeler 2024 – P. Sahle, G. Vogeler, *Criteria for Reviewing Scholarly Digital Editions*, version 1.1, 2024 disponibil la: <https://www.i-d-e.de/publikationen/weitereschriften/criteria-version-1-1/>
- Terras, Nyhan, Vanhoutte 2013 – Melissa Terras, Julianne Nyhan, Edward Vanhoutte (editors), *Defining digital humanities: a reader*, Farnham, 2013
- Untner, Andorfer 2024 – L. Untner, P. Andorfer, *Using GitHub for digital editions. From Transkribus to static websites*, disponibil la: <https://doi.org/10.5281/zenodo.11609624>
- van Zundert 2024 – J. van Zundert, *History of Digital Scholarly Editions*, 2024, disponibil la: [https://www.youtube.com/watch?v=JBF\\_VNjoHVQ](https://www.youtube.com/watch?v=JBF_VNjoHVQ)