



HUNYADI ZSOLT

Szegedi Tudományegyetem Bölcsész- és Társadalomtudományi Kar

Középkorászok a digitális térben

Honnan jövünk és hová tartunk?

ABSTRACT. *Medievalists in the Digital Space – Where Did We Come From, and Where Are We Headed?* The study gives a brief overview of the primary stages of the evolution of the Internet, with a particular focus on the digital infrastructures of medieval history research. It spotlights the main characteristics of the road from the “ancient times” now referred to as Web 1.0 to the current Web 3.0/4.0 environment supported by artificial intelligence, showcasing the primary paradigm shifts for the technology and its users. The second part of the study follows the history of online sources, databases, and technologies available for medieval history research from the 1990s to the present day, examining the development of the early digital catalogs, specialized databases, and semantic data links. A primary question raised by the study is whether the development of web-based technologies had a direct effect on the methodological renewal of medieval history research.

Keywords: *web history, online databases, digital research of medieval history, information infrastructure, semantic web*

Az alábbi tanulmány röviden áttekinti a web fejlődésének főbb szakaszait, különös tekintettel a középkorkutatás digitális infrastruktúráira. Felvillantja a ma már web1.0-ként emlegetett „őskortól” a mesterséges intelligenciával támogatott jelenkori web3.0/4.0-ás környezetig tartó folyamat legfőbb jellemzőit, felvillantva a legfontosabb technológiai és felhasználói paradigmaváltásokat. A tanulmány második része a középkori online forrásgyűjtemények és adatbázisok, illetve technológiák elérhetőségének történetét követi nyomon a ’90-es évektől napjainkig: a korai digitális katalógusok, tematikus adatbázisok, illetve a szemantikus adatkapcsolatok kialakulását. A tanulmány egyik fontos kérdésfeltevése az, hogy a webes technológiák fejlődése milyen módon hatott a középkorkutatás módszertani megújulására, ha egyáltalán hatott. Ehhez példaként az Anjou-kori oklevélregeszták készülő digitális korpuszát fogjuk használni.

A tanulmány a címében ígérteknél bizonyosan kevesebbet fog nyújtani olvasójának. Nem vállalkozik teljes áttekintésre, már csak azért sem, mert az messze túlfeszítené ezen írás kereteit, másrészt a célját is másként határozza meg: trendeket fog felvázolni, ami részben helyzetértékelésen alapul, illetve a változások irányát próbálja feltérképezni. Az informatika felkent papjai felé előre is apológiával fordul, mert a bölcsészeket komolyan próbára tevő technológiák szabatos leírásában nem lesz feltétlenül makulátlan,¹ az érthetőség oltárán időn-

¹ Ugyanakkor igyekszik elkerülni azokat a szakmai közbeszédben is jelen lévő fogalmi csapdákat, mint például az AI/MI elnevezés korlátlan, ily módon pedig sokszor alaptalan használata. Az esetek nagy részében pontosabb a (felügyelt vagy nem felügyelt) gépi tanulás (ML) kifejezést használni.

ként áldozatokat fog hozni. A teljességre törekvésnek már eleve az is határt szabna, hogy az elmúlt évtizedek digitális/online projektjeinek jelentő része efemer életűnek bizonyult – és ez máris a trendek egyikét villantja fel. De ne szaladjunk előre!

Első pillantásra úgy tűnhet, hogy az *Aetas* (est. 1985) már évek óta jelen volt a történelemtudományi folyóiratok „piacán”, amikor Tim Berners-Lee és a CERN útjára indította az információk megosztásának és összekapcsolásának decentralizált protokollját.² A történet azonban ennél egy kissé korábban kezdődött, mi több szegedi, sőt még középkorász vonatkozása is van az események láncolatának. Az 1970-es évek közepén, Horváth Iván nevével fémjelezve, a szegedi JATE-n kezdődtek a régi magyar költészet számítógépes inventáriumának munkálatai, és ennek a vállalkozásnak egy fontos mérföldköve lett 1984, amikor telefonvonalon keresztül online adatbázis-kapcsolat jött létre a JATE és az MTA Irodalomtudományi Intézete között.³ Szintén a '70-es években kezdődött el a Magyar Országos Levéltár (akkor MOL, ma már MNL OL) Mohács előtti gyűjteményének számítógépes feldolgozása elméleti kereteinek kidolgozása Borsa Iván koncepciója alapján, amelynek megvalósítása 1983-ban vette kezdetét.⁴ Az ún. kiadói mutató pedig már a Web „születése” előtt is elérhető lett egy régóta létező, de csak 1983-ban véglegesített protokoll, a *telnet* segítségével. A CERN által kidolgozott protokoll indulásával egyidőben más technológiák is jelen voltak, ezek közül feltétlenül kiemelendő a Minnesotai Egyetemen kifejlesztett *Gopher*, amely egy hierarchikus, szöveges felületen kínált hozzáférést navigációs menükön keresztül kiszolgáló szervereken elhelyezett fájlokhoz. Ezzel szemben a web-protokoll hiperhivatkozásokkal kapcsolta össze az elérhető tartalmakat, ezekhez ráadásul grafikus felületet is kínált, és ez nagyban elősegítette a multimédiás tartalmak elérését is, amire a *Gopher* nem kínált lehetőséget.⁵ Ugyanakkor megkerülhetetlen tény, hogy az első években a *Gopher* nemcsak komoly versenytársa volt a web-nek, de a szöveges tartalmakhoz hozzáférést kereső felhasználói tábor, például a történészek elsősorban ebben a hierarchikus keretrendszerben bolyongtak. Mi több, már ekkor felsejlett annak lehetősége, hogy az online hozzáférés egy új (digitális) kánonképződés egyik első lépése lehet.⁶

A '90-es évek közepére eldőlt a verseny, és a HTML-alapú web(1.0) fokozatosan teret nyert a felhasználók egyre szélesebb táborában, ráadásul a már korábban is létező protokollok egy részét (*telnet*, *ftp*) nem is szorította ki, hanem az URL révén gyakorlatilag integrálta saját keretrendszerébe. Ugyanakkor miközben ez hatalmas technológiai előre lépésnek tekinthető, a mai felhasználók erősen firtorognának a három évtizeddel ezelőtti lehetőségek, illetve korlátok láttán. A legfontosabb jellemzője ennek a mátrixnak a statikusság: a tartalmak „csak olvashatóak” voltak, az interaktivitás még csak nyomokban fordult elő. A tartalom-gyártás (adat-aggregálás) nagy buzgalommal zajlott, de ezek jellemzően intézményi honlapokon váltak elérhetővé, és ún. keresőmotorok (Yahoo!, Altavista) segítségével váltak

² Cf. World Wide Web Consortium - www.w3.org

³ A régi magyar vers repertórium (Répertoire de la poésie hongroise ancienne - RPHA) - f-book.com/rpha/v7/index.php.

⁴ Rác: *A Magyar Országos Levéltárban őrzött 1526 előtti levéltári anyag épülő adatbázisa*, 181.

⁵ A szakma egy része szerint nem pusztán a technológiai „előny” segítette a web-protokollt a végső győzelemhez, hanem az a piaci tény, hogy az 1991-ben megjelenő *Gopher* népszerűsége olyan gyorsan nőtt, hogy az egyetem ezt megpróbálta gazdasági haszonná transzformálni, és licenc-díjat vezetett be, míg a CERN-protokollja nyílt hozzáférésű maradt. Poulter: *The design of World Wide Web search engines*, 133–134.

⁶ A „mit és mikor” digitalizálunk versenyben a Biblia és Shakespeare összes drámája sokáig toronymagasan vezetett, de aránylag hamar felkerült a *Gopher* rendszerébe az *Index Thomisticus*, illetve a kora középkor zenei adatbázisa is.

felfedezhetővé. Persze illetlen lenne alábecsülni annak jelentőségét, hogy mekkora segítséget jelentett a történészeknek, hogy külföldi könyvtárak és levéltárak katalógusaiban és inventáriumaiban is tudtak kutakodni, és ha a közvetlen tartalmakhoz nem is tudtak még hozzáférni, de egy-egy – jellemzően rövid – külföldi kutatást hatékonyan elő tudtak készíteni itt-honról. Mi több, az online tér tágulásával szoros korrelációt mutatott a digitalizáló eszközök (scannerek, digitális fényképezőgépek) megjelenése, fejlődése és terjedése a kereskedelmi forgalomban.⁷

Az is fontos nehézséget jelentett, hogy a generált tartalmak szorosan kötődtek az adott intézményhez (kögyűjtemény, egyetem, etc.), így azok tartós elérhetősége, fellelhetősége függött az adott intézmény működésétől, esetleges átszervezésétől, az adott projektek kifutási idejétől. Az online térben kutató történész szisztematikusan építette online könyvjelzőgyűjteményét, és a rémálmok egyike a „broken link”, avagy a 404-es hiba volt, amely az adott tartalom (megadott címen való) elérhetetlenségét jelezte, ami kutatómódszertani problémákat is felvetett, hiszen az eltűnt vagy elérhetetlen tartalmakra való hivatkozás a rekurzivitás elé gördített akadályokat. Ez pedig csak fokozta a HTML-oldalakra való hivatkozások egyébként is létező problematikáját: nem álltak rendelkezésre megszokott referenciális pontok, azaz a klasszikus oldalszámok, a tartalom megváltozhat(ott), mire az olvasó eljut annak olvasásáig. Részben ez utóbbi kiküszöbölésére terjedt el a hivatkozásokban az utolsó hozzáférés pontos dátumának rögzítése és megadása.⁸

Jelen tanulmány bizonyosan nem fog állást foglalni a web-fejlődés kronológiáját illetően, de arra talán illik utalni, hogy e téren nincs teljes egyetértés ezen sokszereplős szakmai közegekben. A legelfogadottabb avagy legelterjedtebb korszakolás szerint az ezredforduló éveire (leginkább 2001–2005 közé) tehető a web2.0 születése és térnyerése. A korszakolás már csak azért is nehéz feladat, mert pontosan meg kellene határozni azokat az indikátorokat, amelyek segítségével egyértelműen ki lehetne jelölni a határokat. De kénytelenek vagyunk beérni trend-jelző jellemzőkkel. A folyamatosan táguló online tér már nem érte be a „mindent (is) digitalizáljunk” imperatívusszal, hanem a (költség)hatékonyabb elérés és felhasználás felé mozdult el. Ennek látható jelei a tartalomgyártás és -megosztás új útjai lettek: például közös(itett) lekérdezőfelületek⁹ és repozitóriumok (tematikus adatbázisok) kialakítása. De szokás ezt a korszakot a keresőmotorok diadalaként is aposztrofálni, hiszen az online térben elhelyezett és hozzáférhetővé tett tartalmak hatékonyabb megtalálását és elérését nagyban segítette a tartalomszolgáltatók oldaláról induló keresőoptimalizálás (SEO). Ennek végletesen leegyszerűsített példái lehetnek akár ennek a lapszámnak a tanulmányai, amelyben szerepelnek kulcsszavak (*keywords*), segítve az automatizált keresőket, hogy ne értsék félre a szöveget vagy akár annak absztraktját.

Ennek a „korszaknak” persze legalább ennyire fontos jellemzője volt a dinamikus tartalomszolgáltatással (CMS-rendszerek) párosuló interaktivitás kialakulása és rohamos terje-

⁷ Ezeket az eszközöket már jóval korábban kifejlesztették, de amíg tömegesen nem kezdtek el terjedni, – megítélésünk szerint – nem játszottak komoly szerepet a tömeges digitalizálásban.

⁸ Nehéz megállni annak a jó ideje terjedő abszurd gyakorlatnak a kárhoztatását, midőn nyomtatásban megjelent és digitalizált (vagy a printtel egyidőben generált PDF-formátumú) könyvet, tanulmányt a szerzők némelyike ily módon hivatkozta, és gyakran a pontos bibliográfiai adatok megadása helyett pusztán a web-linket és a letöltés dátumát adja meg – mintha ez utóbbinak bármi jelentősége lenne.

⁹ Például KVK – Karlsruhe Virtual Catalog (kvk.bibliothek.kit.edu), vagy idehaza a MOKKA – Magyar Országos Könyvtári Közös Katalógus (mokka.hu).

dése.¹⁰ Ennek egyik beszédes példája a web1.0-ás, rögzített tartalmú lexikonok (például *Britannica online*)¹¹ és a szabadon szerkeszthető, megosztott tartalmat kínáló „webkettes” Wikipedia.¹² A terjedő (PHP, MySQL-alapú) webes applikációk lassan, de biztosan kiszorították a korábbi technológiákat (*ftp, telnet*), ahogyan az e-mailezést is fokozatosan háttérbe szorították a szaporodó fórumok és blogok, amelyek kialakították a közösségi média új világát – a középkorász szakma számára is. Sőt, emellett is lehetne érvelni, hogy itt is elhelyezhető egy korszakhatárt jelölő jelzőkaró, hiszen nagyjából 2010-től nyertek teret az összekapcsolt nyílt adathalmazok (*Linked Open Data* – LOD) és a metaadat-szabványok rögzülésével és terjedésével a szemantikus web már legalább másfél évtizede kopogtatott a kutatók és így módon a történészek ajtaján. Kizárólagos érvényű idővonal bizonyosan nem rajzolható, mert a *cutting edge technology* térnyerése nem volt sem lineáris, sem pedig egyenletes a szakmában.¹³ Ezt példázhatja az RDF-adatmodell (*Resource Description Framework*) születése, illetve elterjedése. Az első technológiai ajánlás már 1999-ben megszületett, de az RDF 1.0 specifikációt 2004-ben, a jelenlegi szemantikus web-struktúrát meghatározó 1.1-et 2014-ben rögzítették, a közeli jövőt meghatározó RDF 1.2 koncepciója pedig már formálódik.¹⁴

Függetlenül attól, hogy az intézményi adatbázisok közötti interoperabilitást lehetővé tévő vagy a közgyűjteményi és kutatási adatbázisok metaadat-hálózatai átjárhatóságát (data linking/harvesting) biztosító technológiákra, ontológiákra (*Web Ontology Language* – OWL) a web3.0 vagy 4.0 nevet aggatjuk, abban többnyire egyetértés mutatkozik, hogy a post-Covid időszak online terét a gépi tanulás,¹⁵ a nagy nyelvi modellek (*Large Language Models* – LLM) létrehozása és elérhetővé tétele, a különböző névterek kialakítása fémjelzi, ahol már átvitt értelemben is a gépek hálózatáról beszélünk, ha a webet jellemezni akarjuk. A tartalomszolgáltató már nem tulajdonosa, de jellemzően még csak nem is megosztója, hanem elsősorban felügyelője a tartalmaknak, amelyeket okosapplikációk segítségével lehet személyre szabni. A FAIR-adatelvek mentén a nyílt hozzáférést (OA) biztosító technológiákat ma már – a teljesség igénye nélkül – az adatbányászat, az adatvizualizáció, a hálózatkutatás és az interaktív térképek létrehozása jellemzi.

A tanulmány címében foglaltak szerint illik e ponton ráközelítenünk arra a kérdésre, hogy a medievisták hol helyezkednek el ebben a sokparaméteres mátrixban. Mi több, mint mindennek, ennek is története van, sőt a kezdetekre már fentebb utaltunk: a MOL/MNL OL középkori állománya számítógépes feldolgozásának terve már az 1970-es évek elején megszületett, a munkálatok pedig a '80-as évek elején el is kezdődtek.¹⁶ A '90-es évek elejére a *Diplomatikai Levéltár* (MOL DL) állományának regesztáit is el lehetett érni *telnet* proto-

¹⁰ Ezt természetesen új technológiák, protokollok megjelenése és terjedése alapozta meg, illetve támogatta: például Flash, Java és a HTML-nél jóval nagyobb rugalmasságot biztosító XML. Ezekkel nem az olvasót akarjuk vegzálni, de arra fontos utalni, hogy a fejlődés irányát időnként komolyan befolyásolják az adott pillanatban rendelkezésre álló technológia határai. Például az ígéretes Flash-technológia biztonsági sérülékenysége miatt húzódott vissza napjainkra.

¹¹ www.britannica.com

¹² hu.wikipedia.org

¹³ Részletezés nélkül szeretnénk utalni arra, hogy a „technológiai ugrás”, azaz a fejlődési fokok kiagyása ezen a területen különös jelentőséggel bír.

¹⁴ A 2025 októberében publikált munkaverziót ld.: www.w3.org/TR/rdf12-concepts

¹⁵ Az AI/MI kifejezést a bevezetésben említettek alapján továbbra is kerülnénk. Lásd még: *Model: Supervised and Unsupervised*, 158.

¹⁶ Lásd: Rácz: *A Magyar Országos Levéltárban őrzött 1526 előtti levéltári anyag épülő adatbázisa*, 181.

kollon keresztül. De a távoli adatbázisok kissé lassabban váltak elérhetővé a középkort kutatók számára (is).

A már a '80-as évektől létező könyvtári OPAC-rendszerek a '90-es évek első felétől fokozatosan web-alapúvá váltak (OPAC, ILS), és némi fáziskéséssel követték őket a levéltárak inventáriumai (ISAD, ISAAR) is. Ugyanakkor kisebb vagy helyi gyűjtőkörű, de nagyon fontos középkori gyűjteménnyel rendelkező intézmények levéltári állományait még csak elvételezhető online kutatni még a fondjegyzékek szintjén is. És a fejlettség fent említett (részben földrajzi) egyenetlensége időnként generált érdekes helyzeteket. A johannita lovagrend középkori levéltára, komoly hányattatások után, 1530-ban Máltára került, és ma az 1776-ban alapított Máltai Nemzeti Könyvtár állományában őrzik. Az 1990-es évek végén azonban az az érdekes helyzet állt elő, hogy a minnesotai *Hill Monastic Manuscript Library* (HMML)¹⁷ bencésai által az 1960-as években lemásolt – Máltán őrzött – anyagot Collegeville-ben feldolgozták, az inventáriumokat digitalizálták és online elérhetővé tették. Bár az anyaggal először 1998-ban dolgoztam, még egy 2002-es máltai *hungarica*-kutatás során is az USA-beli online katalógust¹⁸ lehetett használni az egyébként európai gyűjteményben őrzött állományok kutatásához.

Az elképesztő tempóban digitalizált állományok elérése ugyanakkor jellemzően még nem az online térben történt. Elsősorban fizikai hordozókon (floppy lemezek, CD-ROM, DVD-ROM) és csak bizonyos platformokon működő adatbázisok jelentek meg hónapról hónapra, időnként az eltérő számítógépes rendszerekkel való kompatibilitás vagy az átjárhatóság legkisebb reménye nélkül. Ezek részben forráskiadványok voltak, de időnként kutatási segédletek is korongra kerültek, mint például a nemzetközi együttműködésben (IRHT–HMML–BNF) készült *In principio* kéziratkatalógus¹⁹ vagy a leeds-i egyetem középkorászai által már a '60-as évek második felétől publikált biannuális *International Medieval Bibliography* (IMB),²⁰ amelynek 1995-től éves rendszerességgel jelentek meg friss CD-kiadásai.²¹ Az *In principio* ráadásul tükrözi a korábban jelzett kronológiát: 1993-ban jelent meg az első korong, majd az adatbázis 2009-ben költözött az online térbe, akkor már több mint egymillió középkori kézirat digitális katalógusaként. Míg Európában a *Brepols*, addig a középkorászokat leginkább érintő digitalizálás hazai zászlóshajója kétségtelenül az *Arcanum* kiadó volt,²² amely vállalkozás nagyon sok középkorász életét könnyítette meg CD-ROM és DVD-ROM hordozókon publikált adattáiraival.²³ A technológiai fejlődés e téren is fontos szerepet játszott, amit jól jellemez egy nemzetközi példa: A meglehetősen korán digitalizált *Patrologia Latina* sorozat²⁴ 221 kötete 5 CD-korongon fért el, így ezek használata vagy nagyon körülményes volt, vagy speciális eszközt, például ún. CD-tornyot igényelt, amellyel szimultán kezelhetővé vált a teljes állomány. És messze nem csak a kezelhetőség, illetve a keresettség okozott nehézségeket: a digitális kiadványok borsos ára sajátos versenyhátrányt teremtett az addig sem homogén kutatói körben.

¹⁷ Ma már *Hill Museum & Monastic Library* néven ismert intézmény.

¹⁸ hmml.org/collections/repositories/malta/national-library-of-malta

¹⁹ apps.brepols.net/inpr/Help/EN/INPR_manual_EN.pdf

²⁰ about.brepols.net/imb

²¹ A kiadvány nyomtatott változata 2017-ben megszűnt. ahc.leeds.ac.uk/medieval-research-innovation/doc/international-medieval-bibliography

²² www.arcanum.com/hu

²³ Bár kisebb vitákat generált a *Turul* folyóirat köteteinek oldalszámok nélküli folyó-szöveggént történő első publikálása (2004) vagy a sok-sok elütést tartalmazó Fejér György-féle *Codex diplomaticus* megtisztítása a hibáktól, amely tulajdonképpen új kiadványt hozott létre.

²⁴ Migne: *Patrologia Latina database*, i. m.

A web2.0 a 2000-es évek közepétől ugyanakkor valamelyest közelebb hozta az „ígéret földjét”, amennyiben mind a könyvtári, mind pedig a levéltári katalógusok esetében egyre szaporodtak az integrált könyvtári/levéltári²⁵ katalógusok, amelyek érzékelhetően növelték a keresések hatékonyságát. Fokozatosan szaporodtak a tematikus adatbázisok (például *Monastic Matrix*,²⁶ *Mapping Gothic France*,²⁷ *Regesta Imperii Online*²⁸). Megindult az online térbe költözési hullám: már nemcsak egyes intézmények jelentek meg gyűjteményeikkel, hanem részben EU-s finanszírozás révén látványosan megnőtt a nemzetközi vállalkozások száma és láthatósága (MoM,²⁹ DigiMap,³⁰ Europeana³¹). Az online hozzáférést adatbázisokhoz (Brepols, JSTOR) kialakultak előfizetési modellek, amelyekhez egyénileg vagy konzorciális formában a magyar kutatóhelyek is hozzáférhettek. A hazai digitális tartalomgyártók és -szolgáltatók sem tétlenkedtek. Látványos lendülettel nőni kezdett az online módon elérhető középkori vonatkozású források és feldolgozások száma. Ennek a folyamatnak egy fontos motorja továbbra is az *Arcanum* maradt, amely cég platformján a MOL/MNL OL állományát (oklevelek nagy felbontású digitális fotóit), illetve levéltári segédleteket (például fondjegyzékek), kézikönyveket és forráskiadványokat (okmánytárak, regeszta-kiadványok)³² elérhetővé tettek, jellemzően kétrétegű PDF-állományokban: 2014-ben elindult a digitális közgyűjteményi portál (*Hungaricana*).³³ Kétségtől fontos mérőszámok számított a szakma képviselői, mindenekelőtt természetesen az okleveles forrásokat kutató és feltáró történészek számára. Mindazonáltal a kereslet-kínálat törvényszerűsége itt is megjelent. Minél nagyobb adat-integráció jött létre, a kutatók annál nagyobb igényekkel léptek fel. Bár szinte minden elérhető vált az online térben, az egymással összefüggő adatok rendkívül széttagoltan voltak jelen a rendszerben, és minden egyes kutatás esetén a középkorászok újra és újra létrehozták, sőt definiálták az adatkapcsolatokat a facsimile fotó, a teljes szövegű kiadás, a regeszta, fordítások és kivonatok között, amelyek ugyanazt az entitást reprezentálták: egy eredetiben, átírásban, másolatban, tartalmi átírásban, említésben vagy ezek valamilyen kombinációjában fennmaradt dokumentumot. A hatékonyság növelése önmagában nem hozott módszertani áttörést, és nem sok teret engedett a gépi feldolgozásnak.

Ha nem is hozott azonnali nagy változást, de mindenképpen új megközelítésre kínált lehetőséget a fent említett Monasterium.net (MoM) keretében kialakított virtuális levéltár, amely 2006-ban még pusztán osztrák kezdeményezésként indult, de nagyon hamar nemzetközi együttműködéssé nőtte ki magát. A vállalkozás több szempontból is komoly nívóvumnak számított ezen a tudományterületen. Egyrészt a már említett web2.0-ás trendbe illeszkedett, azaz nemzetközivé válása az integrált levéltári portálok egyik fontos ernyőszervezete, az *Icarus (International Centre for Archival Research)*³⁴ égisze alatt valósult meg. Másrészt, és ez már erősen a web3.0 felé mutató paraméter, a MoM a szó szoros értelmében „virtuális” levéltárrá vált, hiszen az adatokat nem inkorporálta egyetlen fizikai adatbázisba, hanem a

²⁵ A fent említett könyvtári integrált katalógusokon túl mindenképpen megemlítené a *The Archives Portal Europe* (korábban APEx): www.archivesportaleurope.net

²⁶ arts.st-andrews.ac.uk/monasticmatrix

²⁷ mcid.mcah.columbia.edu/mapping-gothic

²⁸ opacfix.regesta-imperii.de

²⁹ www.monasterium.net/mom/home

³⁰ digimap.edina.ac.uk

³¹ www.europeana.eu

³² library.hungaricana.hu/hu/collection/kozepkori_magyar_okmanytarak

³³ www.hungaricana.hu/hu A portálon keresztül elérhető adatbázisok jóval szélesebb kört fednek le, itt elsősorban a kifejezetten középkori relevanciájú elemekre utaltunk.

³⁴ www.icar-us.eu/en

2006-ban kikristályosodó *Linked Open Data* (LOD)³⁵ koncepció ösvényén haladva egymástól távol tárolt adatbázisokat kapcsolt össze, és csak a metaadatokat tárolta/kezelte saját „rendszerében”. De középkori oklevelekről lévén szó érdemes egy pillanatra megállni, és újabb szálakat bekapcsolni, amelyek végül elvezetnek az adatok hálózatához (*web of data*) és a gépi tanuláshoz.

Már a '80-as évek végén létrejött egy szabvány (TEI: *Text Encoding Initiative*), amely elsősorban, de nem kizárólag a humán tudományok által vizsgált szövegek leírására és gépi feldolgozására kínált lehetőséget. Az eredetileg az SMGL-re (*Standard Generalized Markup Language*) épülő szabvány 2002-re nagyságrendekkel rugalmasabbá vált, amikor XML-alapokra (*eXtensible Markup Language*) transzformálták. Ezt a lehetőséget aknázták ki egy 2004-ben a müncheni Ludwig Maximilian Egyetemen Georg Vogeler³⁶ vezetésével indult vállalkozás, amely a TEI kereteit szűknek találta okleveles források leírására, és javaslatot tett egy új szabvány, a CEI (*Charters Encoding Initiative*)³⁷ létrehozására. A CEI szabvány kialakítása 2004 és 2009 között zajlott,³⁸ és a MoM-projekt lett az új szabvány létrehozásának egyik fő hasznélvezője. De koránt sem az egyetlen! A '90-es évek elején a Torontói Egyetem a Michael Gervers vezetésével már 1975 óta működő DEEDS-projektet (*Documents of Early England Data Set*)³⁹ a világhálóra költöztette, és ez a digitális vállalkozás nem pusztán 11–14. századi angol magánjogi oklevelek adatbázisa kívánt lenni, hanem megoldást próbált kínálni arra a nyomasztó problémára, hogy ezen oklevelek nagy része keltezetlenül maradt az utókorra. Az eredetileg Oracle-adatbázisként létrehozott korpuszt, amelyet 1999-ben Magyarországon is bemutattak,⁴⁰ a 2000-es évek közepére XML-alapúvá alakították, majd 2007-től hasznosítani kezdték a CEI-szabványt. A kérdés persze joggal fogalmazódott meg már negyedszázaddal ezelőtt: mit adtak nekünk a kanadaiak? Mert nemcsak a magyar, de jellemzően a közép-európai középkori magánjogi írásbeliség által létrehozott és az utókorra maradt okleveles korpusz is többnyire datált, ha nem is mindig napi keltezéssel. A torontóiak által kidolgozott módszer ugyanakkor nemcsak hamis oklevelek azonosítására volt alkalmas, de a matematikusok és statisztikusok által kidolgozott vagy adaptált módszertan⁴¹ nagy nyelvi korpuszok vizsgálatára kínált lehetőséget a 21. század második évtizedére, és ez már a web3.0 korszaka felé vezetett, a természetes nyelvi feldolgozás (NLP) előszobájának bizonyult.

Ahhoz persze, hogy az elemezhető, vizsgálható szövegekig eljussunk, a középkori kéziratosokat el kell olvasni. Ezek automatizálásának igénye már a web2.0 idején megfogalmazódott, de a technológiai háttér megteremtése időigényesnek bizonyult. A digitális paleográfia

³⁵ Alapelveit a web-ötletet is jegyző Tim Berners-Lee fogalmazta meg. Lásd például: data.europa.eu/sites/default/files/d2.1.2_training_module_1.2_introduction_to_linked_data_en_edp.pdf. Az LOD természetesen nem egyetlen technológia, hanem több szabvány kombinációja, amelyek részletes bemutatásától kímélnénk az olvasót.

³⁶ Ma már a Grazi Egyetem Digitális Bölcsészeti professzoraként koordinál digitalizálási kutatásokat.

³⁷ www.cei.lmu.de/project

³⁸ E sorok írója is szerepet vállalt a szabvány végleges tartalmának kialakításában egy 2007-es müncheni workshop révén: *Digital Diplomats - Digitale Diplomatie: Historical research with medieval charters in a digital world - Die historische Arbeit mit Urkunden in der digitalen Welt*, München, 2007. február 28. – március 2.

³⁹ deeds.utoronto.ca

⁴⁰ Gervers: *The DEEDS Project and the Development of a Computerized Methodology*, 13–36. Lásd még: Gervers et al.: *The Dating of Undated Medieval Charters*, i. m.

⁴¹ Például a Látners/Rejtett Dirichlet Allokáció (LDA/RDA) alkalmazása középkori okleveles korpuszok vizsgálatánál.

előzményei a nyomtatott szövegek felismerésére kidolgozott gépi megoldásokig (*Optical Character Recognition* – OCR) nyúlnak vissza, egészen az 1950-es évekig. A folyamatosan finomdó szoftveres környezet nagyjából 2010-re összeolvadt a kéziratos szövegek gépi olvasásával (*Handwritten Text Recognition/Transcription* – HTR/HTT), és napjainkra a felülyelt gépi tanulás egyik legfontosabb területévé vált. Bár a HTR-motorokra épülő alkalmazások (mindenekelőtt a *Transkribus*)⁴² kora újkori és újkori kézírást felismerésében bizonyult eddig a leginkább hatékonynak, utóbb a középkori kézíratok feldolgozásában is egyre komolyabb eredményekkel kecsegtet. A technológiai előrehaladást számos projekt (Digipal,⁴³ MOA,⁴⁴ BeCoRe,⁴⁵ eScriptorium,⁴⁶ Teklia⁴⁷) igyekszik hasznosítani, kiaknázni. A fejlődés íve lenyűgöző, ma már a *trained machine learning* akár 90% feletti felismerési arányt is képes felmutatni, bár a tapasztalati képlet szerint ehhez legalább 1000 kézirat-sort kell felismertetni a szoftverrel.⁴⁸ A gépi tanulás itt különösen fontos szerepet kap, hiszen nem pusztán az adott kézírás jellegzetességeit kell felismernie a gépnek, hanem tulajdonképpen a forrásnyelven, alapvetően latinul/németül is tudnia kell, hiszen a túlbujáncz rövidítések feloldása egyáltalán nem egyértelmű. A végződések azonosításához a latin *declinatio*, helyenként *coniugatio* alapos ismerete elengedhetetlen. A jó hír persze az, hogy a gép hajlandó latinul (is) megtanulni.

De ezzel már a web3.0/4.0 tereumán, nem középkorásznak való vidéken járunk. Ez már a jelen és a közeli jövő problematikája – a matematikai/statisztikai/nyelvi modellek utat találnak a középkorász kutatásokba: a rejtett/látens Dirichlet alokáció (RDA/LDA) és téma-modellezés (TM) rendszeres alkalmazása, a természetes nyelvi feldolgozás (NLP) – az adatbányászat (DM), a tartalomelemzés (CA), a szemantikus elemzés (SA) és a hálózatelemzés (NA). A nagy nyelvi modellek (LLM) használata és a generatív AI/MI térnyerése azt az illúziót is keltheti, hogy akkor mostantól a gépek küzdenek a források elolvasásával és feldolgozásával. Jó eséllyel a helyzet ennél valamelyest bonyolultabb.

A fentiekből mindenesetre egy első látásra ijesztő trend látszik kirajzolódni. A történésznek/középkorásznak programozóvá kell avanzálnia, hogy képes legyen kezelni, *horribile dictu*: uralni ezt az online teret, amely bár számos előnnyel kecsegtet, de egyúttal komoly felkészültséget igényel. A fent említett DEEDS-projekt alapjait lefektető programozók Oracle-tanfolyamokat tartottak az érdeklődőknek, a projekt vezetőjének pedig folyamatosan humán nyelvre „fordították” a leprogramozott algoritmusok lényegét. Majd egy ponton XML-alapú kódolásba (*encoding*) kezdtek, új kihívások elé állítva a (kétkezi) középkorászokat. 2020-tól az egyik elemi fontosságú kérdés úgy szölt, hogy a történész képes-e artikulálni kutatási kérdéseit oly módon, hogy a „gép nyelven” gondolkodó programozó képes legyen megfelelő algoritmusokkal előállni. Az alternatív ötlet: kihagyható-e a programozó a folyamatból? A kérdés első közelítésre talán provokatívnak tűnhetett, a tágabb értelemben vett szakma ugyanakkor komolyan gondolta. A választ pedig az időközben kialakuló technológiai feltételek szolgáltatták: *small-coding*. A bölcsészeket és ezen a körön belül a történészek egy csoportját megkísértette a lehetőség, hogy nem kell bonyolult programnyelveket elsajátítaniuk, hanem egyszerű környezetet és megoldásokat biztosító programozási nyelvek (például

⁴² www.transkribus.org

⁴³ www.digipal.eu

⁴⁴ www.modelsofauthority.ac.uk

⁴⁵ shs.hal.science/halshs-04968987

⁴⁶ escriptorium.rich.ru.nl

⁴⁷ tekli.com

⁴⁸ Stutzmann, Dominique: „AI, Image Analysis, and Medieval Manuscripts”, előadás, *4th International Conference: Palaeography and Diplomats*, 2025. október 6–7. Évora, Portugália.

Python)⁴⁹ révén kezelhetik adatbázisaikat. AI/MI támogatással ez az elképzelés egyáltalán nem megalapozatlan, néhány hét alatt megismerhetők olyan keretrendszerek, amelyek hatékony adatbázis-kezelésre, adatbányászatra kínálnak lehetőséget – feltéve, hogy kellően hatékonyan és pontosan sikerül promptolni a generatív AI-t.

Végül, de koránt sem utolsósorban arról illene néhány elképzelést vázolni, hogy merre tart az online térben stabilan jelen lévő medievista szakmai közeg. Első közelítés gyanánt egy disztópikus forgatókönyv is összeállítható: a középkorászok már nem olvasnak kéziratot, nem tudnak forrásnyelveken – a gép olvas, a kézírás-felismerő (HTR) átírja, lefordítja, elemzi, majd a generatív AI megírja a tanulmányt veretes konklúziókkal. Ennek eshetőségét e helyt nem szívesen latolgatnánk. Helyette inkább egy – tetszőlegesen kiterjeszthető – modell felvázolására vállalkozunk, amely, reményeink szerint, hatékonyan ki tudja aknázni napjaink online tereinek lehetőségeit.

Illusztratív példa lehet a Szegeden 1990-ben indult *Anjou-kori oklevéltár (AOkl)* regesztá-gyűjteménye,⁵⁰ amely mára 54 nyomtatott kötetben közel 43 000 magyar nyelvű szöveges kivonatot tartalmaz az 1301–1369 közötti időszakból. Az állomány digitális feldolgozására már a 2000-es évek elején, a web2.0 korszakában történt kísérlet, az addigra elkészült kötetek közel 18 000 regesztájából készült egy XML-alapú adatbázis, amely pilot-projekt-ként a világhálón is elérhető volt – az SZTE egyetemi könyvtárának *Eruditio* könyv- és olvasmánytörténeti adatbázisai közé „elrejtve”. A szakma egy része használta is ezt a korpuszt „tesztüzemmódban”. A korábban vázolt szemantikus web immár lehetőséget kínál arra, hogy a dokumentumok helyett az adatok adatbázisa jöjjön létre, amelyben összekapcsolódhat a fennmaradt oklevelek digitális fotója, a dokumentum teljes szövegű (*in extenso*) kiadása akár az *Anjoukori okmánytár*ból (AO),⁵¹ akár más forráskiadványból származik, a fordítások, az *AOkl* regesztái, a facsimile-kiadások és egyéb szöveges vagy képi reprezentációk. A tervezett és elindult vállalkozás jóval szélesebb adatkapcsolati háló kialakítására tesz kísérletet. A regeszták szövegének adatbányászati módszerekkel történő feldolgozása, elemzése lehetőséget kínál a szöveges kivonatokban található személyek, hivatalok/méltóságok, helynevek, az adott dokumentumban tárgyalt ügyek fogalmainak, sőt az oklevélen található pecsétek azonosítására. A kapcsolódási szálak pontos köre teljességre törekvően még nem határozható meg, de a lehetőségek kifejezetten biztatóak. Ehhez persze a legtágabb értelemben vett szakma paradigmaváltására, azaz annak felismerésére is szükség van, hogy a digitális bölcsészet (DH) nem pusztán számítógépes eszközökkel támogatott kutatás, hanem egy számos tekintetben új megközelítés, új módszertani alapállás. Köz helyé egyszerűsítve, nem az adathalmazban lévő kapcsolatokat kell lemodellezni, hanem web-ontológiákra és szemantikus web-technológiákra építve hálózatos reprezentációkat kell kialakítani. A korábban szóba hozott *Linked Open Data* (LOD), illetve például a CIDOC CRM⁵² már nem pusztán adatmezőket definiál, hanem viszonyokat: ki, mikor, hogyan, mivel áll kapcsolatban. Ezt talán ismét célszerű egy példával megvilágítani, bármilyen partikulárisnak tűnik is első közelítésre.

A web2.0-ás pilot-projektben létrehozott Anjou-regeszták adatbázisának továbbépítése és átstrukturálása 2022-ben indult el, és már a tervezés korai fázisában megfogalmazódott a kérdés, hogy az okleveleken található pecsétek – akár ma már nem létező/ismert, de a dokumentumok *corroboratió*jából rekonstruálható és ily módon a regesztákban is megőrzött

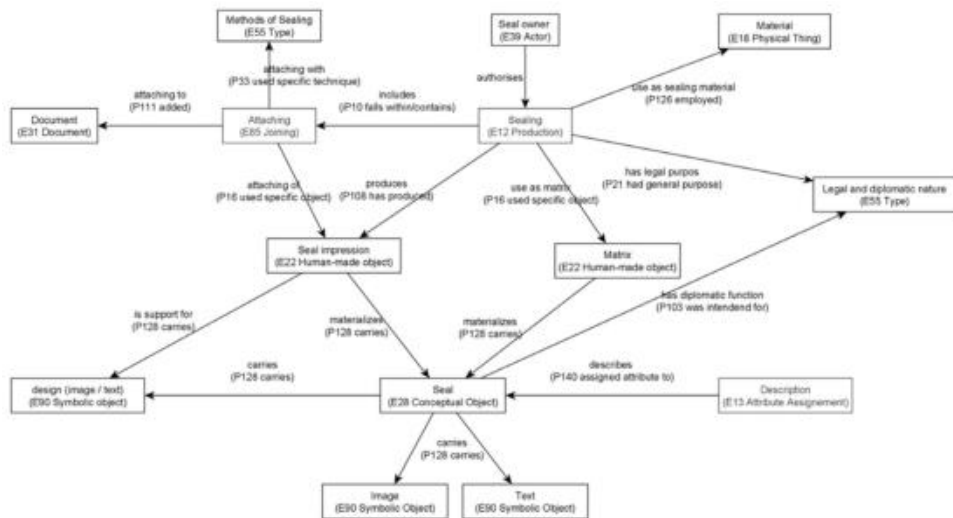
⁴⁹ www.python.org

⁵⁰ Kristó et al.: *Anjou-kori oklevéltár*, i. m.

⁵¹ Nagy et al.: *Anjoukori okmánytár*, i. m.

⁵² Comité international pour la documentation – Conceptual Reference Model (cidoc-crm.org).

hitelesítő eszközök – milyen adatmodellbe illeszthetők be. A problémakör már csak azért is összetett, mert a „hagyományos” módon sem egyszerű a nemzetközi szakmai közegben egyetemesen használt fogalomkészletet minden paraméterre alkalmazni. Erre ugyan már történtek komoly erőfeszítések az 1960-as évek végétől, de csak 1990-ben született meg egy többnyelvű „szótár” (*Vocabulaire international de la sigillographie* – VIS),⁵³ amely nagy segítséget nyújtott a pecsét(kép)ek leírásában. Ugyanakkor már ekkor jelentkezett néhány részben módszertani probléma. Az egyik fontos nehézség az, hogy még a teljességre törekvés igénye sem eredményezett teljes fogalmi korpuszokat, így azokat vagy folyamatosan, vagy valamilyen rendszerességgel bővíteni kell, ahogyan ez meg is történt 2016-ban.⁵⁴ De ez sem oldotta fel azt a módszertani problémát, hogy bár a pecséthasználát nagy vonalaiban többnyire egységes volt a középkori Európában, számos apró eltérés mégis megfigyelhető, például az autentikus pecsét definiálásában.⁵⁵ Ennek áthidalására kezdte Georg Vogeler 2013-ban digitális adatmodellé alakítani a VIS-t,⁵⁶ 2025-re elérve a web-ontológia szintjét, ahol az eltérő lokális jellemzők is érvényesíthetővé válnak, sőt sikerült beilleszteni a CIDOC CRM modellbe. Ez látható az alábbi ábrán, amely az entitások, illetve azok osztályai és tulajdonságai mellett a predikátumokat is reprezentálja, ily módon nem pusztán a pecséteket, hanem a pecséthasználatot is modellezi szemantikusan.



A VOCABULAIRE DE LA SIGILLOGRAPHIE-BŐL LÉTREHOZOTT FOGALMI MODELL ÉS KAPCSOLATA
A CIDOC CRM-MEL.⁵⁷

⁵³ Bautier – Ricci-Noè: *Vocabulaire international de la sigillographie*, i. m.

⁵⁴ Müller–Vrtel: *Vocabularium internationale sigillographicum*, i. m. És akkor még nem említettük a pecsétképek ikonográfiai leírásának problematikáját, amely bizonyosan nem kerülhető meg. Lásd például: Posthumus–Brandhorst. 2006. *ICONCLASS*. 4th ed. (iconclass.org).

⁵⁵ Kumorovitz: *A magyar pecséthasználát története a középkorban*, 13–14.; Vogeler: *Transforming the Vocabulaire de la Sigillographie*, 13.

⁵⁶ Ezt nem közvetlenül, hanem egy ontológia-közi metaadatmodell, ún. SKOS (*Simple Knowledge Organization System*) segítségével. Vogeler: *Von der Terminologie zur Ontologie*, 281–297.

⁵⁷ Vogeler: *Transforming the Vocabulaire de la Sigillographie*, 14.

Hasonló szemantikus mátrixok kialakítását képzeljük el az oklevelekben található hivatal-, illetve méltóságviselőkkel kapcsolatban is, integrálva a létező és elérhető prozopográfiai és archontológiai adatbázisokat. E téren ráadásul a magyar medievisztika nagyon komoly eredményeket tud felmutatni az utóbbi évtizedekben, és a lendület még mindig tart.⁵⁸ Az adatok összekapcsolása látványosan növelné a kutatás hatékonyságát, különösen akkor, ha a szemantikus térben majdan nemcsak az Anjou-kori oklevelek, hanem az Árpád- és Zsigmond-kori dokumentumok publikált korpuszai is megjelennek.

Ugyanakkor a már meglévő és szövegszinten vizsgálható gyűjtemények jelenleg is szóra birtokosak az adat- és szövegbányászat különböző eszközeivel. Például a gépi tanulás (ML) segítségével normalizált szövegek tömeges feldolgozására a szemantikus web kiépítésével párhuzamosan is lehetőség van, sőt a folyamatok egymástól függetlenül is működnek.⁵⁹ Ezt példázza az Anjou-kori oklevélregeszták korpusza is, amelynek adatbányászata a fent említett mátrix kialakításával párhuzamosan zajlik egy a frissen a nemzetközi „piacra” kilépő hazai fejlesztésű *data mining* szoftver, az AVOBMAT segítségével.⁶⁰ A korpusz előkészítése során a legtöbb módszertani kihívással szembesülünk, amely a szemantikus web kialakításánál egyébként is felmerül. Ezt azért is fontos hangsúlyozni, mert még a szakmai közbeszédben is terjedőben van egyfajta hurra optimizmus, amely a generatív AI/MI-t mindenható, de legalábbis mindent (is) megoldó szerepben képzei el. Az intelligenciát szimuláló, ma még többnyire fekete dobozokként működő neurális hálók fejlődésének határai napjainkban még nem feltétlenül megjósolhatóak, de a történészek és részhalmozuk, a középkorászok tábora inkább a statisztikai mintafelismerésen alapuló gépi tanulást használja akár a predikátumok létrehozása, akár az osztályozások során. Természetesen az átjárás az AI/MI felé korántsem kizárt, hiszen a fent említett pecsétkép-leírások már átvezetnek a kép/objektumfelismerés és -leírás birodalmába.⁶¹

Zárszóként újra hangsúlyozni szeretnénk, hogy a fenti áttekintés semmiképpen nem kívánt teljességre törekvő bemutatás lenni. Ahhoz ma már monografikus terjedelem lenne szükséges, mégis azt reméljük, hogy a legfontosabb trendeket sikerült láttatnunk: hogy honnan indult a digitalizáció, meddig jutottak a medievisták, és merre haladnak, mert *navigare necesse est...*

BIBLIOGRÁFIA

Nyomtatott/digitális források

- Kristó Gyula et al. (szerk.): *Anjou-kori oklevéltár*. 1301–1369. I–LV. Budapest–Szeged, 1990–2025.
 Migne, Jean-Paul (ed.): *Patrologia Latina database*. 5 CD-ROM, Alexandria (VA): Chadwyck-Healey, 1995.
 Nagy Imre et al. (szerk.): *Anjoukori okmánytár*. *Codex diplomaticus Hungaricus Andegavensis*. I–VII. (1301–1359). Budapest, 1878–1920.

⁵⁸ Legutóbb lásd C. Tóth: *A váci székeskáptalan archontológiája*, i. m.; Ribi–Thoroczkay: *A középkori magyar egyház intézménytörténeti és archontológiai kézikönyve*, i. m.

⁵⁹ Módszertanilag mindenképpen, de az erőforrások korlátozottsága miatt a költséghatékonyság mindvégig fontos szempont marad bármilyen feldolgozás során.

⁶⁰ Analysis and Visualization of Bibliographic Metadata and Texts (avobmat.hu).

⁶¹ Erre nyújt izgalmas, egészen friss példát a középkori órás-könyvek adatbázisának feldolgozása: heurist.huma-num.fr/stutzmann_horae/web/378891, illetve francia közönyvtárakban őrzött középkori kéziratok iniciáléinak vizsgálata: initiale.irht.cnrs.fr

Szakirodalom

- Bautier, Henri Robert – Stefania Ricci-Noè, Stefania (dir.): *Vocabulaire international de la sigillographie*. Pubblicazioni degli archivi di stato 3. Roma: Ministero per i beni culturali e ambientali, 1990.
- C. Tóth, Norbert: *A váci székeskáptalan archontológiája 1200–1543*. Subsidia ad historiam medii aevi Hungariae inquirendam 16. Budapest: HUN–REN Magyar Kutatási Hálózat, 2023.
- Gervers, Michael: *The DEEDS Project and the Development of a Computerized Methodology for Dating Undated English Private Charters of the Twelfth and Thirteenth Centuries*. In: Dating undated medieval charters. Ed. Michael Gervers. Woodbridges: Boydell Press, 2000, 13–36.
- Gervers, Michael – Tilahun, Gelila – Khoshraftar, Shima – Mitchell, Roderick A.: *The Dating of Undated Medieval Charters*. Archives: The Journal of the British Records Association, 53. (2018) No. 137. 1–33. <https://doi.org/10.3828/archives.2018.7>
- Hodel, Tobias: *Supervised and Unsupervised: Approaches to Machine Learning for Textual Entities*. IE: Archives, Access and Artificial Intelligence: Working with Born-Digital and Digitized Archival Collections, ed. Lise Jaillant. Bielefeld: Transcript, 2022. 157–178.
- Kumorovitz Lajos Bernát. *A magyar pecséthasználat története a középkorban*. Budapest: Magyar Nemzeti Múzeum, 1993.
- Müller, Karel – Vrtel, Ladislav: *Vocabularium internationale sigillographicum*. Comp. Lenka Pavlíková. Bratislava: Veda, vydavateľstvo Slovenskej akadémie vied, 2016.
- Poulter, Alan: *The design of World Wide Web search engines: a critical review*. Program: electronic library and information systems, 31. (1997) No. 2. 131–145. <https://doi.org/10.1108/EUM00000000006884>
- Rácz György: *A Magyar Országos Levéltárban őrzött 1526 előtti levéltári anyag épülő adatbázisa. Beszámoló és javaslat*. In: Tanulmányok Borsa Iván tiszteletére. Szerk. Csukovits Enikő. Budapest: Magyar Országos Levéltár, 1998. 181–196.
- Ribi András – Thoroczky Gábor (szerk.): *A középkori magyar egyház intézménytörténeti és archontológiai kézikönyve, 1000–1387*. I–II. Pécs: Kronosz, 2024.
- Vogeler, Georg: *Transforming the Vocabulaire de la Sigillographie into a Semantic Web Resource*. Digital Medievalist, 18. (2025) No. 1. 1–23. <https://doi.org/10.16995/dm.17137>
- Vogeler, Georg: *Von der Terminologie zur Ontologie. Das 'Vocabulaire international de la diplomatique' als Ressource des Semantic. Web*. Francia, 40. (2013) 281–297. <https://doi.org/10.11588/FR.2013.0.40966>

Internetes források

- Arcanum – www.arcanum.com/hu
- A régi magyar vers repertórium (Répertoire de la poésie hongroise ancienne - RPHA) - f-book.com/rpha/v7/index.php
- AVOBMAT – Analysis and Visualization of Bibliographic Metadata and Texts – avobmat.hu
- BeCoRe – Between Composition and Reception: the Authority of Medieval Charters – shs.hal.science/halshs-04968987
- Britannica online – www.britannica.com
- Charters Encoding Initiative (CEI) – www.cei.lmu.de/project
- CIDOC CRM – Comité international pour la documentation - Conceptual Reference Model – cidoc-crm.org
- DEEDS – deeds.utoronto.ca
- DigiMap – digimap.edina.ac.uk
- DigiPal – www.digipal.eu
- eScriptorium – escriptorium.rich.ru.nl
- Europeana – www.europeana.eu
- Hill Museum & Monastic Library – hmdl.org/collections/repositories/malta/national-library-of-malta
- HORAE – heurist.huma-num.fr/stutzmann_horae/web/378891
- Hungaricana – www.hungaricana.hu
- Icarus – www.icar-us.eu
- Initiale – Catalogue de manuscrits enluminés – initiale.irht.cnrs.fr

In principio – apps.brepolis.net/inpr/Help/EN/INPR_manual_EN.pdf (utolsó hozzáférés: 2025. október 3.)

International Medieval Bibliography – about.brepolis.net/imb, ahc.leeds.ac.uk/medieval-research-innovation/doc/international-medieval-bibliography

KVK – Karlsruhe Virtual Catalog – kvk.bibliothek.kit.edu

Linked Open Data (LOD) – data.europa.eu/sites/default/files/d2.1.2_training_module_1.2_introduction_to_linked_data_en_edp.pdf (utolsó hozzáférés: 2025. október 6.)

Mapping Gothic France – mcid.mcah.columbia.edu/mapping-gothic

MOA (Models of Authority) – www.modelsofauthority.ac.uk

MOKKA – Magyar Országos Könyvtári Közös Katalógus – mokka.hu

MoM – www.monasterium.net/mom/home

Monastic Matrix – arts.st-andrews.ac.uk/monasticmatrix

Posthumus, Etienne – Hans Brandhorst. 2006. *ICONCLASS*. 4th ed. iconclass.org (utolsó hozzáférés: 2025. október 3.)

Python – www.python.org

RDF 1.2 – www.w3.org/TR/rdf12-concepts

Regesta Imperii Online – opacfix.regesta-imperii.de

Teklia – teklia.com

The Archives Portal Europe – www.archivesportaleurope.net

Transkribus – www.transkribus.org

Wikipedia – hu.wikipedia.org

World Wide Web Consortium – www.w3.org